Computational Studies of Electron Transfer in Multi-Heme Proteins

Xiuyun Jiang
Supervisor: Prof. Dr. Jochen Blumberger

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Physics and Astronomy
University College London

December 11, 2019
I, Xiuyun Jiang

Supervisor: Prof. Dr. Jochen Blumberger, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.
Abstract

Multi-heme proteins are fascinating biomolecules that bind several redox-active heme cofactors in close distance to shuttle electrons across the bacterial membrane. Yet, the kinetics and time scales on which these electron transfer (ET) events occur is not well known and difficult to probe experimentally. The central aim of this thesis is to compute and to quantify heme-heme ET rate constants and electron flux through solvated multi-heme proteins. To this end, density functional theory and molecular dynamics simulation are deployed to compute heme-heme ET parameters in the framework of (non-adiabatic) Marcus theory, the central theory underlying such ET events. Three ubiquitous multi-heme proteins have been studied, which bind 4 and 10 heme cofactors. Our calculations revealed that electron transfer through these proteins is strongly enhanced by cysteine side chains that are inserted in the space between heme groups. We believe this to be a general design principle in this family of proteins for acceleration of ET steps that would otherwise be too slow for biological respiration. Our computational protocol has been verified via comparing our predicted time scale of heme-heme ET with the corresponding ET rate constant measured from pump-probe spectroscopy. The maximum, protein-limited electron flux is $\approx 10^5 - 10^6 \text{ s}^{-1}$. Such efficiency in long-range electron transfer indicates that multi-heme proteins are promising candidates for biological nano-electronic devices.
Impact Statement

Certain bacteria survive in anoxic environments by switching from aerobic to anaerobic respiration: in place of O$_2$ they reduce diverse substrates outside the cell. Multi-heme proteins spanning the outer membrane have been identified as the essential building blocks for this process. Yet, detailed underlying mechanism is still a matter of debate.

Here with molecular simulation and quantum-mechanical calculations of three multi-heme proteins, my PhD work provides molecular-level insight into the electron transfer mechanism in this family of proteins. The research work reported herein have partially validated the use of Marcus theory, a commonly-used theory for thermally-mediated electron transfer, in the estimation of electron transfer rate constants in solvated multi-heme proteins. Our study reveals that the electron hopping rate constants in the three multi-heme proteins studied herein could be strongly enhanced by the cysteine side chains of the heme cofactors. We believe this to be a general design principle in multi-heme proteins for acceleration of electron transfer steps that would otherwise be too slow for respiration. Our study uncovers a natural design principle of significance to an entire class of proteins involved in biological electron transfer and could further impact the design of bio-inspired materials.
Acknowledgements

First of all, I owe the greatest thank to my supervisor, Prof. Dr. Jochen Blumberger, who takes full responsibility of supervising including guidance, discussion, inspection, criticizing and support. Apart from my supervisor, the two postdocs working on the same project, Dr. Zdenek Futera and Dr. Antoine Carof, also gave me a lot of constructive suggestions and technical support. They are both patient, friendly, and helpful. Dr. Marian Breuer, the previous PhD student working on this project, is always responsive to my questions through emails in time in detail. Dr. Guido Falk von Rudorff, a fellow student, has given me a lot of suggestions, scientifically, technologically and personally. I thank Prof. Dorothy Duffy and Dr. Andela Saric for the insightful discussions at my upgrade viva; I thank Prof. Alexandra Olaya-Castro and Dr. Edina Rosta for the constructive discussions and comments at my PhD viva.

Next, I would like to thank my main experimental collaborators from University of East Anglia and Central Laser Facility at Rutherford Appleton Laboratory (UK), led by Prof. Julea Butt, who is knowledgeable, open for discussion and friendly. I also had a lot of direct discussions with Dr. Jessica H. van Wonderen, Dr. Christopher R. Hall, and Dr. Katrin Adamczyk who carried out the experimental work. The collaboration was interactive and fruitful.

I am also grateful for our experimental collaborators from Weizmann Institute of Science (Israel), Prof. David Cahen and Dr. Kavita Garg for the insightful discussions and the collaboration.

I would like to thank Dr. R. Thomas Ullmann (Max Planck Institute for Biophysical Chemistry, Gottingen, Germany) and Prof. Emil Alexov (Clemson University, US) for the discussion and technical support on the use of Poisson-Boltzmann solvers.

I am grateful for Jan Elsner for proofreading the whole thesis and providing valuable feedback. I am also grateful for Dr. Wei Fang and Dr. Antoine Carof for proofreading some chapters and their valuable comments on this thesis.

I thank all my current and past labmates for collaboration, discussion and the time spent
together: Dr. Marian Breuer, Dr. Fruzsina Gajdos, Dr. Jacob Spencer, Dr. Guido Falk von Rudorff, Dr. Hui Yang, Samuele Giannini, Matthew Ellis, Chris Ahart, Dr. Antoine Carof, Dr. Souyma Ghosh, Dr. Zdenek Futera, Dr. Ehesan Ali, Dr. Carlo Bortolotti, Dr. Orestis Ziogos, Dr. Xiaojing Wu, Dr. David Dell’Angelo, Jan Elsner, Patrick Gutlein, Peter Cooke, Daniel Holub, Bastian Burger, Oliver Gittus, Kevin Lively, Laura Scalfi, Rasmus Jakobsen, Zak Zheng.

I thank Thomas Young Centre (TYC) for both academic environment and social events. I thank all people who regularly come to the TYC journal club and suggest papers: Matt Ellis, Chris Ahart, David Bowler, Jan Gerit Brandenburg, Jack Pouton, Amy Lai, Samuele Giannini, Xiaojing Wu, Patrick Rowe, Yasmine S. Al-Hamdani, Wei Fang, Philipp Pedevilla, Kane Shenton, Laura Hargreaves, David Schmidt, Guo Rui, Kevin Lively. I learned a lot from the journal discussions and felt strong sense of belonging. I thank Matt Ellis and Chris Ahart who take over the journal club hosting and Yasmine S. Al-Hamdani who pass this role to me. I hope the journal club, as well as TYC will be there forever.

I thank my parents for their understanding of my choice, letting their only daughter to pursue a PhD abroad. I am grateful for their support and push as always.

Finally, I would like to acknowledge my funding entity and computational resources. This PhD work was made possible by a joint PhD studentship funded by the Chinese Scholarship Council and University College London. Via the membership of the UK’s HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202, EP/R029431), I have access to the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk). I am grateful to the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/P020194/1). I acknowledge the use of the UCL Grace High Performance Computing Facility (Grace@UCL), and associated support services, in the completion of my work. I am grateful to London Centre for Nanotechnology for the local machine and Fabrizio Sidoli who provides the technical support.
Contents

1 Introduction 23
  1.1 Thesis statement .................................................................................. 29

2 Theoretical Background and Computational Approaches 31
  2.1 Molecular modelling approaches for biomolecular systems ................. 31
    2.1.1 Molecular dynamics (MD) simulation ........................................... 31
    2.1.2 Density functional theory (DFT) ............................................... 35
    2.1.3 Quantum mechanics/molecular mechanics (QM/MM) ................. 38
  2.2 Marcus theory in non-adiabatic limit .................................................. 40
    2.2.1 Assumptions and validations ...................................................... 41
    2.2.2 Heterogeneous electron transfer (ET) ......................................... 42
  2.3 Computation of electron transfer (ET) parameters .............................. 43
    2.3.1 Electronic coupling $H_{ab}$ ......................................................... 43
    2.3.2 Reorganisation free energy $\lambda$ .............................................. 45
    2.3.3 Driving force $\Delta A$ .............................................................. 47
  2.4 Chemical master equation for macroscopic properties ....................... 49
    2.4.1 Steady-state flux ....................................................................... 49
    2.4.2 Relaxation dynamics .................................................................. 50
  2.5 Typical workflow ................................................................................. 52

3 Electron Transfer in Wild-Type Proteins STC, MtrC, MtrF 53
  3.1 Electronic couplings ............................................................................. 53
    3.1.1 Side chain enhancement of electronic couplings ......................... 53
    3.1.2 Rate enhancement estimation with pathway-tunnelling model ........ 64
    3.1.3 Generalized distance metric for electronic couplings ................. 66
3.2 Calculation of ET driving forces and reorganisation free energies ............. 69
3.3 $H_{db}, \lambda, \Delta A$: which is the most sensitive ET parameter? .................. 70
3.4 Maximum, protein-limited steady-state electron flux ............................. 72
  3.4.1 Influence of interactions between redox centres on ET parameters ....... 73
  3.4.2 Biological implications of ET in deca-heme proteins ..................... 76
3.5 $I - V$ curve modelling ................................................ 79
  3.5.1 $I - V$ from scanning tunnelling microscopy (STM) ...................... 79
  3.5.2 $I - V$ from suspended nanowire ..................................... 82
3.6 Conclusion ........................................................................... 84

4 Photo-Initiated Electron Transfer in Ru(bpy)$_3$-Labelled STC .......................... 87
  4.1 Experimental background .................................................... 87
  4.2 Docking and molecular dynamics simulation ................................. 90
    4.2.1 Structural model for Ru$_{77}$:STC .................................. 90
    4.2.2 Force field parameters ................................................. 92
    4.2.3 MD simulation of Ru$_{77}$:STC and Ru$_{77}$+$^{\cdot}$:STC$^{\cdot}$ (CS$^{\cdot}$) .... 92
  4.3 Estimation of ET parameters for charge separation and charge recombination .... 95
    4.3.1 Reorganisation free energies for CS and CR .......................... 95
    4.3.2 Electronic coupling matrix elements for CS and CR ................. 97
  4.4 Influence of the Ru-dye on heme reduction potentials ..................... 100
    4.4.1 Heme reduction potentials for STC .................................. 101
    4.4.2 Heme reduction potentials for Ru$_{77}$:STC .......................... 102
  4.5 Extraction of heme-heme ET rate constant from transient absorbance data ...... 102
    4.5.1 CS and CR without heme-heme ET .................................. 102
    4.5.2 CS and CR with heme-heme ET .................................... 104
    4.5.3 Discussion of model fits .............................................. 106
  4.6 Summary ........................................................................... 111

5 Effect of Electronic Polarizability on Reorganisation Free Energy .................. 113
  5.1 Electronic polarizability claimed to cause ergodicity-breaking effect ........ 113
  5.2 Results and discussion ..................................................... 117
    5.2.1 Reorganisation free energies from vertical energy gaps ............. 117
    5.2.2 Force fields: CHARMM27 vs AMBER03 ........................... 117
List of Figures

1.1 Crystal structures of deca-heme cytochromes MtrF and MtrC. ..................... 25
1.2 Structural formula of a c-type heme cofactor. ........................................... 26
1.3 Crystal structures of deca-heme cytochrome MtrC and tetra-heme cytochrome STC and their cellular locations. .................................................. 28

2.1 Illustration of a typical protein force field. ................................................. 33
2.2 Free energy curves to illustrate Marcus Theory. .......................................... 40
2.3 Two-dimensional view of the three-dimensional Debye-Hückel model. ........... 48
2.4 Diagram of a typical workflow for electron transfer studies in multi-heme proteins. 52

