ELSEVIER

Inorganica Chimica Acta 243 (1996) 271–282

Inorganica Chimica Acta

Tunneling matrix element in Ru-modified blue copper proteins: pruning the protein in search of electron transfer pathways[1]

John N. Gehlen[a], Iraj Daizadeh[a], A.A. Stuchebrukhov[a], R.A. Marcus[b,*]

[a]*Department of Chemistry, University of California, Davis, CA 95616, USA*
[b]*A.A. Noyes Laboratory of Chemical Physics California Institute of Technology, Pasadena, CA 91125, USA*

Abstract

We investigate with semi-empirical extended Hückel theory calculations the tunneling matrix element for electron transfer in three ruthenium-modified blue copper azurin molecules from the bacterium *Pseudomonas aeruginosa* which have been recently synthesized and studied experimentally by Gray and co-workers. All of the atoms in the protein can be included in the calculations with the method of transition amplitudes that has been developed recently. Our particular focus here, however, is to develop procedures that create a truncated protein much smaller than the initial 2000 atom one, the aim being to retain only those amino acids that are important to the electron tunneling mechanism. Such a procedure, which we refer to as 'pruning', is useful, first because it reduces the size of the problem, perhaps allowing for more accurate techniques to be used on the truncated protein, and second because it allows for the identification of the regions in the protein in which the tunneling electron is localized. The pruning procedures enable us to reduce the number of atoms required in an extended Hückel theory analysis of the tunneling mechanism by approximately a factor of 10 over that in the original protein.

Keywords: Ru-modified; Blue copper protein; Electron transfer

1. Introduction

An area of particular interest in the study of the nature of electron transport in biological media is the effect of biological structures such as β-strands and α-helices on the electron transport mechanism. To examine such issues, Gray and co-workers have recently studied electron transfer in the blue copper protein azurin from the bacterium *Pseudomonas aeruginosa* [1]. This protein is composed primarily of a β-barrel of eight nearly parallel β-strands. The electron transfer reaction in their studies is between a buried Cu^+ ion and an externally attached Ru^{3+} complex. The site of attachment for the ruthenium complex is different in the three molecules they have synthesized, and they were able to examine the distance dependence of the super-exchange coupling matrix element H_{DA} when the direct line for electron transfer is nearly parallel to several β-strands.

The distance dependence that Gray and co-workers have found experimentally is weaker than that predicted for electron transfer in proteins by Dutton et al., whose predictions were based on studies of a variety of proteins [2]. This observation has the possible implication that β-strands are more efficient electron transport secondary structures than a somewhat more randomly oriented biological medium. In the present study, we examine electron tunneling in the azurin molecules theoretically, focusing on the determination of the nature of the tunneling barrier and its relation to the β-strands in the azurin molecule.

An exact examination of the tunneling mechanism for the transferring charge would involve quantum mechanical analysis of a system composed of tens of thousands of coupled electrons. Such an analysis is well beyond the current capabilities of theoretical methods. As such, many approximate approaches have been proposed and used for treating biological electron transfer [3–22]. The atomic pathway picture of Beratan and Onuchic represents one possible approach [3–10]. In this empirical picture, which

* Corresponding author.

[1] This paper is dedicated to Professor Harry B. Gray.

SSDI 0020-1693(95)04917-7

avoids any electronic structure calculations, it is assumed that a linear chain of atoms can be identified as the primary path through which the tunneling electron travels. These authors and others have performed analyses of several biological electron transfer reactions in which it is proposed that the tunneling mechanism is dominated by a finite number of such atomic chains [1,3–10]. Such an extreme localization has not in itself been confirmed in electronic structure calculations, but the simplicity of the formulation is very attractive.

With a related perspective, Siddarth and Marcus have developed a pathway picture in which the constituents of the pathways are whole amino acids [18–22]. They discovered in a number of Ru-modified cytochromes and myoglobins that the tunneling electron is localized in a relatively small number of amino acids that they identified using artificial intelligence search procedures. The semi-empirical one-electron extended Hückel Hamiltonian of such a reduced protein, whose parameters have been optimized so as to best mimic multi-electron effects at a mean field level [26–27], could then be directly diagonalized yielding the matrix element for electron transfer. This approach, though pathway in nature, differs from that of Beratan and Onuchic in that it allows the electron to be delocalized over full amino acids rather than over a sequence of atoms in an atomic chain. The utility of using amino acids as the basic units in a pathway was established by comparison with perturbation theory [20].

Recently, Stuchebrukhov and Marcus have introduced the method of transition amplitudes [16,17] which permits the calculation of the tunneling matrix element for electron transfer in the full protein. Its savings come from the realization that the full Hückel Hamiltonian need not be diagonalized to evaluate the coupling matrix element; a sparse matrix linear algebra procedure can be used instead, as we describe later. The method of solution places no limitations on the tunneling mechanism. As such, the method can provide a powerful tool in the evaluation of the validity of pathway pictures. Additionally, both electron and hole transfer are taken into account on the same footing, as also in Refs. [18–22].

The present work utilizes the method of transition amplitudes in all-atom calculations for the full protein (about 2000 atoms). Our particular focus here, however, is to develop procedures that create a truncated protein much smaller than the initial 2000 atom molecule, with the aim of truncating the protein in such a way as to keep only those amino acids that are important for the electron tunneling mechanism. The importance of different amino acids is examined by comparing the tunneling matrix element in truncated molecules in which the amino acids have been removed with that of the full protein. Such a procedure, which we refer to as 'pruning', is useful, first because it reduces the size of the problem, perhaps allowing for more accurate techniques to be used on the truncated protein, and second because the actual location of the remaining amino acids can be used to analyze the tunneling mechanism. This procedure allows us to examine issues such as localized versus delocalized tunneling mechanisms and interference effects between multiple pathways. Our pruning procedure results in a striking reduction of the number of atoms; the remaining atoms, representing approximately 10% of the total number of atoms present in the protein itself, reproduce the full protein matrix element with less than 10% error.

2. Pruning procedure and results

2.1. Full protein calculations

We treat the tunneling with an analysis akin to the first Born approximation of scattering theory, yielding the coupling matrix element [16]:

$$H_{DA} = \langle A|\hat{V}\frac{1}{E-\hat{H}_B}\hat{V}|D\rangle \quad (1)$$

This formulation neglects the direct coupling between donor and acceptor states. Here, H_{DA} is the superexchange coupling matrix element between donor D and acceptor A, $\hat{H}_B$ is the electronic Hamiltonian of the bridging protein, E is the energy of the tunneling electron and V is the coupling between the donor and the acceptor and the bridging protein. The states $|D\rangle$ and $|A\rangle$ are the localized donor and acceptor electronic states, respectively. In the work presented in this paper, we assume the donor and acceptor states to be orbitals centered at the donor and acceptor atoms of s-orbital symmetry. The use of s-orbitals permits a more direct comparison with the pathway results in Ref. [1], which employed σ-pathways. A more detailed analysis in which the valence d shell of the donor and acceptor atoms is included is a feasible extension and has been included in other work on biological electron transfer [16]. We do not discuss this issue in this paper; results for d-orbital calculations on the azurin molecules will be considered elsewhere. Within the Hückel approximation, the matrix $(E-\hat{H}_B)^{-1}$ can be diagonalized and the matrix element reduced to

$$H_{DA}\sum_{\alpha}\langle A|\hat{V}|\alpha\rangle\frac{1}{E-E_\alpha}\langle\alpha|\hat{V}|D\rangle \quad (2)$$

The states $|\alpha\rangle$ are the one-electron eigenstates of the bridging protein with the corresponding eigenvalues E_α.