3.1 Frontier orbitals of a typical snapshot in DFT calculations. ............................ 56
3.2 Schematic illustration of the computational models for heme-heme electronic coupling calculations of protein STC. .................................................. 59
3.3 Position of substituents $R_i$, $i = 1, \ldots, 8$ in the heme models defined in Table 3.1. 60
3.4 $|H_{ab}|$ values for heme pair 1-2, 2-3, 3-4 of protein STC. ........................... 61
3.5 Crystal structure of proteins STC and MtrC with the heme cofactors and the Fe-heme frontier orbital pairs mediating the electron transfer highlighted. .......... 62
3.6 $|H_{ab}|$ convergence with respect to the QM model size for heme pair 1-6 with minimized MtrF crystal structure. .................................................. 64
3.7 Comparison of wave functions of a frontier orbital of an isolated heme with cysteine side chain and its artificial mutants with serine and methylated alanine side chains. .................................................. 66
3.8 The distance dependence of heme-heme electronic coupling matrix elements, using the data of MtrC, MtrF and STC. ........................................... 67
3.9 MD averaged distance separation between each heme pair in MtrC and MtrF. .... 68
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.10</td>
<td>Diagram of the refinement of the heme-heme ET rates in steady-state flux calculations.</td>
</tr>
<tr>
<td>3.11</td>
<td>Kinetics of trifurcated electron flow in solvated MtrC and MtrF.</td>
</tr>
<tr>
<td>3.12</td>
<td>Experimental and computed $I - V$ curves for MtrF and MtrC in STM experiment.</td>
</tr>
<tr>
<td>3.13</td>
<td>$I - V$ curves of STC, MtrF and Azurin measured by suspended nanowire method and modelled by coherent tunnelling model and incoherent hopping model.</td>
</tr>
<tr>
<td>4.1</td>
<td>Illustration of Ru(II)tris(bipyridine)-labelled STC protein.</td>
</tr>
<tr>
<td>4.2</td>
<td>Photochemistry reaction scheme of Ru-dye labelled STC and the evolution and interconversion of each reaction component measured from the transient absorbance (TA) spectroscopy.</td>
</tr>
<tr>
<td>4.3</td>
<td>Illustration of Ru(bpy)$_2$(4-methylbipyridene-4'-CH$<em>2$-) docking in Ru$</em>{77}$:STC.</td>
</tr>
<tr>
<td>4.4</td>
<td>Molecular models of the reactive conformers of Ru$_{77}$:STC.</td>
</tr>
<tr>
<td>4.5</td>
<td>Conformation interconversion of the Ru-dye in the four MD trajectories of Ru$_{77}$:STC.</td>
</tr>
<tr>
<td>4.6</td>
<td>Non-ergodicity correction for outer-sphere reorganisation free energies.</td>
</tr>
<tr>
<td>4.7</td>
<td>Electronic couplings of CS and CR in Ru$_{77}$:STC.</td>
</tr>
<tr>
<td>4.8</td>
<td>Reduction potentials of STC and Ru$_{77}$:STC computed from linearized Poisson-Boltzmann equations.</td>
</tr>
<tr>
<td>4.9</td>
<td>Experimental TA data and the best-fit results for Ru$<em>{77}$:STC and Ru$</em>{23}$:STC.</td>
</tr>
<tr>
<td>4.10</td>
<td>Summary of the fit to the experimental TA data for Ru$<em>{77}$:STC, Ru$</em>{23}$:STC and Ru$_{10}$:STC using multiple-conformer models without and with inclusion of heme-heme electron transfer.</td>
</tr>
<tr>
<td>5.1</td>
<td>Illustration of the ergodicity-breaking effect in redox-active proteins.</td>
</tr>
<tr>
<td>5.2</td>
<td>Illustration of our model set-up in the calculation of reorganisation free energy for oxidation of solvated cytochrome c at MM and QM/MM levels.</td>
</tr>
<tr>
<td>5.3</td>
<td>Fluctuations and convergence of the vertical energy gap $\Delta E$ for oxidation of solvated cytochrome c.</td>
</tr>
<tr>
<td>5.4</td>
<td>Histograms of vertical energy gap $\Delta E$ for oxidation of solvated cytochrome c for different computational models and force fields.</td>
</tr>
<tr>
<td>5.5</td>
<td>Frontier orbitals of a test snapshot in QM/MM calculations.</td>
</tr>
</tbody>
</table>
5.6 Accumulating average of reorganisation free energies for oxidation of solvated cytochrome c for different computational models and force fields. . . . . . . . . 124
5.7 Gap energy auto-correlation function and its time integral as obtained for oxidation of cytochrome c in reduced and oxidized states. . . . . . . . . . . . . . . . 126
# List of Tables

3.1 Computational models for electronic coupling calculations with the test heme pairs \( i:j \) in STC and MtrF. .............................................. 58

3.2 Averaged heme-heme electronic coupling matrix element \( \langle |H_{ij}|^2 \rangle^{1/2} \) for STC. ... 60

3.3 Distances between heme pairs in STC. ........................................ 63

3.4 Heme-to-heme tunnelling enhancement factors in MtrC, MtrF and STC from POD/PBE50 and pathway calculations. ............................... 65

3.5 Summary of the mean, maximum and minimum of the computed heme-heme ET parameters in proteins STC, MtrC, MtrF. ........................................... 72

3.6 Sensitivity of the heme-heme ET rate constants with respect to the change of each individual Marcus ET parameter. ............................... 72

3.7 Computed protein-limited steady-state electron fluxes in STC. ................. 76

3.8 Maximum, protein-limited electron flux through MtrC and MtrF in different flow directions. ................................................................. 77

3.9 Summary of the fitting results for the \( I − V \) curves of proteins STC, MtrF and Azurin with coherent tunnelling mechanism. ................................. 83

4.1 Structural characterization of conformers 1-4 in state A (Ru:STC) and state B (Ru\(^+\)STC\(^−\)) for Ru\(_{77}\):STC. .................................................. 93

4.2 Summary of ET parameters and rate constants in CS and CR for the different conformers 1-4. ................................................................. 98

4.3 Summary of fitting results for the experimental TA data of Ru\(_{77}\):STC. ........ 107

4.4 Summary of fitting results for the experimental TA data of Ru\(_{23}\):STC. ........ 108

4.5 Summary of fitting results for the experimental TA data of Ru\(_{10}\):STC. ........ 109

4.6 Electron transfer rate constants for the Ru\(_{77}\):STC photocycle. ................ 110
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Convergence of electronic polarizability of the electronic ground state of reduced and oxidized heme c cofactor with number of excited states.</td>
<td>127</td>
</tr>
<tr>
<td>5.2</td>
<td>Convergence of reorganisation free energies with number of excited states used in QM(PMM)/MM(np) approach.</td>
<td>127</td>
</tr>
<tr>
<td>5.3</td>
<td>Computed reorganisation free energies for oxidation of solvated horse-heart cytochrome c.</td>
<td>131</td>
</tr>
<tr>
<td>C.1</td>
<td>Summary of ET parameters and rate constants for heme-to-heme ET in STC.</td>
<td>141</td>
</tr>
<tr>
<td>C.2</td>
<td>Summary of computed ET parameters and rate constants for heme-to-heme ET in MtrC and MtrF.</td>
<td>142</td>
</tr>
</tbody>
</table>
Publications

Parts of this dissertation have been published as follows:


Chapter 1

Introduction

“Life is nothing but an electron looking for a place to rest”, as Nobel Prize winner Albert Szent-Gyorgyi puts it, electron transfer (ET) processes are ubiquitous in nature and vital to the basics of life, such as photosynthesis [1], respiration [2] and corrosion [3]. Chemically, an electron transfer reaction (also termed reduction-oxidation reaction), is defined as a type of chemical reaction in which the oxidation states of atoms are changed.

Electron transfer is widely studied in various biological systems, such as proteins [4, 5], DNA [6, 7], ATP [8, 9], photosynthesis I and II [10, 11], iron-sulfur clusters [12, 13], etc. Not only do these studies unveil fundamental aspects of life, but they also influence the directions of research and engineering fields such as biomimetic catalysis design and bio-inspired materials.

Our lab has been studying a type of proteins called multi-heme proteins for several years. The proteins are isolated from Gram-negative bacteria Shewanella oneidensis [14, 15], which is a marine species initially found in Lake Oneida, NY, USA. Shewanella can respire and survive anaerobically for up to a month by extending conductive pili outside the bacteria cell to a number of insoluble substrates, such as rocks or other sediments in the ocean [14]. The secret lies in the heme cofactors – densely packed, redox-active organic alloys facilitating electrons transport from inside the bacterial cell to extracellular space for distances up to 100 Å. Apart from the long-range capability, this extracellular electron transfer is surprisingly efficient. While the electron exchange between ions Fe$^{2+}$ and Fe$^{3+}$ happens on the time scale of seconds in aqueous solutions [16], the time scale reduces to microseconds or even less when they are centered in heme cofactors embedded inside these proteins [17]. Studying this kind of long-range, fast ET in these bacterial nanowires is important in understanding the geochemical cycling of various metals. It is also helpful in the design of promising bio-materials such as biological catalysts [18, 19], biological nano-electronic devices [20] or microbial fuel cells [21].
The structural, biochemical, and electrochemical properties of multi-heme proteins from *Shewanella* have been studied with different experimental techniques. On the cellular level, El-Naggar *et al.* have used live fluorescence and electron cryotomography to reveal that those conductive pili of bacteria *Shewanella* are in fact vesicular extension of outer-membrane and periplasmic extracellular electron transport components [22, 23]. By constructing field-effect transistors based on individual bacteria nanowires, Leung and coworkers have measured high field-effect mobility of these nanowires comparable to devices based on synthetic organic semiconductors [24]. Proteoliposome experiments [25–27] carried out by Richardson *et al.* have identified the key component mediating the extracellular electron transport. A “porin-cytochrome” model of the transmembrane electron transfer protein complex has been proposed in which the deca-heme protein MtrA, facing the periplasm, and the deca-heme MtrC, located outside the bacteria cell surface, form close contact within the porin MtrB, so that electrons might be transferred across the outer membrane in either direction.

In recent years, there has been an increasing interest in the mechanistic study of the electron transfer and electron transport processes in these redox proteins on the experimental side. Current-voltage characteristic curves have been measured on the protein level with various types of protein junctions at different conditions [28–32]. More recently, efforts have been made to study the electron transfer mechanism in fully solvated multi-heme proteins with ultra-fast pump-probe spectroscopy [33]. However, the detailed electron transfer and transport mechanism in these multi-heme proteins is still not fully understood.

The (anaerobic respiratory) multi-heme proteins in *Shewanella* are mainly found on the inner and outer membranes and in the periplasm. One of the most widely-studied multi-heme proteins, the deca-heme protein MtrF (resolved in 2011 [34], see Figure 1.1, blue), features 10 tightly packed c-type hemes arranged in the peculiar shape of a staggered cross: a vertically aligned octa-heme chain is intersected horizontally by a tetra-heme chain. Located on the surface of the outer membrane, MtrF, homologous to MtrC (see Figure 1.1, red, and description below in System of focus), is supposed to be directly involved in the extracellular electron transfer [35]. Each heme cofactor consists of a central iron on the porphyrin ring and two propionate groups. As shown in Figure 1.2, the heme cofactor is bound to the protein backbone via two cysteine linkages and one of the axial histidine groups.

By deploying and combining a range of computational approaches, Marian Breuer, the previous PhD student working on this project in our lab, has computed all the thermodynamic and
kinetic parameters for heme-to-heme ET in deca-heme cytochrome MtrF. This was the first computational study on this multi-heme protein [37–39], providing molecular-level insights that are difficult to obtain from experiments. In his studies, a multi-scale modelling framework, combining non-adiabatic Marcus theory and chemical master equations, was established to model the electron transfer along the heme chain as sequential electron hopping events between neighbouring heme cofactors. The computed heme-heme ET rate constants, however, remain to be verified, because experimental measurements are only able to probe macroscopic conductivity [25,40–43], and so far have not been carried out with fully solvated protein.

Not until recently, through appropriate combinations of light-triggered electron transfer and time-resolved spectroscopy, pump-probe spectroscopies offer a powerful way to resolve pathways and dynamics of protein electron transfer, which opens up the possibility to measure heme-heme ET rate constants and obtain mechanistic insight of electron transfer in multi-heme proteins from experiment [33, 44, 45]. In order to verify our computational protocol and the computed heme-heme ET rate constants, we collaborated with Prof. Julea Butt’s lab at University of East Anglia, who are experts in protein engineering and ultra-fast pump-probe spectroscopy.

Instead of initially focusing on deca-hemes MtrF and MtrC, we have chosen to start with
Figure 1.2: Structural formula of a c-type heme cofactor, which consists of a porphyrin with an iron atom at its centre, and two propionate groups. The iron is 6-coordinated with 4 nitrogen atoms from the porphyrin ring and 2 N-methyl imidazoles from the histidine side chains. The heme cofactor is covalently bound to the protein backbone through one axial histidine and the two cysteine linkages. The structure is drawn with web-based engine at molview.org.

the tetra-heme protein STC (see Figure 1.3 B, and protein description below in System of focus), which is a much simpler protein, thus ideal for fundamental studies. To this end, our experimental collaborators docked a Ru-dye close to a terminal heme of STC and used ultra-fast pump-probe spectroscopy to monitor a single electron injection in the protein and subsequent relaxation dynamics. The heme-heme ET rate constants of STC have been computed following the previous protocol and compared to the pump-probe measured results.

In 2015, the crystal structure of MtrC (see Figure 1.1, red) has been resolved, which largely resembles MtrF (Figure 1.1, blue). Aligning the crystal structures of the two deca-heme proteins, the root-mean-square deviation (RMSD) of the heme cofactors is only 1.83 Å. Although it has been shown that MtrF can functionally replace MtrC [46,47], under anoxic conditions, only MtrC is expressed and MtrF is not expressed. This raises the question of whether different expression levels of the two proteins are due to differences in their electron transfer properties. To answer this question, the ET parameters of these two proteins have been computed and compared with each other.
Meanwhile there is a claim in the literature which states that certain redox proteins, including cytochrome c, could operate in an ergodicity-breaking regime so as to facilitate biological electron transfer on the physiological time scale [48–52]. If this were really the case, it would profoundly alter the classical picture of Marcus theory that our calculations are based on. Thus, we systematically investigated this claim for the mono-heme protein, cytochrome c, with a range of state-of-the-art methodologies. Our simulation results turn out to be assuring.

**System of main focus.** There are more than 100 types of multi-heme proteins in *Shewanella*, of which about a dozen have been resolved with a crystal structure, such as OmcA [53], UndA [54], etc. In my PhD I have focused mainly on two multi-heme proteins isolated from bacteria *Shewanella*, whose crystal structures are available with relatively high resolution – outer-membrane deca-heme protein MtrC and periplasmic tetra-heme protein STC. Their crystal structures and cellular locations are illustrated in Figure 1.3.

The crystal structure of MtrC (PDB code: 4LM8 [36], 626 amino acids, 5150 atoms, resolution 1.8 Å) was resolved from *Shewanella* in 2015. MtrC contains 10 tightly packed c-type hemes arranged in a staggered cross, with a vertically aligned octa-heme chain as its main axis, intersected horizontally by a tetra-heme chain in the middle, as does its homologue MtrF. The heme pairs in MtrC and MtrF are arranged in three types of packing motifs (see Figure 1.1 insets): T-shaped (heme pairs 1-3, 6-8), coplanar (heme pairs 1-2, 1-6, 6-7), and (partially) stacked (heme pairs 3-4, 4-5, 8-9, 9-10).

Together with the transmembrane proteins MtrA and MtrB, the extracellular electron transfer protein complex MtrC and MtrAB (see Figure 1.3 C) spans over 150 Å, where MtrA (another deca-heme protein) is partially shielded by beta barrel membrane protein MtrB, together forming the transmembrane part inserted in the bacterial outer membrane, while MtrC is fully exposed to solution outside the bacteria cell [25, 47]. Adjacent MtrC and MtrAB complexes are thought to interact via the tetra-heme chains of MtrC to facilitate micro-meter long electron transfer along the outer membrane as observed by c-AFM [43], while the octa-heme chains support ET away from the membrane and onto extracellular substrates. In this way, the heme cross motif helps supply the surface of the membrane with electrons while reducing extracellular substrates. MtrC and MtrA are evidenced to be the key components in *Shewanella* extracellular electron transport pathway [23, 25, 27].

The crystal structure of the small tetra-heme cytochrome (STC, PDB code: 1M1Q [55], 91 amino acids, 949 atoms, resolution 0.97 Å) was resolved from *Shewanella* in 2002. It contains
only 4 heme cofactors embedded in single protein domain, which is one of the smallest representatives of the multi-heme protein family with an experimentally-resolved structure. The 91 amino acids fold in a series of α-helices. Its structure is shown in Figure 1.3 B, where the four heme cofactors are embedded semi-linearly in the protein with the two terminal heme pairs packed in a T-shaped motif and the middle pair stacked. All hemes (with slight exception of heme III) are solvent-exposed, and the protein surface is homogeneously negatively charged, allowing any of the four hemes to interact with redox partners. Although this protein is suggested to function as a soluble electron carrier rather than a tetra-heme wire [55] in periplasm, it features very similar
heme-heme packing motifs as the deca-heme proteins MtrF [34] and MtrC [36]. Hence, STC is a good multi-heme protein candidate for fundamental studies because of its smaller size and simpler heme arrangement.

1.1 Thesis statement

This thesis is organized as follows.

In Chapter 2 I will first introduce three commonly-used molecular modelling approaches for biomolecular systems that I have used in my PhD – classical molecular dynamics simulation, density functional theory, and the hybrid quantum mechanics/molecular mechanics approach. The central theory for the description of electron transfer in solvated multi-heme proteins, Marcus theory (in non-adiabatic limit), will then be introduced in detail, including its assumptions, derivation, and the computational approaches for each ET parameter in that theory. Then I will introduce chemical master equations which allow us to predict macroscopic properties (i.e., experimental observable), such as the electron flux through the protein, using the heme-heme ET rate constants computed microscopically. This chapter will conclude with a typical workflow for electron transfer studies in multi-heme proteins.

Chapter 3 forms one of the essential parts of this thesis, where I will report our simulation procedure and our computed ET parameters for three multi-heme proteins: STC, MtrC, MtrF. Emphasis will be put on the computation of electronic coupling matrix elements, highlighting the non-negligible coupling enhancement effect from the cysteine side chain groups that are covalently bound to the heme groups in the sub-optimal heme pair packing motifs. Using the computed ET parameters, the maximum, protein-limited electron flux for all possible electron flow directions in the three proteins will then be estimated. I will then move on with our study of the $I-V$ curves for the three proteins, and briefly discuss the more complicated electron transport mechanism in ambient protein junctions. Two limiting electron transport mechanisms are investigated: the sequential (incoherent) hopping model and the coherent tunnelling model.

In Chapter 4, I will present our joint study (with Prof. Julea Butt’s team at University of East Anglia and at Rutherford Appleton Laboratory) on the electron transfer dynamics in Ru(bpy)$_3$-labelled STC variants, combining ultra-fast pump-probe spectroscopy and molecular simulations. After providing the experimental background, I will report the molecular modelling of exactly the same experiment, including model set-up of the labelled STC variants, prediction of charge injection and charge recombination rate constants from/to the Ru-dye, and the influence of the Ru-dye on heme reduction potentials. This will be followed by an interpretation of the experi-
mental signal, in which I will detail different trial kinetic models for the reaction scheme, and our fitting procedure. The fit heme-heme ET rate constant is in good agreement with our predicted time scale using the electron hopping model, which partially validates our computational protocol for solvated multi-heme proteins.

Chapter 5 deals with the effect of electronic polarizability on the prediction of reorganization free energy. We have systematically computed this Marcus ET parameter for the oxidation of a prototypical redox protein, cytochrome c, with a variety of state-of-the-art methodologies ranging from non-polarizable MD simulations, QM/MM calculations with the polarizable QM centre described by two different methods, and polarizable MD simulation. I will clearly show that the overestimation of experimental reorganisation free energies for cytochrome c typically obtained with non-polarizable force fields is due to missing explicit electronic polarisation of the protein and water in the force fields used, in contrast to a recent claim which states that certain redox proteins, including cytochrome c, could operate in an ergodicity-breaking regime so as to facilitate biological electron transfer on the physiological time scale [50–52].

Finally, Chapter 6 concludes the findings of my PhD and provides an outlook on extensions and continuations of the work presented here.
Chapter 2

Theoretical Background and Computational Approaches

In this chapter, I will first introduce the most commonly-used computational approaches for molecular modelling in biomolecular systems. These include: molecular dynamics simulation (subsection 2.1.1), density functional theory (subsection 2.1.2), and the hybrid quantum mechanics/molecular mechanics (QM/MM) method (subsection 2.1.3). Next, I will focus on the particular approaches for studying electron transfer (ET) problems. These include Marcus theory (section 2.2), and the computational approaches used to compute the Marcus ET parameters: electronic coupling matrix elements (subsection 2.3.1), reorganisation free energies (subsection 2.3.2) and driving forces, or free energy differences (subsection 2.3.3). In section 2.4, I will introduce chemical master equations which link the computed microscopic Marcus ET rates to macroscopic observables. Finally, I summarise the typical workflow for ET property calculation in section 2.5.

2.1 Molecular modelling approaches for biomolecular systems

2.1.1 Molecular dynamics (MD) simulation

Classical molecular dynamics (MD) simulation applies Newton’s second law to each individual atom in a molecular/atomic system. Atoms and molecules evolve in a known potential and the forces acting on them are first derivatives of this potential \( V(\{R\}) \),

\[
F_i(\{R\}) = -\nabla_i V(\{R\}) = m_i \cdot \dot{v}_i
\]  

(2.1)

where \( F_i \) is the force acting on atom \( i \), \( \{R\} \) represents the nuclear coordinates, \( V(\{R\}) \) is the potential including all the interactions, \( m_i \) is the mass of atom \( i \) and \( v_i \) is the corresponding
velocity.

In classical MD, potential \( V(\{R\}) \) is described by some functional forms with a set of parameters, together termed force field (FF, see subsection 2.1.1.1). The parameters are obtained from fits to experimental data (such as vibrational spectra, densities, heats of vaporization) and/or high-level ab-initio calculations (such as vibrational spectra, minimum energy geometries, dipole moments) [56]. In ab-initio MD, forces are computed directly from quantum mechanical calculations for each snapshot and \( V(\{R\}) \) is updated on the fly; however, this is not the method used in my work.

Velocity Verlet, with the advantage of good numerical stability and moderate memory cost, is a common algorithm to propagate the dynamics [57],

\[
\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t) \Delta t + \frac{1}{2} \vec{a}(t) \Delta t^2
\]

(2.2)

\[
\vec{v}(t + \Delta t) = \vec{v}(t) + \frac{\vec{a}(t) + \vec{a}(t + \Delta t)}{2} \Delta t
\]

(2.3)

in which coordinates \( \vec{x} \) and velocities \( \vec{v} \) are updated at the same step, and the global error in both of them is \( O(\Delta t^2) \). \( \vec{a} \) is the acceleration and \( \Delta t \) is the timestep. Velocity Verlet is also the algorithm implemented in NAMD [58] and AMBER TOOLS [59] which I use throughout this work.