Eq. (2) represents a complete prescription for the evaluation of the effective coupling matrix element between donor and acceptor. For proteins of even modest size, however, the Hückel basis set becomes too large for diagonalization on current computer hardware. As a consequence, Eq. (2) can no longer be used for such systems. Recently, the method of transition amplitudes has been developed in which the matrix element is evaluated by the

solution of a sparse set of linear equations rather than by diagonalization [16,17]. The procedure that is utilized in this case is to re-express Eq. (1) as

$$H_{DA} = \sum_{ij} A_i^{\dagger}(E\hat{S} - \hat{H}_B)_{ij}^{-1} D_j \quad (3)$$

where $A_i^{\dagger} = \langle A | \hat{V} | i \rangle$, $D_j = \langle j | \hat{V} | D \rangle$ and S is the overlap matrix. The states $|i\rangle$ represent a complete and possibly non-orthogonal basis set for the bridge Hamiltonian. We utilize the non-orthogonal Hückel basis set. The value of H_{DA} can then be found by rewriting Eq. (3) in the form

$$H_{DA} = \sum_i A_i^{\dagger} T_i \quad (4)$$

where the T_is are defined by the linear equations

$$D_j = \sum_i [ES_{ji} - (H_B)_{ji}]T_i \quad (5)$$

Here, S_{ij} is the overlap matrix element $\langle i | j \rangle$ in the chosen basis set.

With the Hückel basis set and Hamiltonian, the values of D_j, $A_i^{\dagger}$ and $[ES_{ji} - (H_B)_{ji}]$ are readily evaluated. The linear equations given by Eq. (5) can then be solved, yielding the values of T_i. These values are then introduced into Eq. (4.), producing the desired effective coupling matrix element H_{DA}. With the Hückel Hamiltonian and basis set, there is a tremendous savings in computer memory usage over the alternative diagonalization contained in Eq. (2), due to the fact that sparse matrix techniques can be used in the solution of Eq. (5). This savings allows even modest modern workstations to be used for solving for effective coupling matrix elements for realistic biological systems.

The computational advantage of the method of transition amplitudes is not without consequence. The diagonalization procedure implied by Eq. (2) yields both the states $|\alpha\rangle$ and their energies E_α. For a given tunneling energy, the importance of a given state $|\alpha\rangle$ can be ascertained from Eq. (2). This feature is particularly useful, because it allows for a determination of the regions of the biological molecule important to the tunneling mechanism, since the wavefunctions $|\alpha\rangle$ are known. In the method of transition amplitudes, however, this information is lost, since the states $|\alpha\rangle$ are not explicitly calculated. It is plausible, however, that many of the atoms that constitute the protein that surrounds the donor and acceptor ions are not important in the actual tunneling mechanism and that the important atoms might in fact constitute only a small percentage of the full protein. If these atoms can be identified using the method of transition amplitudes, the resulting number of atoms may in fact be sufficiently small that a full diagonalization of the Hückel basis as described in Eq. (2) can be performed on the truncated protein, allowing for a detailed study of the electron tunneling mechanism.

In this paper, a two-step process is presented for obtaining such information. This process prunes the protein, so that the essential atoms can be identified. By 'pruning' it is meant that parts of the protein are removed using a given criterion for establishing their importance in the tunneling mechanism. In our pruning procedure, whole amino acids are kept intact. That is, either all atoms of an amino acid are included or they are all excluded. If, due to the removal of an amino acid, an amide linkage between a remaining amino acid and a deleted amino acid is destroyed, a hydrogen is placed at the location of the removed carbon or nitrogen atom of the peptide bond. Thus, if the carbon is removed, a $-NH_2$ group remains in the previously connected amino acid, and if the nitrogen is removed an aldehyde group, –COH, remains. If a ligand of a metal center is removed, the atom which is directly ligated to the metal is replaced with a hydrogen atom. The final pruned protein is similar to the reduced protein model of Siddarth and Marcus [18–22], who used instead an artificial intelligence search to select the important parts of the protein. We discuss next the two-step pruning procedure and its results when applied to the Ru-modified azurin molecules.

2.2. *Protein truncation: tunneling tubes*

Our final goal is to remove from the original protein those amino acids that are not important for the electron transfer process. The primary tool used in achieving this goal is a pruning procedure described in detail in the next section. This procedure examines the importance of each amino acid individually. As a consequence, if the entire protein is pruned with this procedure, the determination of the relevant amino acids requires a matrix element evaluation for each amino acid constituting the protein. This computationally costly process is unnecessary because most of the amino acids in the protein are so far spatially removed from the electron transfer pathway that they will not contribute to the electron tunneling mechanism.

For this reason, as a first step in the protein pruning process, we introduce in this section a truncation procedure which allows one to rapidly eliminate a large portion of the protein matrix that is irrelevant to the tunneling process due to its spatial separation from the electron transfer pathway. This procedure, as we discuss shortly, does so with a relatively small number of matrix element evaluations.

The protein truncation procedure we employ is similar to that described by Okada et al. [15]. The portion of the protein relevant to the tunneling mechanism is selected as follows: a tube of cross-sectional radius R with a curvilinear axis defined by a chosen atomic pathway is created. At its two ends are spherical caps, also of radius R, cen-

tered at the donor and acceptor metal centers, hence creating a tubular region in space. Any amino acid with any atom lying within this tube is retained, all other amino acids being discarded. This selection leaves roughly a bent cylindrically shaped region of the protein of cross-sectional radius R. The truncated protein created in this way will have a cross-sectional radius larger than R, since the full amino acids are retained even when the bulk of the amino acid remains outside the selected tube. An extended Hückel evaluation of the matrix element for this truncated protein is then performed with the method of transition amplitudes, the goal being to find the minimal R for which the matrix element of the full protein can be reproduced. This procedure quickly removes unimportant amino acids because if a tubular region of radius R results in a matrix element converged to the full protein value, all amino acids outside of the tube can be removed with one matrix element evaluation.

Thus, we construct tubes in the protein medium in such a way that the axis or 'backbone' of the tube is a curved path connecting D and A through the protein. The paths we use as backbones of the tubes could be the atom-to-atom paths described in the theory of Beratan and Onuchic. Rather than using their search procedure, we selected intuitively several reasonable and direct atomic chains, qualitatively using their ideas regarding the relative importance of through-bond and through-space jumps in the tunneling process. The selection of these atomic paths is not important for our whole procedure except as a computational tool. That is, the better the initial guess for an atomic pathway, the more computationally efficient our calculations will be in finding the final truncated protein. It should be emphasized that in the methods used in the present article, and in the work of Siddarth and Marcus [18–22], the amino acids are kept whole and, as a consequence, the method can never achieve an atomic wire limit which would require breaking individual side chains that belong to the amino acids.

In order to describe the rational for the paths we have so constructed, it is useful to describe some aspects of the modified azurin molecules [28]. The molecular model discussed here and used in our calculations is the same one used by Gray and co-workers in their atomic pathway analysis [1,28]. The electron transfer reaction in these molecules is between a buried Cu^+ ion and the $Ru^{3+}(bpy)_2im$ complex attached to an exterior histidine of the modified protein. The copper ion is surrounded by five ligands in a distorted trigonal bipyramidal structure. The three equatorial ligands are Cys^{112}, His^{46} and His^{117} and the axial ligands are Gly^{45} and Met^{121} [1].

The sulfur atom of the axial ligand Met^{121} lies a relatively large distance of 3.1 Å from the copper ion in both of the molecules we consider here. Extending from Met^{121} is a β-strand which terminates at Lys^{128}. It is along this β-strand that the ruthenium complex has been attached in the three molecules synthesized by Gray and co-workers. Using site-directed mutagenesis, the protein was altered by the replacement of an amino acid by a histidine to which the Ru-complex was then attached. In three synthesized molecules, one histidine was placed at the amino acid sequence number 122, 124 and 126 with center-to-center donor-acceptor separation distances of 14.9, 20.7 and 26.7 Å, respectively. In this paper, for brevity, only the molecules having the shortest (His^{122}) and the longest (His^{126}) donor-acceptor distance are considered. A more detailed account of the calculations will be presented elsewhere. Gray and co-workers have proposed an atomic pathway for electron transfer that involves a jump across the large distance from the copper atom to the sulfur atom of Met^{121} followed by through-bond jumps along the β-strand extending from this methionine to the histidine to which the ruthenium is ligated [1]. A final jump is made from the histidine to the ruthenium atom. This atomic path will be denoted by P_{121}, due to the involvement of Met^{121}. This pathway, whose atomic composition is given in Table 1, serves as a curvilinear axis for our tubular truncation procedure.