One also needs to choose an ensemble to run the simulation (i.e., choose a reasonable thermostat and barostat under a particular temperature and/or pressure). Common ensembles include NVE (microcanonical ensemble, no thermostat or barostat), NVT (canonical ensemble, thermostat added) and NPT (isothermalisobaric ensemble, thermostat and barostat added). Common thermostats include velocity rescaling, Langevin thermostat, Nose-Hoover thermostat, etc. Common barostats include Nose-Hoover barostat, Langevin barostat and so on [60].

Finally, the initial structure for MD must be chosen carefully. The timestep has to be at least 10 times smaller than the highest vibrational frequency in the system of interest. A timestep of 2 fs is commonly used in biomolecular MD simulation when O-H bonds are constrained during the MD through the use of SHAKE [61] or LINCS [62] algorithms. Unless one wants to purposefully carry out non-equilibrium MD, the simulation system has to be equilibrated in certain ensemble before any meaningful result can be obtained. To this end, total energy, potential energy, temperature, pressure, densities and root-mean-square deviation (RMSD) are common observables to check. MD trajectories sample configurations in certain statistical ensembles. The
time series of these configurations also allows for the study of dynamic properties and processes.

2.1.1.1 Classical force field

In classical force field (FF), atoms are explicitly treated as mass points, while electrons and environmental effect are implicitly treated by assigning a partial charge to each atom. A topology file is created to record the atom type (e.g., \( sp^3 \) carbon, \( sp^2 \) nitrogen) and linkage information of the simulation system. A parameter file stores all the information of pair-wise interactions among the atoms. Common FFs for biomolecular systems include AMBER [59], CHARMM [64], GROMOS [65] etc. They are parameterized with different experimental and QM data. For example, AMBER partial charges are parameterized to reproduce molecules’ gas-phase electrostatic potentials, while CHARMM partial charges are targeted to reproduce dimerization energies and minimum-energy interaction distances for small molecule-water dimers in the condensed phase. This means that each force field most likely performs best at reproducing the particular data for which its partial charges were parameterized [56]. Despite the differences, force fields usually contain the following interaction terms:

\[
U = \sum_{\text{bonds}} \frac{1}{2} K_b (b - b_0)^2 + \sum_{\text{angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \sum_{\text{torsions}} K_\phi \left[ 1 - \cos(n\phi + \delta) \right]
\]
\[ + \sum_{\text{non-bonded} i,j} \left[ \varepsilon \left( \frac{(r_{0ij}/r_{ij})^{12}}{(r_{0ij}/r_{ij})^6} - 2 \right) + q_i q_j/r_{ij} \right] \]  

(2.4)

where \( K_b, K_\theta, K_\phi, \varepsilon \) are force constants of bond vibrations, bond bendings, torsions and Van der Waals interactions, \( b, \theta, r_{ij} \) are bond length, angle, non-bonded atom-atom distance and \( b_0, \theta_0, r_{0ij} \) are the corresponding equilibrium values. \( \phi \) is the dihedral period, \( \delta \) is dihedral phase shift, and \( q_i \) is the partial charge of atom \( i \). As illustrated in Figure 2.1, the five terms in the potential represent bond vibration interactions between any two bonded atoms, angle bending interactions between any two adjacent bonds, dihedral angle (torsion) interactions between any two adjacent angles expressed as a Fourier series, Van der Waals interactions between any two unbonded atoms (in the form of the Lennard-Jones potential) and electrostatic (Coulomb) interactions between any two unbonded atoms, respectively. A cut-off is applied so that Van der Waals interactions between two atoms that are far apart (e.g., > 12 Å) are not taken into account. Electrostatic interactions, however, decays more slowly, as \( 1/r \), so cannot be omitted at large distance. Coulomb interactions that are within cut-off are computed with direct summation in real space, while the remaining long-range interactions are computed in reciprocal space to accelerate the computation. This is referred to as Ewald summation [66]. The potential, equation 2.4, is given in atomic units.

2.1.1.2 Polarizable force field

Electronic polarization, as a ubiquitous phenomenon in condensed phases, originates from the redistribution of electron density of a certain molecule due to interaction with other molecules. While polarization is effectively included in the version of non-polarizable molecular mechanical force fields, which can usually achieve sufficiently accurate description on properties such as geometry and energy [67, 68], there are certainly other situations (e.g., polar, heterogeneous systems) where explicit treatment of electronic polarization is necessary [69–71]. The explicit treatment of polarization, to some extent, introduces a nonadditive \( n \)-body effect [72].

At present, there are four main types of polarizable force field: polarizable continuum model, fluctuating charge model, Drude oscillator model and induced point dipole model [72]. Induced dipole model is the most commonly used, due to its analytical dipole term, high transferability and straightforward generalization to include higher-order induced multipoles. Protein force fields that use induced dipole model include AMBER02 [59], AMOEBA [73] and OPLS/PFF [74]. POL3 [75] is a common polarizable water model.
2.1. Molecular modelling approaches for biomolecular systems

In the induced dipole model, there is an extra term $U_{\text{pol}}$ contributing to the potential energy,

$$ U = U_{\text{bonded}} + U_{\text{nonbonded}} + U_{\text{pol}} \quad (2.5) $$

$$ U_{\text{pol}} = -\frac{1}{2} \sum \mu_i \cdot E_i^0 = -\frac{1}{2} \sum \alpha_i E_i \cdot E_i^0 \quad (2.6) $$

where $\mu_i$ is the induced dipole moment of atom $i$, $\alpha_i$ is the isotropic point polarizability of atom $i$, $E_i$ is the overall electrostatic field at atom $i$ due to charges and induced dipoles of all other atoms, and $E_i^0$ is the unperturbed electrostatic field at atom $i$ due to the partial charges of all other atoms. Unlike the non-polarizable force field, the point charge interaction in the $U_{\text{nonbonded}}$ is now parameterized from DFT in gas-phase, as the polarization effect in the condensed phase context is included explicitly through the $U_{\text{pol}}$ term.

The existence of $U_{\text{pol}}$ means that the potential energy is no longer additive. This is because the induced dipole of a given atom $i$ depends on the electrostatic field $E_i$ of all other atoms, and $E_i$ further depends on the induced dipoles at all other atom sites apart from atom $i$. To properly calculate the interaction energy would require an iterative approach to self-consistently solve the induced dipoles of all atoms, which is computationally demanding when dealing with long MD trajectories. Therefore, an extended Lagrangian formalism, available in AMBER TOOLS [59], analogous to Car-Parinello dynamics, is often used to propagate the dipoles as additional dynamical variables in MD simulations.

2.1.2 Density functional theory (DFT)

Density functional theory (DFT), due to its excellent balance between accuracy and computational cost, is so far the most widely-used quantum chemistry approach for condensed matter systems. Instead of working with ground state many-body wave functions, the main idea of DFT is to solve the ground state electron density $\rho(r)$ by minimizing the total energy with respect to the density. This is based on the two Hohenberg-Kohn theorems [76], which state that the external potential (and hence the total energy), is uniquely determined (up to a trivial constant) by the ground state electron density, and that the ground state energy and the corresponding density is obtained by minimizing the density functional with respect to density.

The core of DFT is the Kohn-Sham equation that solves for a (non-interacting) single-electron in an effective potential,

$$ \left[ -\frac{\nabla^2}{2} + v_{\text{eff}}(r) \right] \psi_n(r) = \varepsilon_n \psi_n(r) \quad (2.7) $$
\( v_{\text{eff}}(r) = v_{\text{ext}}(r) + v_{\text{Hart}}(r) + v_{\text{xc}}(r) \) (2.8)

where \( \psi_n(r) \), \( \varepsilon_n \) are the Kohn-Sham orbitals and Kohn-Sham energies, \( v_{\text{ext}} \) is the electric field due to nuclei and external field (such as electrodes), \( v_{\text{Hart}} \) and \( v_{\text{xc}} \) are the Hartree potential and exchange-correlation potential, respectively.

\[
v_{\text{Hart}}(r) = \frac{\delta E_{\text{Hart}}}{\delta \rho(r)} = \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|r - r'|} \tag{2.9}
\]

\[
v_{\text{xc}}(r) = \frac{\delta E_{\text{xc}}}{\delta \rho(r)} \tag{2.10}
\]

The Kohn-Sham equations have to be solved in a self-consistent (SC) manner because \( v_{\text{eff}} \) depends on electron density which further depends on \( \psi_n(r) \). Starting from an initial trial density \( \rho_0 \), \( v_{\text{Hart}} \) and \( v_{\text{xc}} \) are computed using equations 2.9 and 2.10, and inserted in equation 2.7 to solve for \( \psi_n(r) \) and \( \varepsilon_n \). The electron density \( \rho(r) = 2\sum_{\text{occ}} |\psi_n(r)|^2 \) is then constructed by filling up the electrons in the lowest \( N \) Kohn-Sham spin orbitals (\( N \) equals the total number of electrons). This updated density is used in the next SC step to solve the Kohn-Sham equation until the density gradient is below some threshold (e.g., \( 10^{-6} \) a.u.). Finally, the total energy is expressed in terms of electron density and Kohn-Sham energies:

\[
E_{\text{total}} = \sum_i^N \varepsilon_i - E_{\text{Hart}}[\rho] + E_{\text{xc}}[\rho] - \int \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})} \rho(\mathbf{r}) d\mathbf{r} \tag{2.11}
\]

Note that up to here everything is exact. The only issue is that the functional form of exchange-correlation potential \( v_{\text{xc}} \) is unknown, which is where approximations come in.

Researchers have put much effort into developing DFT functionals for all kinds of materials since the birth of DFT [77, 78]. There is a hierarchy of exchange-correlation functionals modify this: local density approximation (LDA) functional only depending on the density at a given point, gradient density approximation (GGA) functional adding the density gradient, meta-GGA functional adding the second derivative, hybrid functional mixing some portion of exact Hartree-Fock exchange to pure GGA functional, random phase approximation (RPA) adding the unoccupied orbitals. Generally speaking, computational accuracy increases when more information is included. However, there is no functional that works best for all systems; one always needs to test DFT functionals for a specific system.

In my PhD I have used PBE [79], PBE0 [80] and PBE50 functionals for heme cofactors and some other isolated molecules. PBE is a widely-used GGA functional which is derived only
with the information from the slowly varying electron density and known exact conditions on the functional [78]. PBE0 and PBE50 are hybrid functionals with 25% and 50% of exchange functional substituted by exact Hartree-Fock exchange $E_{x}^{\text{HF}}$, as defined in equations 2.13 and 2.14. Introduction of exact exchange eases the self-interaction error in DFT [81,82] which is an artifact that electron feels itself locally because of the density representation.

$$E_{x}^{\text{HF}} = -\frac{1}{2} \int dr \int dr' \psi_i^{*}(r) \psi_j^{*}(r') \frac{1}{|r-r'|} \psi_j(r) \psi_i(r')$$  \hspace{1cm} (2.12)

$$E_{xc}^{\text{PBE0}} = \frac{1}{4} E_{x}^{\text{HF}} + \frac{3}{4} E_{x}^{\text{PBE}} + E_{c}^{\text{PBE}}$$  \hspace{1cm} (2.13)

$$E_{xc}^{\text{PBE50}} = \frac{1}{2} E_{x}^{\text{HF}} + \frac{1}{2} E_{x}^{\text{PBE}} + E_{c}^{\text{PBE}}$$  \hspace{1cm} (2.14)

where $E_{x}^{\text{HF}}, E_{x}^{\text{PBE}}, E_{c}^{\text{PBE}}$ are exact exchange, PBE exchange and PBE correlation energies, respectively. Computationally, hybrid functionals are about one order of magnitude more expensive than GGA functionals. However, the hybrid scheme demonstrates improved predictions of energetics, particularly suited for organic molecules [77,83].