One of the equatorial ligands of the copper atom is the sulfur atom of Cys^{112}. The Cu–S bond lengths for this ligand bond in the molecular models we use are 2.25 Å and 2.45 Å for the molecules with the complex attached to His^{122} and His^{126}, respectively [28,29]. These bond distances are considerably shorter than the axial Cu–S bond distance mentioned above. As such, the electronic coupling between the copper atom and the surrounding medium may be much stronger through this ligand. The strong coupling may result in a possibility of another tunneling pathway, different from P_{121} described above. The strong coupling between the Cu^+ ion and the sulfur atom of Cys^{112} is also supported by recent CNDO calculations for the copper complex [31] and by more detailed studies of the electronic structure of the Cu center in blue copper proteins [32,33]. The Cys^{112} ligand is a part of an additional β-strand that runs parallel to the β-strand to which the ruthenium is attached and through which the pathway P_{121} passes.

In order to explore the possible contribution of the Cys^{112} ligand in the tunneling process in the azurin molecule we construct the pathway P_{112} which begins with a relatively short jump from the copper atom to the sulfur atom of Cys^{112}, followed by through-bond jumps along the parallel β-strand. A through-space jump from this β-strand to the β-strand to which the ruthenium complex is attached must then occur and we choose as the site for this jump a close contact that lies near the histidine ligand that ligates the ruthenium atom. Many other locations for the β-strand to β-strand jump are also possible, but we again stress that the particular choice of the atomic path does not effect the quality of the final result; it only effects the computational effort necessary to arrive at the result. Our particular choice was made because it seemed

Table 1
Pathways based on atomic routes

Molecule	Pathway	Amino acid sequence	% of full protein atoms	$H_{DA}/H_{DA}^{(FULL)}$
His^{122}	P_{121}	$Cu^{+/2+} \rightarrow Met^{121} \rightarrow His^{122} \rightarrow Ru^{3+/2+}$	2.3	0.06 (−10.0 eV) 0.20 (−11.0 eV)
His^{122}	P_{112}	$Cu^{+/2+} \rightarrow Cys^{112} \rightarrow Phe^{111} \rightarrow Phe^{110} \rightarrow Met^{109} \rightarrow His^{122} \rightarrow Ru^{3+/2+}$	5.0	0.28 (−10.0 eV) 1.23 (−11.0 eV)
His^{126}	P_{121}	$Cu^{+/2+} \rightarrow Met^{121} \rightarrow Lys^{122} \rightarrow Gly^{123} \rightarrow Thr^{124} \rightarrow Leu^{125} \rightarrow His^{126} \rightarrow Ru^{3+/2+}$	5.5	−0.12 (−10.0 eV) 2.50 (−11.0 eV)
His^{126}	P_{112}	$Cu^{+/2+} \rightarrow Cys^{112} \rightarrow Phe^{111} \rightarrow Phe^{110} \rightarrow Met^{109} \rightarrow Tyr^{108} \rightarrow Gln^{107} \rightarrow His^{126} \rightarrow Ru^{3+/2+}$	6.9	−0.05 (−10.0 eV) −0.15 (−11.0 eV)

Note: The exact atomic routes corresponding to the above set of amino acids are: His^{122} (P_{121}): $Cu^{+/2+} \rightarrow S_\delta\, C_\gamma\, C_\beta\, C_\alpha\, C \rightarrow N\, C_\alpha\, C_\beta\, C\gamma\, C_\delta\, N_\varepsilon \rightarrow Ru^{3+/2+}$; His^{122} (P_{112}): $Cu^{+/2+} \rightarrow S_\gamma\, C_\beta\, C_\alpha N \rightarrow C\, C_\alpha\, N \rightarrow C\, C_\alpha\, N \rightarrow C\, C_\alpha\, N \rightarrow C_\alpha\, C_\beta\, C_\gamma\, C_\delta\, N_\varepsilon \rightarrow Ru^{3+/2+}$; His^{126} (P_{121}): $Cu^{+/2+} \rightarrow S_\delta\, C_\gamma\, C_\beta\, C_\alpha\, C \rightarrow N\, C_\alpha\, C \rightarrow N\, C_\alpha\, C \rightarrow N\, C_\alpha\, C \rightarrow N\, C_\alpha\, C \rightarrow N\, C_\alpha\, C_\beta\, C_\gamma\, C_\delta\, N_\varepsilon \rightarrow Ru^{3+/2+}$; His^{126} (P_{112}): $Cu^{+/2+} \rightarrow S_\gamma\, C_\beta\, C_\alpha\, N \rightarrow C\, C_\alpha\, N \rightarrow C\, C_\alpha N \rightarrow C\, C_\alpha\, N \rightarrow C\, C_\alpha\, N \rightarrow C\, C_\alpha\, N \rightarrow C_\alpha\, C_\beta\, C_\gamma\, C_\delta\, N_\varepsilon \rightarrow Ru^{3+/2+}$.

a reasonable one. The detailed atomic pathway for P_{112} is given in Table 1.

Finally, the pathway $P_{121+112}$ was constructed which is composed of the atoms that constitute both paths P_{121} and P_{112}. This path is useful because it allows both paths to contribute simultaneously to the electron transfer mechanism.

The results for the tubes constructed around the three pathways, P_{121}, P_{112} and $P_{121+112}$, are presented in Fig. 1. For comparison, calculations for tubes constructed around the straight line connecting the centers of the donor and acceptor metal centers are also shown [15]. The ratio $H_{DA}(R)/H_{DA}^{(FULL)}$ is plotted versus R, where R is the cross-sectional radius of the tube, $H_{DA}(R)$ is the effective coupling matrix element when the donor and acceptor are surrounded by a truncated protein with a tube radius R and $H_{DA}^{(FULL)}$ is the effective coupling matrix element when the donor and acceptor are surrounded by the full protein. Hence, in the limit of large R, all pathways should tend to a value 1, as is displayed in Fig. 1. A tube of radius $R \approx 15$ Å encloses all the amino acids constituting the protein.

All matrix elements shown in Fig. 1 were computed with the method of transition amplitudes as given by Eqs. (3)–(5) with a tunneling energy E of −11.0 eV. The convergence results are similar for all tunneling energies within the gap between the occupied and unoccupied orbitals, i.e. for energies between −11.7 eV and −9.8 eV in our model calculations. When the relevant charge transfer spectrum is known, the tunneling energy can be

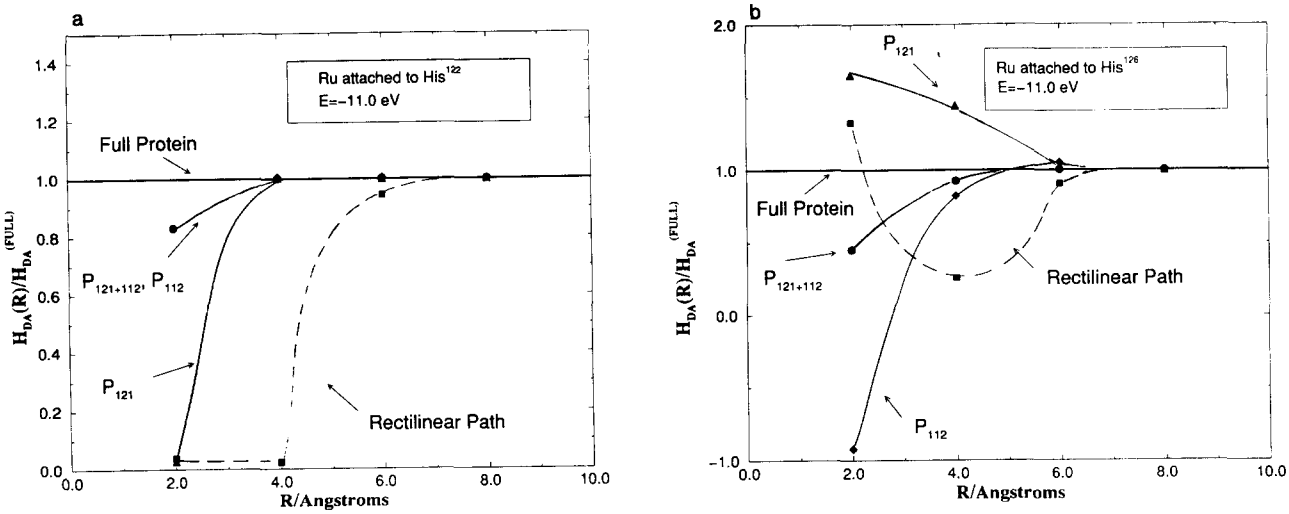

Fig. 1. (a) Convergence of the effective coupling matrix element as a function of the tunneling tube radius, R in His^{122}. The quantity $H_{DA}(R)/H_{DA}^{(FULL)}$ for tubes of radii R of 2, 4, 6 and 8 Å about amino acid path P_{112}, P_{121} and $P_{121+112}$ are presented. These paths are defined by the sequence of atoms given in Table 1. The results for the straight line tube path between donor and acceptor are also presented. The horizontal solid line is the full protein result. The results for P_{112} and $P_{121+112}$ coincide with one another to the resolution of the graph. (b) Same as (a), but for His^{126}. Here, the results for P_{112} and $P_{121+112}$ do not coincide.

determined [16,18–22]. In the present paper, the Cu^{+}-L $\rightarrow$ Cu^{2+}-L^{-} charge transfer spectrum is not known, but results are available on the Cu^{+}-L $\rightarrow$ Cu-L^{+} spectrum [32,33], which is relevant to the hole transport contribution in the protein. We plan to explore its use in the present system in the later and more detailed article. In the meanwhile, we have performed calculation for several tunneling energies within the energy gap.

In Fig. 1a, results are presented for the molecule in which the ruthenium complex is attached at the His^{122} site and the results for the His^{126} location are given in Fig. 1b. A detailed examination of these results is given in the next section. It is noted here that the convergence to the full protein result is qualitatively similar for both molecules, with a convergence to the full protein results with R approximately 4 Å for the His^{122} attachment and approximately 6 Å for the attachment at His^{126}. These 4 and 6 Å tubes contain, respectively, approximately 450 and 775 atoms, or about 20 and 40% of the full protein's 2000 atoms. The reduction in number of atoms is significant.

It is also noted that the convergence is qualitatively similar for all of the paths shown in Fig. 1. This is a result of the significant overlap of the generated tubes around the pathways P_{112} and P_{121}. Neither of the pathways alone accurately reproduces the full protein result as can be seen in Table 1. Hence, it can be concluded that the pathways P_{112} and P_{121} themselves, though in the region of the protein important to the tunneling mechanism, do not incorporate all necessary atoms. One exception is the amino acid sequence corresponding to P_{112} in the His^{122} molecule. At a tunneling energy of −11.0 eV, this intuitively chosen pathway reproduces the full protein matrix element with less than 25% error as is shown in Table 1. Inspection of Tables 1 and 2 show that the amino acid sequence corresponding to P_{112} is almost the same as that generated by the amino acid pruning procedure.

The greatest difference in convergence rates between pathways is seen in Fig. 1a where P_{112} gives converged results for $R = 4$ Å and the straight line path converges with the relatively large $R = 8$ Å. This may appear to be a small difference, but computationally it is significant because, in a tube, the number of amino acids remaining can be expected to grow as the square of R. Hence, in the 8 Å tube required for the straight line path, many more amino acids will remain than in the 4 Å tube required by P_{112}. This can result in large computational savings in the pruning procedure presented in the next section. The results presented in Fig. 1b, however, show little improvement over the straight line path for the atom-to-atom pathways we have chosen. Hence, intuitively chosen curvilinear pathways give results that are no worse than the straight line path (Fig. 1b) and on some occasions, the convergence of the results is considerably better, Fig. 1a. Nevertheless, even though the straight line pathway sometimes has slower convergence with R than the intuitively chosen pathways, its simplicity makes it our preferred choice. The truncation procedure results, regardless of axis pathway, in a dramatic reduction in the number of amino acids in the truncated molecule and the first pruning step is extremely useful in reducing the computational effort required to prune the protein.

Subtle but important differences that can be noticed with a tube radius of 4 Å between the differing pathways and differing molecules in Fig. 1 are discussed in Section 3. Implications of these effects on the tunneling mechanism are also discussed in that section.

2.3. Amino acid pruning

Though by changing the curvilinear axis of the tubes, the truncation method serves as a valuable probe of the inhomogeneity of the tunneling mechanism, further insight into the tunneling process can be obtained by calculating the relative importance of individual amino acids residing in the truncated molecule. Such an amino acid based perspective in the analysis of electron transfer in proteins has been considered previously by Siddarth and Marcus [18–22], Okada et al. [15], and others [11]. In this section a pruning procedure is devised in which the con-

Table 2
Pruned molecules with $P_{cut} = 0.1$

Molecule	E	Pruned amino acid sequence	% of full protein atoms	$H_{DA}/H_{DA}^{(FULL)}$
His^{122}	−10.0 eV	$Cu^{+/2+}$, Phe^{111}, Cys^{112}, Met^{121}, His^{122}, Gly^{123}, $Ru^{3+/2+}$ [All Ru ligands are present]	6.5	0.90
His^{122}	−11.0 eV	$Cu^{+/2+}$, Phe^{111}, Cys^{112}, His^{122}, Gly^{123}, $Ru^{3+/2+}$ [All Ru ligands are present]	5.7	0.98
His^{126}	−10.0 eV	$Cu^{+/2+}$, His^{46}, Gln^{107}, Tyr^{108}, Met^{109}, Phe^{110}, Phe^{111}, Cys^{112}, Thr^{113}, Met^{121}, Lys^{122}, Thr^{124}, Leu^{125}, $Ru^{3+/2+}$ [All Ru ligands except for imidazole and histidine are present]	13.1	1.03
His^{126}	−11.0 eV	$Cu^{+/2+}$, His^{46} Gly^{105}, Glu^{106}, Gln^{107}, Tyr^{108}, Met^{109}, Phe^{110}, Phe^{111}, Cys^{112}, Thr^{113}, Met^{121}, Lys^{122}, Gly^{123}, Thr^{124}, Leu^{125}, His^{126}, $Ru^{3+/2+}$ [All Ru ligands except for imidazole are present]	15.3	0.98

tribution of individual amino acids to the tunneling mechanism is explored.

We begin with the truncated protein from the previous section. An amino acid is then selected and removed from the truncated protein. The rupture of the polypeptide bonds is accompanied by hydrogen atom substitution as described in the previous section. Removal of this amino acid generates a more truncated protein to which the semi-empirical Hückel calculation previously discussed is applied, yielding an effective coupling matrix element $H_{DA}^{(i)}$, which denotes the matrix element with the deletion of the amino acid designated by the number i. If the deleted amino acid is important to the tunneling mechanism in the full protein, $H_{DA}^{(i)}$ will be appreciably different from $H_{DA}^{(FULL)}$. On the other hand, if the amino acid is unimportant for the tunneling, $H_{DA}^{(i)}$ will be nearly equal to $H_{DA}^{(FULL)}$.

This procedure is repeated until each of the amino acids in the original truncated molecule has been so examined. For example, if the starting protein contains 40 amino acids, 40 molecules are created each of which is missing a single amino acid. The matrix element $H_{DA}(i)$ is evaluated in each of these molecules. We construct from this result the quantity [34].

$$p_i = \frac{H_{DA}^{(i)} - H_{DA}^{(FULL)}}{H_{DA}^{(FULL)}} \tag{6}$$

A large magnitude for p_i indicates that the amino acid labeled by i is important to the tunneling mechanism, and a small value implies, of course, an unimportant amino acid. We use the value of p_i as a pruning criterion, as discussed later. The importance of the first truncation step in the pruning procedure can now be appreciated, for in the detailed pruning of this section, the removal of each amino acid requires an additional evaluation of a matrix element, while in the previously discussed truncation procedure, many amino acids can be removed with a single evaluation.