Finally, to actually solve the Kohn-Sham equation using linear algebra, a basis set has to be used. Inspired by condensed matter systems, people have developed two kinds of basis sets: localized basis sets, such as Gaussian basis sets and Slater basis sets, which resemble electron orbitals in isolated molecules, and plane-wave basis sets which resemble Bloch states in periodic solid state crystals. Since it is impossible to use infinite number of functions (i.e., achieve completeness) in an actual computation, one needs to test the convergence with respect to the size of basis set.

When dealing with plane-wave basis sets, a common computational trick in modern DFT calculations is the utilisation of pseudopotentials. The idea of pseudopotentials is to get rid of the computation of high-energy, oscillatory interaction between core electrons and the nucleus, which is extremely expensive for plane waves (because it requires a high energy cut-off) but relatively unimportant since only valence electrons are involved in most chemical reactions. To this end, an effective potential [84] is constructed assuming the core electrons are frozen and taken in by the nucleus. This potential should not alter the electric field that the valence electrons feel. Within a cut-off radius $r_C$ of the nucleus, the pseudopotential is much shallower than the full potential and those valence electrons are described by pseudo-wave functions with significantly fewer nodes. Different kinds of pseudopotentials use different constraints upon construction,
therefore needs to test the basis set convergence when choosing a new pseudopotential.

2.1.3 Quantum mechanics/molecular mechanics (QM/MM)

The modelling of certain chemical reactions in the condensed phase ideally requires a precise description of the rearrangement of electron distribution between reactants and products, as well as efficient inclusion of the wider environment and its effects on reaction energetics [85]. While reactive force fields can incorporate the information of bond breaking and bond formation without expensive QM calculations [86], this often requires many empirical parameters [87], which results in difficulties when considering complex biological systems. In QM/MM, the reaction centre is modelled at QM level (usually DFT) while the rest of the system is modelled at MM level (usually classical MD). On the one hand, QM methods are able to address the required precision but are usually restricted to a few hundred atoms; on the other hand, MM is capable of treating up to hundreds of thousands of atoms in MD simulations over tens of nanoseconds [88], with good enough accuracy to describe the region that is more distant from the reaction centre. QM/MM combines the strength of both levels of theories and particularly suits the study of enzyme reactions where only embedded protein cofactors located in small regions have to be treated at QM level.

Since the birth of QM/MM [89], and especially in the past two decades when QM/MM has been implemented into many codes as either a standard module [90–92] or high-performance interface [93, 94], there are many established studies using QM/MM such as hydrogen bond network, spectroscopic properties, catalyst design, predictions of ligand binding affinities and modelling enzyme mechanisms [85, 95]. QM/MM was awarded the Nobel Prize in Chemistry in 2013.

The QM/MM energy for the entire system, $E_{\text{QM/MM}}$, is defined by:

$$E_{\text{QM/MM}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM-MM}}$$

(2.15)

where $E_{\text{QM}}$, $E_{\text{MM}}$ are the energies of QM and MM subsystems, respectively, and $E_{\text{QM-MM}}$ is the interaction energy between the QM and MM subsystems. Bonded and non-bonded Van der Waals interactions of $E_{\text{QM-MM}}$ are treated at MM level. Depending on how the non-bonded electrostatic interaction (which usually dominates $E_{\text{QM-MM}}$) is treated between the QM and MM subsystems, there are three main types of QM/MM schemes: mechanical embedding, electrostatic embedding, polarizable embedding. In mechanical embedding (also termed subtractive QM/MM
2.1. Molecular modelling approaches for biomolecular systems

In electrostatic embedding (also termed additive QM/MM scheme), the MM subsystem is included as external field in the QM calculation so that the QM subsystem is polarized by the MM subsystem. In polarizable embedding, MM subsystems are described with polarizable force field. Mutual polarization between the QM and MM subsystems is invoked and QM electron density and MM multipoles are optimized together self-consistently.

The electrostatic embedding scheme is the standard implementation in most QM/MM codes such as CP2K [90] which I used in my PhD. In electrostatic embedding, the electrostatic interaction between the QM and MM subsystem is defined by:

$$E_{\text{elec}, \text{QM-MM}}[\rho(x), Z_J, R_J, q_I, R_I] = \sum_{I=1}^{\text{MM}} q_I \int_{\text{QM}} \rho(x) \frac{d^3x}{|x - R_I|} + \sum_{I=1}^{\text{MM}} \sum_{J=1}^{\text{QM}} \frac{q_I Z_J}{|R_I - R_J|} \tag{2.16}$$

where the first term in the right hand side represents the particle-mesh interaction between QM electron density and MM partial charges, and the second term represents the particle-particle interaction between QM nuclei and MM atoms. $\rho(x)$ is electron density in QM subsystem, $Z_J, R_J$ are charge and position of nucleus $J$ in the QM subsystem, $q_I, R_I$ are partial charge and position of atom $I$ in the MM subsystem. $E_{\text{elec}, \text{QM-MM}}$ together with $E_{\text{QM}}$ can be computed by an extended DFT calculation of the QM subsystem in the external Coulomb potential generated by the MM partial charges.

One tricky question that arises is how to choose the active QM subsystem. In most cases, one cannot avoid cutting a covalent bond when dividing the total system to QM and MM subsystems. To avoid artificial overpolarization at QM boundaries, nonpolar C-C bonds are suggested to be cut and set as the QM-MM boundaries. There are several approaches to treat the boundary atoms, such as the link atom method [96], the generalized hybrid orbital method (GHO) [97] and optimized effective core pseudopotentials [98]. The GHO method is the standard approach implemented in CP2K, in which the four $sp^3$ orbitals of the boundary MM atom are included in the QM calculation. However, only the one that was forming the covalent bond with the boundary QM atom is active in the SCF optimizations, while the other three auxiliary orbitals remain frozen. Such charge reoptimization of the boundary MM atom allows GHO to prevent, or at least reduce the overpolarization which is sometimes found in the link atom approach even when QM-MM boundaries are properly set up between two aliphatic carbon atoms [95]. Finally, QM/MM calculations have to be tested with respect to the increase of size of QM subsystem. The chosen QM simulation box should be large enough so that the QM subsystem (i.e., reaction
centre and some of its neighbouring atoms) should stay stably inside the QM box throughout the entire MD simulation, unless the adaptive buffered force QM/MM method [99] is used.

### 2.2 Marcus theory in non-adiabatic limit

The non-adiabatic Marcus theory [100–102] is a semi-classical theory to describe thermally induced electron transfer (ET) processes mediated by electron hopping from one chemical species (termed donor) to another (termed acceptor) in aqueous solution. It is usually explained in the representation of charge localized or diabatic states with a pair of parabolas [4]. As shown in Figure 2.2, the two blue parabolas represent the diabatic free energy surfaces for initial and final states. The vertical energy gap $\Delta E$ is chosen as the reaction coordinate because it leads to an important linear free energy relation,

$$
\Delta A(\Delta E) := A_B(\Delta E) - A_A(\Delta E) = \Delta E
$$

(2.17)

i.e., the vertical free energy gap is equal to the vertical energy gap for any $\Delta E$. This linear relationship is valid as long as the sampling obeys a Boltzmann distribution [4]. Marcus theory combines the idea of transition state theory [103] and the quantum transition probability from Landau-Zener theory [104]. In the initial diabatic state $A$, the excess electron is localized at the donor site and the system is oscillating around $\Delta E^{\text{min}}_A$. When certain (rare) thermal fluctuations

![Figure 2.2: Free energy curves for electron transfer between electron donor and electron acceptor in a dielectric environment (solution or protein). Reprint from Ref. [4] with permission.](https://pubs.acs.org/doi/10.1021/acs.chemrev.5b00298) Further permissions related to the material excerpted should be directed to the ACS.)
bring the system to the crossing point (i.e., $\Delta E = 0$) such that the energy levels of the initial and the final states match. ET can occur at a certain probability ($\sim \langle |H_{ab}|^2 \rangle$) by hopping from one potential energy surface to the other, followed by relaxation to the equilibrium of the final state $\Delta E_B^{\text{min}}$. A more detailed derivation is available in Appendix A.

Marcus theory provides an estimate of ET rate as a function of four parameters (equation 2.18): temperature $T$, free energy difference (or driving force) $\Delta A$, reorganisation free energy $\lambda$, and electronic coupling matrix element $H_{ab}$ between initial and final diabatic states,

$$k_{ET} = \frac{2\pi}{\hbar} \langle |H_{ab}|^2 \rangle (4\pi \lambda k_B T)^{-\frac{1}{2}} \exp\left(-\frac{(\Delta A + \lambda)^2}{4\lambda k_B T}\right)$$

(2.18)

where $\hbar$ is the reduced Plank constant and $k_B$ is the Boltzmann constant.

There are two notable differences between Marcus theory and transition state theory. One is the $\langle |H_{ab}|^2 \rangle$ prefactor which comes from Landau-Zener transition probability in the non-adiabatic limit (or weak coupling limit). The square dependence of electronic coupling in the rate expression makes $H_{ab}$ an important parameter in the ET rate estimation. The other difference is the reaction barrier, $\frac{(\Delta A + \lambda)^2}{4\lambda}$, which is a combination of thermodynamic driving force and the rearrangement of solution molecules. The square in the numerator, $(\Delta A + \lambda)^2$, indicates the existence of a maximum barrier-less point as $\Delta A$ goes to large negative values (i.e., $-\Delta A = \lambda$), and an inverted region (i.e., $-\Delta A > \lambda$) where ET rate decreases with increasing driving force. Marcus theory was awarded the Nobel Prize in Chemistry in 1992 after the inverted region was unequivocally verified experimentally in 1984 [105].

Note that although ET theory is introduced in the diabatic picture, the diabatic states are not well-defined and are not even unique. Instead, adiabatic states (plotted with red parabolas in Figure 2.2) are the eigenstates of the Hamiltonian, which are often solved for first (e.g., with quantum chemical methods [106]) to assist in the construction of diabatic states.

### 2.2.1 Assumptions and validations

In this section I justify the validity of Marcus theory in multi-heme protein systems. The following assumptions are made in Marcus theory:

- The electron is localized on the donor (acceptor) in the initial (final) state, which is valid in our multi-heme proteins as we usually have $|H_{ab}| \ll \lambda$ [107], and donor and acceptor heme cofactors are geometrically well-separated by a centre-to-centre distance of $\sim 15 \text{Å}$.
- Background fluctuations satisfy a Gaussian distribution. This can be well-justified as the
protein systems of interest usually contain \( \sim 100,000 \) atoms (including water molecules). According to central limit theorem the fluctuation distribution is expected to be Gaussian. Indeed, as shown in Chapter 5 the vertical energy gap fluctuations satisfy a Gaussian distribution.

- Initial and final diabatic free energy curves have the same parabolic curvatures. The potential energy surfaces are parabolas under Gaussian statistics. It is reasonable to assume same curvature for the initial and final states as the majority of the system does not experience significant change before and after ET.

- ET has to be slower than coupled molecular motions so that the system is well-relaxed before and after ET. Most of the heme-heme ET in our system is on 100 ps - 10 \( \mu s \) time scale which is much slower than molecular vibration frequencies which are usually on fs to ps time scales [108].