The values of p_i as given by Eq. (6) are presented in Fig. 2. The amino acid number i is a dummy index and is not equal to the amino acid sequence number. The sequence number for those amino acids with large $|p_i|$ are given in Fig. 2 and in its legend, and also in Table 2. In Fig. 2a,b, results are given when we begin with the molecule defined with the 6 Å tube about the pathway $P_{121+112}$. The results in these figures are for the molecules in which the ruthenium complex is attached at His^{122} and His^{126}, respectively. In the His^{122} truncated molecule, there are approximately 40 remaining amino acids and in the His^{126} molecule there are approximately 55 amino acids, out of a total of 128 amino acids in the full protein. The tunneling matrix element between donor and acceptor in these 6 Å tubular molecules differs from the tunneling matrix element for the full protein by less than 1%. Results for tunneling energies of −11.0 eV and −10.0 eV are given in Fig. 2a,b and in Table 2. A striking result is that only a very small fraction of the amino acids constituting the truncated protein appreciably affects the tunneling matrix element when deleted.

In the pruning procedure the magnitude of p_i is used as the pruning criterion and the fate of an amino acid is determined with the rules:

$$\begin{aligned} |p_i| &\geq p_{cut} \quad \text{Retain amino acid} \\ |p_i| &< p_{cut} \quad \text{Delete amino acid} \end{aligned} \tag{7}$$

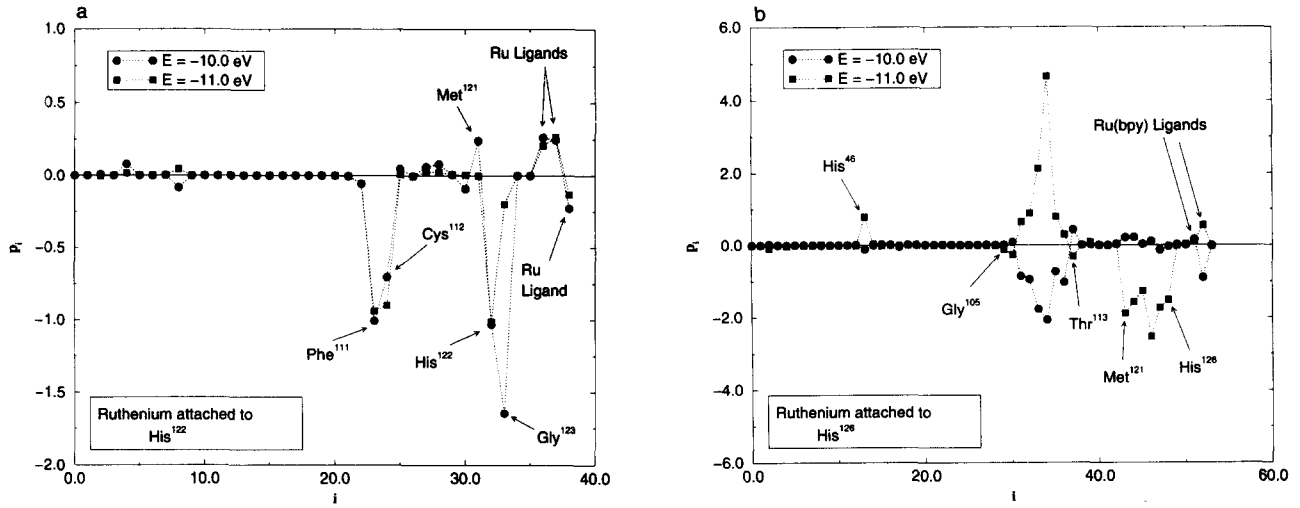

Fig. 2. (a) The value of p_i given by Eq. (6) plotted versus the removal of the amino acid labeled i from the truncated protein in His^{122} is given. The dotted line with solid circles is for $E = -10.0$ V and the dotted line with solid squares is for $E = -11.0$ V. All amino acids with $|P_{cut}| \geq 0.1$ are labeled with their amino acid type and sequence number. (b) Same as (a). Several amino acids with $|P_{cut}| \geq 0.1$ are labeled in the figure and a complete identification of the important amino acids can be obtained by noting that the amino acids that lie between the labeled Gly^{105} and Thr^{113} are, from left to right, Glu^{106}, Gln^{107} Tyr^{108} Met^{109} Phe^{110} Phe^{111} and Cys^{112} and those between the labeled Met^{121} and His^{126} are, from left to right, Lys^{122}, Gly^{123}, Thr^{124} and Leu^{125}.

Here, the p_{cut} is the pruning parameter whose value is varied to control the extent to which the protein is pruned and the vertical bars denote absolute value. A small value of p_{cut} will keep the original truncated molecule intact, while a large value will keep only 'key' amino acids.

The pruning procedure governed by Eqs. (6) and (7) dramatically reduces the number of amino acids required to reproduce the full protein matrix element, generating a savings of more than a factor of three in number of atoms over the tubular pruning method. We find that for the molecules we consider, a value $p_{cut} \approx 0.1$ yields pruned molecules that reproduce the full protein matrix element with a relative error of less than 10%, often significantly less as can be seen in Table 2. The resulting collections of amino acids in the pruned molecules, with $p_{cut} = 0.1$, are listed in Table 2. Since the tunneling pathways are composed of whole amino acids in our approach, this collection of amino acids in the pruned protein could be called a pruned pathway. The amino acids for the pruned molecules are listed in Table 2. These pruned molecules, most notably the small number of amino acids of which they are composed, and the pruning procedure we use are the principal results of the present paper. We discuss the implications of our results in the next section.

3. Discussion and conclusions

In this section, several aspects of the results from this work are discussed. A qualitative discussion is given of the success of the attempt to reduce the number of amino acids needed to reproduce the full protein matrix element, and of its potential utility. The section concludes with a discussion of the implications for the actual tunneling mechanism, focusing, in particular, on the presence of amino acid pathways.

A two-step pruning algorithm was presented above for the determination of the minimum number of atoms required to reproduce the full protein matrix element. The first step, which generated a tubular portion of the protein in which the tunneling electron is localized, was successful in reducing the number of atoms from the full protein by a factor of 3–5. The ideas of Okada et al. were extended by allowing for a more general pathway axis than a single straight line path between donor and acceptor, namely a curvilinear one that can be chosen to follow a chosen atom-to-atom pathway between donor and acceptor. The results in Fig. 1a,b indicate that qualitatively there was no considerable improvement by incorporating these new tubular pathways. The first step is important because it significantly reduces the number of amino ac-

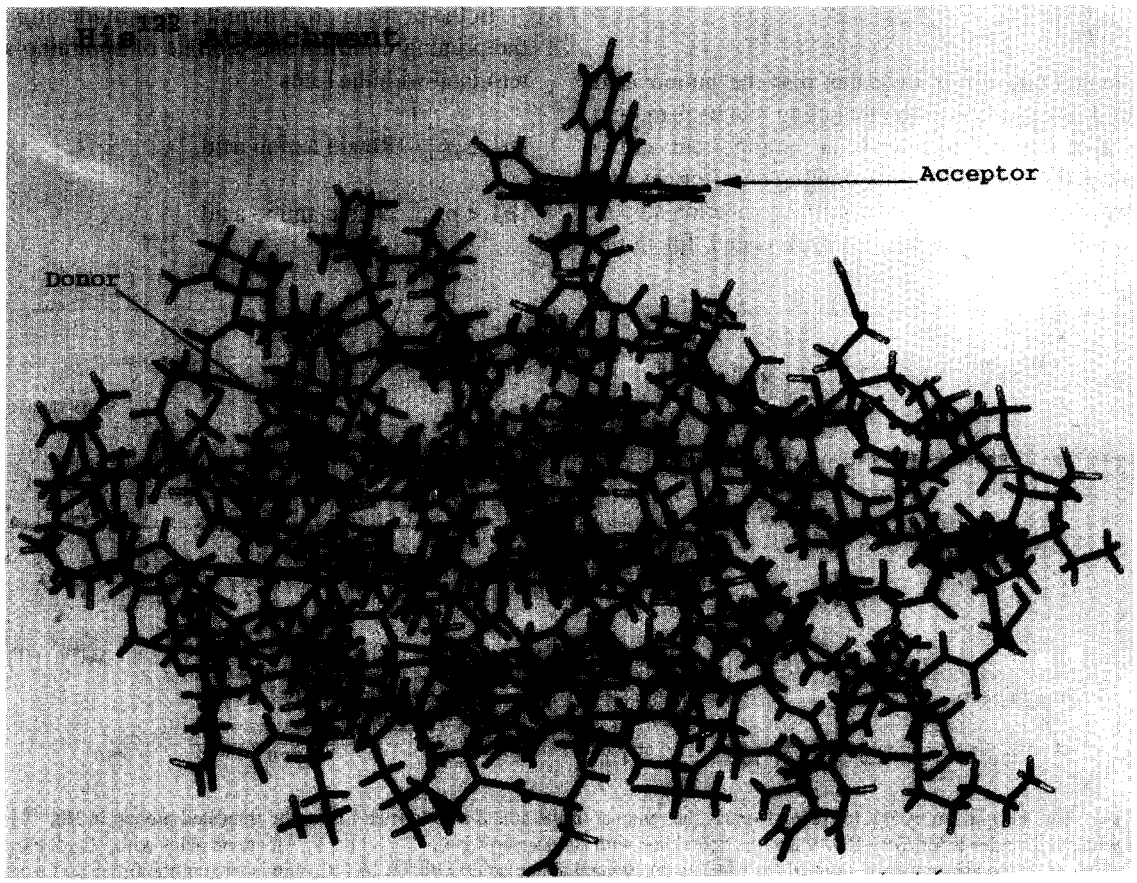

Fig. 3. Protein molecule in which the ruthenium complex is ligated to His^{122}. The total number of atoms in the protein is 1971. The arrows point to the donor and acceptor metal centers.

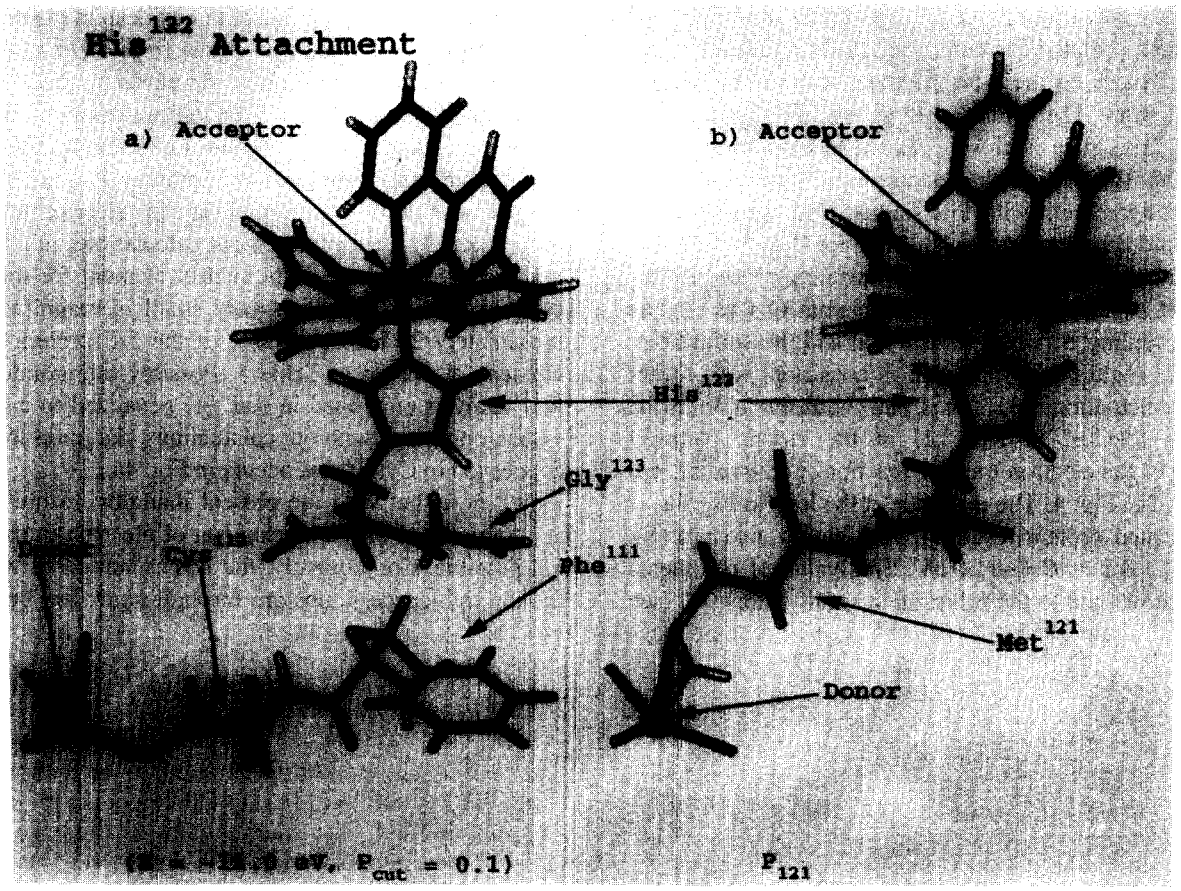

Fig. 4. Truncated proteins for His^{122}. (a) Left: the molecule results from amino acid pruning governed by Eqs. (6) and (7) at $E = -11.0$ eV with $P_{cut} = 0.1$. Its matrix element differs from that of the full protein by an less than 3%. (b) Right: the amino acids that are important in the pathway P_{121} are shown (see Table 1). The molecules in (a) and (b) are made up of 112 and 90 atoms, respectively.

ids that are used in the more computationally expensive second step of the procedure. In the second step, a pruning procedure, designed to localize the tunneling pathway further, determines more precisely whether individual amino acids within the tubular region should be retained or deleted, based only on their direct effect on the matrix element. This procedure does not require, in principle, any initial guess for the tunneling pathway; it finds the path on its own. The results with the pruning procedure indicate a further reduction of the number of atoms by a factor of 3 over the tube truncation method, resulting in approximately 10% of the original protein atoms being present in the final pruned molecule.

This reduction of the problem by roughly a factor of 10 reduces a large protein to a system of only a few hundred atoms. It is striking how small a number of amino acids is needed to reproduce the matrix element of the full protein to a very good accuracy, as shown in Table 2. The number of atoms composing the pruned protein is sufficiently small that much more sophisticated quantum calculations potentially including many-electron effects [31,35–37] could be performed on the remaining system. This avenue is worthy of further examination. The use of such techniques, perhaps even at the Hartee–Fock level, can test the accuracy of the currently used more approximate methods, such as the extended Hückel theory of the present study.

Implications of the present results regarding the pathways for electron transfer are considered next, focusing first on the electron transfer reaction in the molecule in which the ruthenium complex is attached to His^{122}. The full molecule is shown in Fig. 3. The pruning procedure results in a remarkable reduction of the tunneling region of the protein as illustrated by Fig. 4. A tunneling energy of −11.0 eV is used as an example for the purpose of the discussion.

In Fig. 2a, the labels indicate that the most important amino acids are present in the β-strands associated primarily with P_{121} and P_{112}. At −11.0 eV (dotted line with solid squares), the two amino acids, Phe^{111} and Cys^{112} along the Cys^{112} β-strand are each seen to have a large effect on the effective coupling matrix element. Their effect may be contrasted with those amino acids in the β-strand originating with Met^{121}. At −11.0 eV only one amino acid along this β-strand is seen to have a large effect on the matrix element, and it is His^{122}, which is directly ligated to the ruthenium complex. From these results it is deduced that the tunneling pathway involves a

jump from the copper atom to Cys^{112}, followed by travel along the associated β-strand, and finally a through-space jump to His^{122} on the other β-strand and so into the ruthenium atom. The pruned path so identified is listed in Table 2. This path is closely related to the pathway P_{112} defined in Table 1 which was constructed on an intuitive basis. In agreement with this finding, Fig. 1a displays a more rapid convergence for P_{112} than for P_{121}.