- The vertical energy gap \( \Delta E \) depends linearly on time during ET. This is a reasonable assumption for a small piece of a quadratic function (i.e., near the crossing point).

- Diabatic states are time independent. This is a good assumption on the time scale of ET.

- The flux through transition region is constant. This is a reasonable statistical assumption on the time scale of ET.

2.2.2 Heterogeneous electron transfer (ET)

The heterogeneous ET expression estimates the interface ET rates between an absorbed molecule on a metal electrode and the electrode. It takes into account electron hopping between frontier orbitals of the molecule and all possible states in the conduction band of the electrode, weighted by Fermi-Dirac distribution. The heterogeneous ET rates are given by the electrochemical nonadiabatic ET rate equation [37, 109],

\[
\begin{align*}
k_{\text{in}} &= C_{\text{elec}} \int_{-\infty}^{+\infty} \exp \left[ - \left( x - \frac{\lambda + e(E - \epsilon)}{k_B T} \right)^2 \left( \frac{k_B T}{4\lambda} \right) \right] \frac{1}{1 + \exp(x)} dx \\
k_{\text{out}} &= C_{\text{elec}} \int_{-\infty}^{+\infty} \exp \left[ - \left( x - \frac{\lambda - e(E - \epsilon)}{k_B T} \right)^2 \left( \frac{k_B T}{4\lambda} \right) \right] \frac{1}{1 + \exp(x)} dx
\end{align*}
\]
2.3 Computation of electron transfer (ET) parameters

2.3.1 Electronic coupling $H_{ab}$

The electronic coupling, $H_{ab} = \langle \psi_a | H | \psi_b \rangle$, is a quantum mechanical ET parameter describing the transition probability from the initial state $\psi_a$ to the final state $\psi_b$ near the crossing point of the two diabatic potential energy surfaces.

The diabatic state wave functions can be constructed either directly from DFT calculations, such as in constrained DFT [110], scaled fragment-orbital DFT (sFODFT) [111], or from DFT-calculated adiabatic ground states applying further diabatization schemes, such as projector operator-based diabatization (POD) [106]. In my PhD I have used sFODFT and POD methods.

2.3.1.1 Scaled fragment-orbital density functional theory (sFODFT)

In the fragment-orbital density functional theory (FODFT) approach [112] the diabatic states are constructed from frontier orbitals of isolated donor or acceptor fragments. It is based on the approximation that the initial and final wave functions only differ in their highest occupied molecular orbitals (HOMO). This approach retains the essential physics and is free from any spurious delocalization between the two fragments because polarization between these two fragments is completely excluded [39].

Consider the following ET process $D^- + A \rightarrow D + A^-$. For simplicity we assume D and A have the same number of electrons $N$. The total number of electrons in this system is thus $2N + 1$. The initial and final state diabatic wave functions $\psi_a$ and $\psi_b$ are described by Slater determinant of $2N + 1$ spin orbitals $\phi$, which are taken from the spin orbitals of isolated (non-interacting) $D^-$ and $A^-$ fragment:

$$
\psi_a \approx \psi_{DA}^a = \frac{1}{\sqrt{(2N+1)!}} det(\phi^1_D, \ldots, \phi^{N+1}_D, \ldots, \phi^1_A, \ldots, \phi^N_A) 
$$

$$
\psi_b \approx \psi_{DA}^b = \frac{1}{\sqrt{(2N+1)!}} det(\phi^1_D, \ldots, \phi^N_D, \ldots, \phi^1_A, \ldots, \phi^{N+1}_A)
$$

FODFT is implemented in CPMD codes [91]. First, two DFT calculations are carried out, one for the reduced donor and one for the reduced acceptor, giving two sets of Kohn-Sham
orbitals. The two sets of orbitals are then bi-orthogonalized [113] and used for construction of the Kohn-Sham Hamiltonian of the total system comprised of donor and acceptor. The coupling matrix element is finally given by equation 2.26:

\[ H_{ab} = \langle \psi_a | H | \psi_b \rangle \]  
\[ \approx \langle \psi_a | H_{b,KS}^{|KS} | \psi_b \rangle \]  
\[ \approx \langle \psi_{DA}^a | H_{b,KS}^{|KS} | \psi_{DA}^b \rangle \]  
\[ = \langle \phi_{N+1}^{D} | H_{b,KS}^{|KS} | \phi_{A}^{N+1} \rangle \]  

where \( H \) is the full Hamiltonian, \( H_{b,KS}^{|KS} = \sum_{i=1}^{2N+1} \epsilon_i | \psi_i \rangle \langle \psi_i | \) is the one-particle Kohn-Sham Hamiltonian constructed from \( 2N+1 \) orbitals of \( \psi_{DA}^b \), \( \psi_{a,b} \) represent initial and final state wave functions, \( \psi_{DA}^a,b \) represent initial and final state wave functions constructed from bi-orthogonalized fragment orbitals, \( \phi_{N+1}^{D} \) represents the highest occupied donor orbital, and \( \phi_{A}^{N+1} \) represents the lowest unoccupied acceptor orbital, which is further approximated by the HOMO of the reduced acceptor.

A benchmark study of FODFT at PBE level against high-level ab-initio calculations showed that FODFT couplings are systematically underestimated by a factor of 1.348 [114] due to the missing electronic polarization between donor and acceptor in this method. The raw coupling values are multiplied by this scaling factor. This method is thus denoted as scaled fragment-orbital density functional theory (sFODFT).

2.3.1.2 Projector operator-based diabatization (POD)

The projector operator-based diabatization (POD) method [115] constructs diabatic states from the Kohn-Sham (KS) adiabatic states

\[ \hat{H} | \psi_i \rangle = \epsilon_i | \psi_i \rangle, \]  

which are obtained by standard DFT calculations of the entire system. The adiabatic states are then represented in an orthonormalized basis set of atom-centre localized functions \( \{ \phi_i \} \), and the Hamiltonian is partitioned to donor and acceptor blocks according to the donor-acceptor separation. After a separate diagonalization of the donor and acceptor blocks of the partitioned
2.3. Computation of electron transfer (ET) parameters

Hamiltonian, the system Hamiltonian is transformed to the following form

\[
\hat{H} = \begin{bmatrix}
\varepsilon_{D,1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \varepsilon_{D,N} \\
\varepsilon_{A,1} & \cdots & 0 \\
\hat{H}_{DA} & \vdots & \vdots \\
0 & \cdots & \varepsilon_{A,M}
\end{bmatrix}
\] (2.28)

While there are \(N\) donor and \(M\) acceptor one-electron energies \(\varepsilon_{\alpha,i}\) of diabatic states on the main diagonal, the off-diagonal blocks \(\hat{H}_{DA}, \hat{H}_{AD}\) contain electronic coupling elements between donor and acceptor diabatic states.

Compared to sFODFT, POD has the correct number of electrons during electronic structure calculations and fully includes donor-acceptor interaction. The POD method [106] is recently implemented in the CP2K software package [90], which allows for efficient calculation of electronic coupling elements \(H_{ab}\) at hybrid-functional level of DFT.

2.3.2 Reorganisation free energy \(\lambda\)

Reorganisation free energy can be interpreted as the free energy needed to rearrange the nuclei configuration from the initial state to the final state while keeping the electron localized at the initial state, \(\lambda = A_a(\Delta E_{b}^{\text{min}}) - A_a(\Delta E_{a}^{\text{min}})\), as illustrated in Figure 2.2, where \(A_a\) represents the free energy surface of the initial state, \(\Delta E_{M}^{\text{min}}, M=a,b\) represents the equilibrium geometries of the initial and final states. In the limit of linear response, one can show that this definition is equivalent to \(\lambda^\text{st}\) and \(\lambda^\text{var}\) defined below.

There are two formulas for computing this ET parameter, Stokes and variance reorganisation free energies, \(\lambda^{\text{st}}\) and \(\lambda^{\text{var}}_{M}\) respectively, which are defined by

\[
\lambda^{\text{st}} = (\langle \Delta E \rangle_A - \langle \Delta E \rangle_B)/2 \\
\lambda^{\text{var}}_{M} = \frac{\sigma_{M}^2}{2k_B T}, \quad M = A, B,
\] (2.29, 2.30)

where \(\Delta E\) is the vertical energy gap,

\[
\Delta E(R^N) = E_B(R^N) - E_A(R^N),
\] (2.31)
\( E_M(\mathbf{R}^N) \) is the potential energy of initial \((M = A)\) or final \((M = B)\) state at the nuclear configuration \(\mathbf{R}^N\), \(\langle \cdot \cdot \cdot \rangle_M\) denotes the thermal average on the potential energy surface of redox state \(M\), and

\[
\sigma^2_M = \langle (\Delta E - \langle \Delta E \rangle_M)^2 \rangle_M \tag{2.32}
\]

is the variance of the vertical energy gap fluctuations. \(\lambda_{\text{st}}\) is equal to half the horizontal offset of the free energy curves of initial and final diabatic states (the two blue parabolas in Figure 2.2), and \(\lambda_{\text{var}}\) is related to their curvatures (or force constant, \(k \propto 1/\lambda_{\text{var}}\)). Assuming Gaussian gap fluctuations [116] and ergodic sampling, and using the fluctuation-dissipation theorem [117], it can be shown that \(\lambda_{\text{st}} = \lambda_{\text{var}} = \lambda_{\text{var}}^R = \lambda_{\text{var}}^O\) [4], so that only one reorganisation free energy (\(\lambda\)) appears in Marcus theory and the initial and final state free energy parabolas share the same curvature.

The reorganisation free energies can be obtained by sampling the vertical energy gap \(\Delta E\) along MD trajectories in initial and final diabatic states. \(\langle \Delta E \rangle_A\) and \(\sigma_A\) are obtained by replaying state A trajectory with state B potential. \(\langle \Delta E \rangle_B\) and \(\sigma_B\) can be obtained similarly through evaluation of the energy gap \(\Delta E\) along state B trajectory. The reorganisation free energies for ET between states A and B are then obtained as equations 2.29 or 2.30.

\(\lambda_{\text{st}}\) can be decomposed to outer-sphere and inner-sphere contributions, \(\lambda_{\text{st}} = \lambda_o + \lambda_i\), where inner-sphere reorganisation free energy \(\lambda_i\) comes from the contribution of donor and acceptor species, which is usually the redox-active cofactor, while outer-sphere reorganisation free energy \(\lambda_o\) is the contribution of the rest of the simulation system (i.e., protein and solvent).

Marcus continuum formula can be used to estimate the outer-sphere reorganisation free energy \(\lambda_o\). In atomic unit:

\[
\lambda_o = \left( \frac{1}{2r_1} + \frac{1}{2r_2} - \frac{1}{R} \right) \cdot \left( \varepsilon_{\text{op}}^{-1} - \varepsilon_s^{-1} \right) \cdot \Delta q^2 \tag{2.33}
\]

where \(r_1, r_2\) are radii of donor and acceptor conducting spheres, and \(R\) is their separation, \(\varepsilon_{\text{op}}\) and \(\varepsilon_s\) are optic (high frequency) and static dielectric constant of the solvent, respectively, \(\Delta q\) is the amount of the charge transferred. \((\varepsilon_{\text{op}}^{-1} - \varepsilon_s^{-1})^{-1}\) is also termed “Pekar factor” which implicitly takes into account solvent polarization.