The tunneling path in the above pruning procedure began with a short jump to the sulfur atom of Cys^{112}. The strong coupling between the Cu^+ ion and the sulfur atom of Cys^{112} required for this path is consistent with recent CNDO calculations [31]. This path is different from the pathway selected in Ref. [1] in the atomic pathway analysis of the electron transfer reaction. These results are illustrated in Fig. 4. The molecule at the left (in Fig. 4a) is the truncated protein resulting from amino acid pruning with $p_{cut} = 0.1$ at $E = -11.0$ eV. This value of p_{cut} yields results converged to the full protein result within 3% (see Table 2). This accuracy was obtained despite the dramatic reduction in the number of atoms, from 1971 to 112, as compared to the full protein displayed in Fig. 3. The amino acid Met^{121} is not even present in the truncated protein displayed in Fig. 4. The molecule at right (Fig. 4b), by contrast, represents the pathway P_{121} whose curvilinear axis is the atomic pathway for electron transport used in Ref. [1]. However, in our calculations, at a tunneling energy of −11.0 eV, the matrix element for the truncated protein in Fig. 4b is very small in magnitude, approximately 20% of the value of the full protein matrix element, as shown in Table 1. Possibly the atomic pathway analysis employed in Ref. [1] penalized to a greater extent, compared with our calculations, the jump from the copper ion to Cys^{112} that occurs in Fig. 4a.

We note that with the present methods, resolution on an atomic level of the mechanism of electron transport is not possible. Consequently, the determination of whether the through-space jumps are through hydrogen bonds or

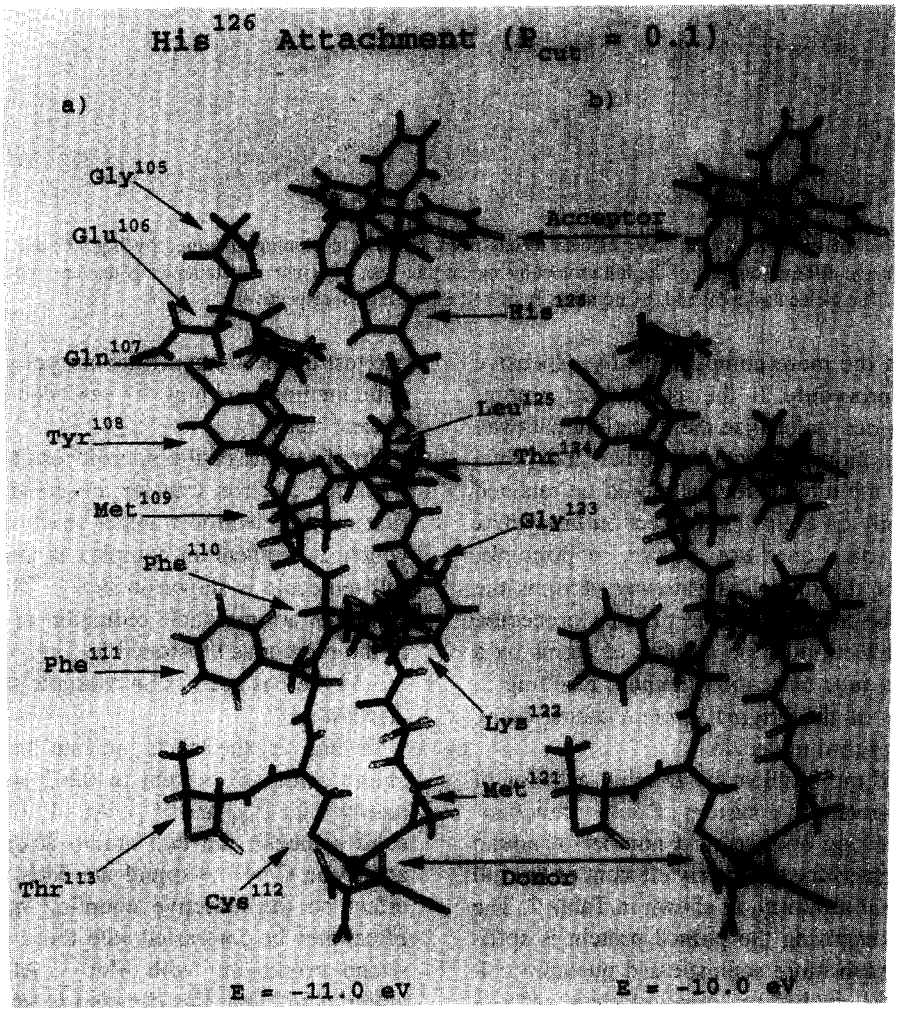

Fig. 5. Truncated proteins for the molecule in which the ruthenium complex is ligated to His^{126}. In (a) the pruning is at $E = -11.0$ eV, using Eqs. (6) and (7) with $P_{cut} = 0.1$. In (b), this pruning is at −10.0 eV. The pruned molecules contain 303 and 260 atoms, respectively. The right hand β-strand in each of the panels corresponds to P_{121} and the left hand one to P_{112}.

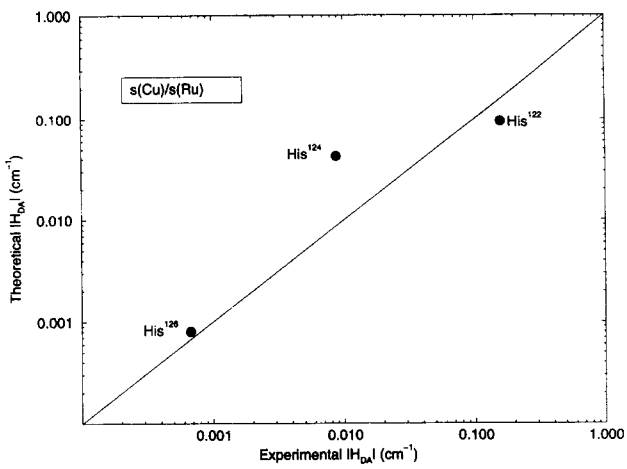

Fig. 6. Comparison of the experimental results [1] and the calculations of the electronic coupling matrix element for tunneling energy −10.0 eV. Effective s-orbitals were used on the donor and acceptor sites.

not cannot be assessed at this point. In the future, we plan to develop techniques that address such questions with quantum calculations on the small pruned molecules.

Analysis of the electron transfer at −11.0 eV in the system with the large donor-acceptor separation distance, the His^{126} molecule, yields strikingly different results. Inspection of Fig. 2b at this energy reveals that both Met^{121} and Cys^{112} β-strands now have a large effect on the coupling matrix element. Moreover, all amino acids contributing to the β-strands are now important, indicating that both P_{112} and P_{121} pathways are important simultaneously at this energy. This conclusion is reinforced in Fig. 1b, in which the convergence with pathway $P_{121+112}$ is more rapid at $R = 4$ Å than either P_{112} or P_{121} alone. At $R = 2$ Å, the tube surrounding P_{121} has a matrix element that is nearly a factor of two larger than the full protein result. It is clear from the figure that the contribution from P_{112}, which is opposite in sign, destructively interferes with this pathway, and so results in a smaller overall matrix element.

The results given in Fig. 2b for the tunneling energy −10.0 eV are considered next and are compared to those for $E = -11.0$ eV. As noted above, at −11.0 eV, the tunneling mechanism in the molecule with the ruthenium complex attached to His^{126} appears to involve quantum interference effects between two pathways running along two parallel β-strands. At −10.0 eV, however, the picture is quite different. Inspection of Fig. 2b reveals that only the β-strand associated with Cys^{112}, makes a major contribution. This conclusion is supported by Fig. 5, in which the results for amino acid pruning with $p_{cut} = 0.1$ for tunneling energies −10.0 eV (Fig. 5b) and −11.0 eV (Fig. 5a) are presented. (This value of p_{cut} gives convergence of H_{DA} to the full protein value within 4% at both energies, as seen in Table 2). Inspection of the molecules in Fig. 5 shows that while for −11.0 eV both β-strands are involved in tunneling, at −10.0 eV, the β-strand associated with Met^{121} is fragmented, and so the tunneling mechanism utilizes primarily an amino acid path that is close to the pathway denoted P_{112}.

It is also interesting to note that the His^{126} that ligates the ruthenium atom is absent in the pruned molecule shown in Fig. 5b. The pathway at −10.0 eV leading to the ruthenium ion now involves instead a through-space jump between Gln^{107} of the Cys^{112} β-strand and one of the bipyridines that ligates the ruthenium ion. The closest approach distance between the bipyridine and Gln^{107} is 2.0 Å. These points illustrate the fact that as the tunneling energy is varied, the principal pathway for the transfer may change, even to the extent of causing different parts of the protein to become significant.