Since the relaxation of different dynamic processes in proteins spans a wide range of time scales from fs to s [108], the trajectory for calculating reorganisation free energy should in principle be on the same time scale of the charge transfer process of interest. For ultrafast (picosecond) processes such as Photosystem II (PS II) [48, 118–120], the reorganisation free energy will
be strongly overestimated if the vertical energy gap is sampled on time scales longer than the actual ET event [118, 119]. A potential solution to this problem is to apply a self-consistent non-ergodicity correction to filter out frequency components that are slower than the ET event [4,121].

2.3. Computation of electron transfer (ET) parameters

2.3.3 Driving force $\Delta A$

The driving force $\Delta A$ (or free energy difference) for the reaction can be estimated from the difference in reduction potentials. The reduction potential can be computed by thermodynamic integration (TI) or solving the linearized or nonlinear Poisson-Boltzmann (P-B) equations.

In the TI approach, the reduction potential is calculated by integrating over an arbitrary “coupling parameter” $\varepsilon$ which alchemically moves the system from the initial state to the final state:

$$\Delta G = G_1 - G_0 = \int_0^1 \frac{dG_\varepsilon}{\varepsilon} d\varepsilon$$  \hspace{1cm} (2.34)

where $G_0$ is the Gibbs free energy of the initial (reduced) state and $G_1$ is the Gibbs free energy of the final (oxidized) state.

In the P-B equation approach, the electrostatic potential of the interest molecule in reduced and oxidized states are solved, and reduction potential is the electrostatic energy difference in oxidized and reduced states.

Poisson-Boltzmann equation solves the distribution of electric potential of a solute molecule in dielectric continuum with mobile ions. Starting from the Poisson equation 2.35,

$$\nabla[\varepsilon(r) \nabla \phi(r)] = -\rho(r)$$ \hspace{1cm} (2.35)

where $\rho(r)$ is the total charge density, $\phi(r)$ is the electric potential of the total system generated by both solute molecule and mobile charges, $\varepsilon$ is the permittivity. The charge density of the solute molecule is described as $\rho_s(r) = \sum q_i \delta(r - r_i)$, which sums over all atoms of that molecule, each with charge $q_i$ located at $r_i$. The distribution of mobile ion of species $j$, $n_j$, obeys a Boltzmann distribution under the assumption of thermodynamic equilibrium, $n_j(r) = n_j^0 e^{-\frac{q_j \phi(r)}{k_B T}}$, where $n_j^0$ is the mean concentration of ion species $j$ with charge $q_j$, $k_B$ is the Boltzmann constant, $T$ is the temperature. $q_j \phi(r)$ is the work required to move an ion of $q_j$ from $|r = \infty|$ to the point $r$.

Further expressing $\rho(r)$ as the sum of two above-mentioned contributions gives the Poisson-Boltzmann equation,

$$\nabla[\varepsilon(r) \nabla \phi(r)] = -\sum_{j=1}^N q_j n_j(r) - \rho_s(r),$$
\[
= - \sum_{j=1}^{N} q_j n_j^0 e^{-\frac{q_j \phi(r)}{k_B T}} - \sum_i q_i \delta(r - r_i),
\]
(2.36)

In the high temperature (weak coupling) limit, i.e., \(q_j \phi(r) \ll k_B T\), \(n_j(r)\) can be Taylor expanded to first order,

\[
e^{-\frac{q_j \phi(r)}{k_B T}} \approx 1 - \frac{q_j \phi(r)}{k_B T}.
\]
(2.37)

This yields the linearized Poisson-Boltzmann equation,

\[
\nabla [\varepsilon(r) \nabla \phi(r)] = \sum_{j=1}^{N} \frac{n_j^0 q_j^2}{k_B T} \phi(r) - \sum_{j=1}^{N} n_j^0 q_j - \sum_i q_i \delta(r - r_i),
\]
(2.38)

which is also known as the Debye-Huckel equation. The second term on the right hand side vanishes in neutral solution systems.

Define the Debye-Huckel length \(\lambda_D = \left(\frac{\varepsilon(r) k_B T}{\sum_{j=1}^{N} n_j^0 q_j^2}\right)\) as a characteristic length scale, which describes the persistence of charge carrier’s electrostatic effect. For an electrically neutral system, the linearized Poisson-Boltzmann equation can be concisely formulated as

\[
\nabla^2 \phi(r) = \lambda_D^{-2} \phi(r) - \frac{\rho_s(r)}{\varepsilon(r)}
\]
(2.39)

\[\text{Figure 2.3: Two-dimentional view of the three-dimensional Debye-Huckel model. Figure adapted from Figure 1.1 in Ref. [122]. © Michael J. Holst.}\]

A typical system can often be divided into three regions with different electric characters, as illustrated in Figure 2.3. Region I represents the solute molecule with dielectric constant \(\varepsilon_1\), region II is a thin solvent layer without any mobile ion and with solvent dielectric \(\varepsilon_2\), region III is the bulk solvent region which consists of solvent with dielectric constant \(\varepsilon_2\), assumed to contain
mobile ions. The linearized Poisson-Boltzmann equation has to be solved in all three regions applying continuous boundary conditions of $\phi(r)$ at the interfaces of the regions.

Equation 2.39 is usually solved numerically in an iterative, self-consistent manner.

Compared to TI, P-B equation allows fast estimation of the reduction potential with a controllable convergence criteria. However, the major drawback of this method is that it only takes into account one representative static configuration of the interest molecule. On the other hand, TI simulations often suffer from poor convergence [123]. Whenever available, microscopic experimental $\Delta A$ is used instead.

### 2.4 Chemical master equation for macroscopic properties

The electron transfer rates between two redox sites are microscopic properties that are not experimental observables in general. However, chemical master equations can translate the microscopic rates to macroscopic observables, such as electron flux, in a simple, mean-field way.

#### 2.4.1 Steady-state flux

The steady-state flux $J$ of the system is evaluated by inserting the ET rates into a set of rate equations for electron transfer between neighbouring sites (also referred to as chemical master equations):

$$J = J_{i+1,i} = k_{i+1,i}P_i(1 - P_{i+1}) - k_{i,i+1}P_{i+1}(1 - P_i), \quad i = 1, 2, 3, \ldots N \quad (2.40)$$

$k_{i+1,i}$ and $J_{i+1,i}$ are the ET rate and electron flux from site $i$ to site $i+1$, $P_i$ is the electron occupation probability at site $i$. The term $(1 - P_i)$ takes into account the fact that each heme site can only hold a maximum of one excess electron [37]. Assuming unlimited electron supply at source in and sink out, and fast electron injection and ejection at the terminal sites, the ET at two terminal steps only depend on the probability of the terminal sites themselves (i.e, no dependence on the probability of source in or sink out),

$$J_{1,\text{in}} = k_{1,\text{in}}(1 - P_1) - k_{\text{in},1}P_1, \quad (2.41)$$

$$J_{\text{out},N} = k_{\text{out},N}P_N - k_{N,\text{out}}(1 - P_N). \quad (2.42)$$

In steady-state condition, i.e., $J_{i+1,i} = J_{1,\text{in}} = J_{\text{out},N}$ for all $i$, the set of N equations can be solved either through an iterative approach or analytically. The steady-state flux $J$ and the electron occupation probability $P_i$ of each site $i$ are the variables to solve for, while all the ET
rate constants are either computed (e.g., \(k_{ji}\)) or given (e.g., \(k_{1,\text{in}}\)). By equating equations 2.40 and 2.41 (with a carefully chosen initial guess \(P_1\)), an iterative approach allows one to express the electron occupation probability \(P_{i+1}\) recursively as a function of \(P_1\) and \(P_i\),

\[
P_{i+1} = \frac{k_{i+1,i}P_i - k_{1,\text{in}}(1 - P_1) + k_{\text{in},1}P_1}{[k_{i,i+1} + P_i(k_{i+1,i} - k_{i,i+1})]},
\]

(2.43)

and use \(J_{\text{out},N}\) from equation 2.42 to update \(J_{1,\text{in}}\) and \(P_1\) until certain convergence criteria \(|J_{\text{out},N} - J_{1,\text{in}}| < \text{threshold}\) has been fulfilled. Alternatively, the electron occupation probability can be expressed as a function of steady-state flux \(J\), by expressing \(P_1\) as a function of \(J\),

\[
P_1 = \frac{k_{1,\text{in}} - J}{k_{1,\text{in}} + k_{\text{in},1}}.
\]

(2.44)

Inserting in equation 2.43 gives \(P_2\) as a function of \(J\), and recursively inserting \(P_1\) in equation 2.43 gives \(P_N\) as a function of \(J\), which can be inserted in equation 2.42 to build a polynomial equation of \(J\), giving an analytical solution (please refer to Appendix B for an example using the analytical approach).

2.4.1.1 Maximum system-limited steady-state flux

Using the further assumption of irreversible electron flux in and out of the system at terminal sites (i.e., \(k_{\text{in},1},k_{N,\text{out}} = 0\)), equations 2.41, 2.42, 2.43 can be reduced to

\[
J_{1,\text{in}} = k_{1,\text{in}}(1 - P_1),
\]

(2.45)

\[
J_{\text{out},N} = k_{\text{out},N}P_N.
\]

(2.46)

\[
P_{i+1} = \frac{k_{i+1,i}P_i - k_{1,\text{in}}(1 - P_1) + P_i(k_{i+1,i} - k_{i,i+1})}{[k_{i,i+1} + P_i(k_{i+1,i} - k_{i,i+1})]},
\]

(2.47)

To solve the equations, \(k_{1,\text{in}}\) and \(k_{\text{out},N}\) are set to values that are much faster than any ET rate constant within the system, \(k_{1,\text{in}},k_{\text{out},N} >> k_{ji} \forall i, j\). This situation corresponds to the maximum, protein-limited steady-state flux, \(J_{\text{max}}\), which is expected to be slightly slower than the rate-determinant ET step in a certain direction.

2.4.2 Relaxation dynamics

In non-equilibrium situations, the original master equation formalism is adopted, i.e., solving \(\frac{dP_i}{dt}\) at each site \(i\) instead of the universal steady-state flux \(J\). This, in general, has to be evaluated numerically.
If there is only one excess electron in the whole system, such as in the case of pump-probe experiments, the chemical master equation becomes a first order linear equation:

\[
\frac{dP_i}{dt} = -(\sum_j k_{ji})P_i + \sum_j k_{ij}P_j.
\] (2.48)

Equation 2.48 can be solved either analytically or numerically.
2.5 Typical workflow

Figure 2.4: Diagram of typical workflow for electron transfer studies in multi-heme proteins. For each ET step, molecular dynamics (MD) simulations are carried out to sample initial and final ET states. The MD trajectories are used to compute the reorganisation free energies with equation 2.29 or 2.30 as introduced in subsection 2.3.2. A subset of snapshots from the MD trajectories is extracted to compute the thermodynamic average of electronic coupling matrix element between that heme pair with either the POD or sFODFT method as introduced in subsection 2.3.1. The reduction potentials and hence the driving forces of the redox sites can be obtained either from simulations (e.g., thermodynamic integration, or solving Poisson-Boltzmann equations) or from experiments. These three ET parameters (i.e., reorganisation free energy, electronic coupling matrix element, driving force) are inserted into the non-adiabatic Marcus ET formula (equation 2.18). Finally, the microscopic ET rates for all steps are used together to model the macroscopic electron flux through a certain heme chain in the multi-heme protein with chemical master equations, as introduced in section 2.4.
Chapter 3

Electron Transfer in Wild-Type Proteins STC, MtrC, MtrF

In this chapter, I will report our simulation procedure and our computed ET parameters for the three multi-heme proteins: STC, MtrC, MtrF. I will start with the calculation of electronic coupling matrix elements, where different QM models have been investigated (subsection 3.1.1). The calculations for the other two ET parameters, reorganisation free energies and driving forces, will be reported in section 3.2. Using these computed ET parameters, I will discuss our predicted maximum, protein-limited electron flux for all possible electron flow directions in the three proteins (section 3.4). I will then move on with our modellings of the experimental $I-V$ curves for the multi-heme proteins (section 3.5), and briefly discuss the more complicated electron transport mechanism in protein junctions.