There are a number of uncertainties at present both in the calculated and experimental absolute values of the tunneling matrix element. (For example, it would be useful to measure the dependence of the electron transfer rate constant on the thermodynamic driving force, $-\Delta G°$.) Nevertheless, it is interesting to observe that within the present model, it is possible to describe semi-quantitatively the overall dependence of the coupling matrix element that is observed in the experiment [1]. No adjustable parameters for individual molecules, apart from the uncertainty in the tunneling energy, which can be resolved in the future, have been used. Results are shown in Fig. 6 of the calculations with effective s-orbitals on donor and acceptor, and with a tunneling energy of −10.0 eV. Given the approximations made and the uncertainties in the model, the overall agreement is encouraging, and perhaps indicates that the main physics of the process is captured correctly. However, it should be stressed that more reliable calculations should include a detailed treatment of the electronic structure of the copper center and its environment [31–33]. A further discussion of the distance dependence and the comparison with experimental data will be given in a detailed forthcoming paper.

In summary, it appears that the pruning procedures interpreted as above can lead to valuable information regarding the pathways of electron transfer. It is worthy of future theoretical consideration to examine whether results markedly different from the present ones with β-strands would be found in α-helices and other non-random biological structures. Also, it would be interesting to apply our pruning procedure to other ruthenium-modified mutants of azurin, for which experimental results have been obtained recently by Gray and co-workers [38,39].

Acknowledgements

It is a pleasure to dedicate this article to our colleague and friend, Harry Gray, whose work has provided so much inspiration and stimulation to us. We also thank him, Jay Winkler and Ralf Langen for many useful

discussions. Part of the research conducted at UC Davis was supported by the grant of the Petroleum Research Fund and the start-up fund provided by the Department of Chemistry to AAS. One of us (RAM) was supported by grants from the National Science Foundation, the Office of Naval Research, and NEDO. Calculations presented in this paper were performed on the CRAY supercomputer provided by Jet Propulsion Laboratory, Pasadena, CA.

References

[1] R. Langen, I. Chang, J.P. Germanas, J.H. Richards, J.R. Winkler and H.B. Gray, *Science, 268* (1995) 1733.
[2] C.C. Moser, J.M. Keske, K. Warncke, R.S. Farid and P.L. Dutton, *Nature, 355* (1992) 796.
[3] S. Risser, D.N. Beratan and T. Mead, *J. Am. Chem. Soc., 115* (1993) 2508.
[4] J.N. Onuchic and D.N. Beratan, *J. Chem. Phys., 92* (1990) 722.
[5] J.N. Onuchic, P.C. Andrade and D.N. Beratan, *J. Chem. Phys., 95* (1991) 1131.
[6] J.N. Onuchic, D.N. Beratan, J.R. Winkler and H.B. Gray, *Annu. Rev. Biophys. Biomol. Struct., 21* (1992) 349.
[7] S.S. Skourtis, J.J. Regan and J.N. Onuchic, *J. Phys. Chem., 98* (1994) 3379.
[8] J.N. Betts, D.N. Beratan and J.N. Onuchic, *J. Am. Chem. Soc., 114* (1992) 4043.
[9] J.N. Onuchic, D.N. Beratan, J.R. Winkler and H.B. Gray, *Science, 258* (1992) 1740.
[10] D.N. Beratan, J.N. Betts and J.N. Onuchic, *Science, 252* (1991) 1285.
[11] M. Gruschus and A. Kuki, *J. Phys. Chem., 97* (1993) 5581.
[12] A. Kuki and P.G. Wolynes, *Science, 236* (1987) 1647.
[13] J.W. Evenson and M. Karplus, *Science, 262* (1993) 1247.
[14] J.W. Evenson and M. Karplus, *J. Chem. Phys., 96* (1992) 5272.
[15] A. Okada, T. Kakitani and J. Inoue, *J. Phys. Chem., 99* (1995) 2946.
[16] A.A. Stuchebrukhov and R.A. Marcus, *J. Phys. Chem., 99* (1995) 7581.
[17] A.A. Stuchebrukhov, *Chem. Phys. Lett., 225* (1994) 55.
[18] P. Siddarth and R.A. Marcus, *J. Phys. Chem., 94* (1993) 8430.
[19] P. Siddarth and R.A. Marcus, *J. Phys. Chem., 94* (1993) 2985.
[20] P. Siddarth and R.A. Marcus, *J. Phys. Chem., 96* (1992) 3213.
[21] P. Siddarth and R.A. Marcus, *J. Phys. Chem., 97* (1993) 2400.
[22] P. Siddarth and R.A. Marcus, *J. Phys. Chem., 97* (1993) 13078.
[23] A.W. Axup, M. Albin, S.L. Mayo, R.J. Crutchley and H.B. Gray, *J. Am. Chem. Soc., 110* (1988) 435.
[24] D.S. Wuttke, M.J. Bjerrum, I. Chang, J.R. Winkler and H.B. Gray, *Biochim. Biophys. Acta, 1101* (1991) 168.
[25] D.R. Casimiro, J.H. Richards, J.R. Winkler and H.B. Gray, *J. Phys. Chem., 97* (1993) 13073.
[26] R. Hoffmann, *J. Chem. Phys., 39* (1963) 1397.
[27] (a) R. Hoffmann, *J. Chem. Phys., 40* (1964) 2245; (b) R. Hoffmann, *J. Chem. Phys., 40* (1964) 2474; (c) R. Hoffmann, *J. Chem. Phys., 40* (1964) 2480.
[28] The Biograph data files containing atomic coordinates of the modified azurin molecules were kindly provided by Ralf Langen, a member of Harry Gray's research group.
[29] We point out that the geometries of the ligands around the copper center in the files described in Ref. [28], which represent energy minimized structures of the mutated molecules, are different than that found in the X-ray study of the unmodified azurin molecule in Ref. [30]. The distances between the copper atom and the sulfur atom of Met^{121}, the sulfur atom of Cys^{112}, the nitrogen atom of His^{46}, the nitrogen atom of His^{117} and the carbonyl oxygen of Gly^{45} are 3.18, 2.27, 2.00, 2.11 and 2.84 Å, respectively, in the modified molecule with the complex attached to His^{122}. The respective distances in the modified molecule with the complex attached to His^{126} are 3.14, 2.45, 2.00, 2.12 and 2.82 Å. Finally, the respective distances in the X-ray structure of the unmodified azurin molecule are 3.15, 2.25, 2.11, 2.03 and 2.97 Å. Note, the X-ray structure was taken at pH 5.5 with the copper ion in its oxidized form.
[30] H. Nar, A. Messerschmidt, R. Huber, M. van de Kamp and G.W. Canters, *J. Mol. Biol., 221* (1991) 765.
[31] S. Larsson, A. Broo and L. Sjölin, *J. Phys. Chem., 99* (1995) 4860.
[32] A.A. Gewirth and E.I. Solomon, *J. Am. Chem. Soc., 110* (1988) 3811.
[33] J.A. Guckert, M.D. Lowery and E.I. Solomon, *J. Am. Chem. Soc., 117* (1995) 2817.
[34] We note that our p_i is similar in spirit to the parameter I_i defined by Okada et al. [15] in their Eq. (6). Presumably for amino acids for which p_i is large, I_i will be large and when p_i is small, so too will be I_i.
[35] M.D. Newton, *Int. J. Quantum Chem. Phys., Quantum Chem. Symp., 14* (1980) 363.
[36] M.D. Newton, *Chem. Rev., 91* (1991) 767.
[37] C.X. Liang and M.D. Newton, *J. Phys. Chem., 97* (1993) 3199.
[38] R. Langen, J.L. Colon, D.R. Casimiro, T.B. Karpishin, J.R. Winkler and H.B. Gray, *J. Biol. Inorg. Chem.*, (1996) in press.
[39] R. Langen, *Ph.D. Thesis*, California Institute of Technology, 1995.