3.1 Electronic couplings

3.1.1 Side chain enhancement of electronic couplings

Background. In the previous calculations carried out in our lab [37], the computed $I-V$ curve fell short of reproducing the approximately nano-Ampere currents reported in STM measurements [29]. Even after accounting for partial protein hydration in these experiments, the computed STM currents remained underestimated by about two orders of magnitude.

This discrepancy has motivated us to take a closer look at the staggered cross heme motif built into MtrC and MtrF. The trifurcation of the electron flow is established by two junctions in the middle of the protein comprised of T-shaped (8-6, 1-3) and co-planar heme pairs (1-6, 6-7, 1-2). Inspection of the crystal structure reveals relatively large heme edge-to-edge distances (defined by the minimum distance between any two heavy atoms in the porphin ring) in these motifs, suggesting that the ET steps across the junctions may limit the overall electron flow
Chapter 3. Electron Transfer in Wild-Type Proteins STC, MtrC, MtrF

through the protein. Indeed, the co-planar heme pair 1-6 in the middle was estimated to be the rate-determining ET step in protein-limited electron flux of MtrF [37]. While the heme edge-to-edge distances of stacked pairs on the two terminals of the octa-heme chain are around 4 Å, the edge separations of the T-shaped and co-planar pairs could reach as much as 6 Å or even more. According to the pathway model of Beratan and Onuchic, tunnelling is assumed to be mediated by consecutive electronic interactions between atoms connecting donor with acceptor [124,125]. Electronic coupling along a given pathway connecting donor and acceptor is written as a product of a (hypothetical) closest contact term, \( H_{ab}^0 \), times an attenuation factor \( \epsilon \) that is a product of decay factors for consecutive tunnelling across covalent bonds (\( \epsilon_{ci} \)), hydrogen bonds (\( \epsilon_{hb}^{ij} \)) or vacuum (\( \epsilon_{ts}^{ij} \)), ts for ‘through space’,

\[
H_{ab} = H_{ab}^0 \epsilon
\]

\[
\epsilon = \prod_i \epsilon_{ci}^{ij} \prod_j \epsilon_{hb}^{ij} \prod_k \epsilon_{ts}^{ij}
\]

where \( \epsilon_{ci} = 0.6 \), \( \epsilon_{hb}^{ij} = 0.36 \exp[-1.7(r/\AA - 2.8)] \) and \( \epsilon_{ts}^{ij} = 0.6 \exp[-1.7(r/\AA - 1.4)] \) and \( r \) the distance between heavy atoms. Unlike in the partially stacked pairs where the most efficient tunnelling pathway is probably through-space direct tunnelling, the tunnelling pathways in T-shaped and co-planar heme pairs are less clear: there are some side chains (either from the porphyrin substituents or from other backbone amino acids) in the space between donor and acceptor hemes, suggesting that the actual tunnelling pathway may be combination of more than one mechanism.

These side chains have never been considered in previous calculations of electronic coupling matrix elements between the heme pairs, which calls for a revisit on this parameter, especially in the context of the persisting mismatch between computation and STM-measured currents [29]. The electronic coupling has been paid much attention because the square of this parameter appears in the rates expression, so the estimated ET rate is fairly sensitive to it. To this end, the impact of the heme side chains on electronic coupling is investigated by successively increasing the size of the quantum-mechanical (QM) model of the heme cofactor. The systematic tests have been carried out with all the three heme pairs in STC (i.e., T-shaped pairs 1-2, 3-4, and stacked pair 2-3), as well as the co-planar heme pair 1-6 in MtrF, covering all three types of packing motifs. In the following I will mainly focus on the T-shaped heme pair 3-4 in STC as an example to discuss the computational procedures and our findings in detail.

**Computational details.** The electronic coupling matrix elements (\( H_{ab} \)) were calculated using the projector operator-based diabatization (POD) approach [106,115] implemented in CP2K [90]
3.1. Electronic couplings

as well as the fragment-orbital density functional theory (FODFT) method implemented in the plane-wave code CPMD [91]. For the POD calculations, the KS states were calculated with PBE [79], PBE0 [80] and a modified PBE functional where 50% of GGA exchange is replaced with Hartree-Fock exchange (denoted as PBE50 in the following). A previous benchmark study on the HAB11 database has shown that POD in combination with PBE50 gives excellent agreement with high-level ab-initio calculations with a small mean unsigned relative error of 9% [106, 114]. Valence electronic states were expanded in a DZVP basis set, core electrons were replaced by GTH atomic pseudopotentials [126]. The auxiliary density matrix method (ADMM) [127] was applied to speed up calculations involving Hartree-Fock exchange (HFX). All the calculations were done in a supercell, treating the long-range interactions according to Martyna and Tuckerman [128]. The wave functions were optimized for a given model of the low-spin Fe$^{2+}$-heme Fe$^{3+}$-heme dimer in vacuum on geometries taken from the 50 ns MD production runs where this pair of hemes was modelled half-reduced with the other hemes oxidized. Calculations were carried out for the doublet ground state, applying a tight convergence criterion of $10^{-7}$ a.u. Thereafter, POD was applied resulting in a matrix $\bar{H}$ for the spin-up (majority spin) and spin-down (minority spin) electrons. The energy levels of the two diabatized fragments were filled with electrons in ascending order, determining the oxidation state of each fragment (i.e., donor and acceptor are thus defined after diabatization for each snapshot). The electronic coupling is independent of whether the excess charge carrier is localized on donor or acceptor fragment, as it should be. For Fe$^{2+}$-hemes, we find that the minority spin HOMO, HOMO-1 and HOMO-2 are quasi-degenerate in each snapshot, lying within 2-3 $k_BT$ at room temperature. Similarly for Fe$^{3+}$-hemes, the LUMO, HOMO and HOMO-1 are quasi-degenerate. These six frontier orbitals, denoted in the following $d^X_i$, $i=1,2,3$, $X=D$ or $A$, are composed of the Fe $d$ orbitals, $d_{xz}$, $d_{yz}$ and $d_{xy}$ and heme ring orbitals. As they are quasi-degenerate, the coupling of the $3 \times 3$ orbital pairs was averaged to an effective electronic coupling matrix element $\langle H_{ab} \rangle = |1/9 \sum_{i,j=1,2,3} |\langle d^A_i|\bar{H}^{KS}|d^D_j \rangle|^2|^{1/2}$. For a certain heme pair, the electronic couplings were square averaged from 25 snapshots extracted along the 50-ns MD trajectory at an interval of 2 ns. The MD trajectory approximates the transition state for ET between the heme pair, with the charges of the two heme cofactors in question set to the averages of the charges for the reduced and oxidized state, with all other hemes modelled in the oxidized state. These transition-state MD trajectories were generated by previous members in our lab. All the coupling calculations were carried out in vacuum as the effect of the extended protein environment on couplings was


shown to be negligibly small in previous QM/MM coupling calculations at the FODFT level [37]. The resultant couplings are denoted as POD/PBE50.

![Figure 3.1: Frontier orbitals of a typical snapshot in DFT calculations. The upper panel (A,B,C) shows the HOMO, HOMO-1, HOMO-2 computed with FODFT and the lower panel (D,E,F) shows the corresponding orbitals computed with POD. The QM model includes all side chains (except for the propionates) of the porphyrin rings (i.e., model 6 for MtrC and MtrF, see Table 3.1 and Figures 3.3 and 3.6 for model description). The orbitals are plotted with isosurface values of ±0.02. The heme cofactors are shown with stick representation (Fe, pink; S, yellow; O, red; N, blue; C, cyan).]

For the heme pairs in STC, the FODFT coupling calculations were carried out on the same heme models and geometries investigated with the POD method with functional PBE [79]. The core electrons of second row atoms were treated with GTH valence pseudopotentials. For Fe a 16-electron GTH semi-core pseudopotential was used. Calculations were carried out for the isolated system (no periodic boundary conditions). I first tested the size of the unit cell and the plane wave cutoff needed to converge the coupling calculations with PBE for the smallest model 1. The test results suggest that a distance of 4.5 Å between any atom and the closest box edge is sufficient to converge $H_{\text{ab}}$ to less than ± 0.05 meV and a cutoff of 160 Ry is sufficient to converge the coupling to less than ± 0.1 meV. These parameters were used in all FODFT coupling calculations reported. The wave functions were optimized so that the gradient was below $10^{-6}$ Ha. As in POD, the orbital order was checked for each MD snapshot. Typically, the two (quasi-)degenerate $d_\pi$ orbitals composed of Fe $d_{xz}$ and $d_{yz}$ and heme ring orbitals, denoted $d_{\pi,1}$ and $d_{\pi,2}$, are HOMO-1 and HOMO-2 in the FODFT/PBE calculations, whereas the HOMO is of
3.1. Electronic couplings

$d_{xy}$ character (see Figure 3.1 (A, B, C)). The effective coupling in this method is obtained from the $2 \times 2$ $d_{x^2-y^2}$ orbital pairs, $|H_{ab}| = \frac{1}{4} \sum_{i,j=1,2} \left| \langle d_{A,i}^{\pi} | H_{KS}^\dagger | d_{B,j}^{\pi} \rangle \right|^2$ \frac{1}{2}$. This choice was motivated by magnetic spectroscopy experiments showing that the spin density of a single oxidized heme is composed of 2 quasi-degenerate $d_{x^2-y^2}$ orbitals [129]. Note that our definition of the effective coupling in FODFT was different from the one used in POD. In POD the $d$-orbital manifold no longer clearly splits into $d_{x^2-y^2}$ and $d_{||}$ orbitals, but is mixed instead (see Figure 3.1 (D, E, F)). This is a consequence of the different diabatization method in FODFT and POD. The former method uses orbitals obtained for the isolated donor and acceptor in vacuum whereas the latter starts from orbitals delocalized over both donor and acceptor. A benchmark study of FODFT/PBE against high-level ab-initio calculations showed that the couplings in this method are systematically underestimated by a factor of 1.348 [114] due to the missing electronic polarization between donor and acceptor. The raw coupling values obtained with FODFT were multiplied by this scaling factor. The resultant couplings are denoted as sFODFT/PBE (“s” for scaled). Again for a certain heme pair, the electronic couplings were square averaged from the same 25 snapshots in sFODFT/PBE calculations. The plane-wave calculations become extremely expensive at hybrid functional level of theory. A typical PBE0 [80] calculation restarted from a PBE calculation takes less than 200 SCF cycles to converge. However, each SCF cycle is 10 times more expensive when computed at PBE0 level (i.e., around 20 seconds with 768 cores) than at PBE level (i.e., around 8 seconds with 192 cores). Therefore, FODFT calculations at functional PBE0 [80] level were only carried out with model 1 and the final model on a single test snapshot for STC.

The electronic coupling calculations for MtrF and MtrC were carried out with POD/PBE50 using CP2K [90]. CP2K uses Gaussian basis set, and plane waves as auxiliary basis [130], which allows efficient DFT calculations at the hybrid functional level of theory.

Results. I will first focus on the T-shaped heme pair 3-4 in STC. Different QM models that have been investigated for the electronic coupling calculations in STC are shown in Figure 3.2 and listed in Table 3.1. The simplest heme model 1, Fe-porphin axially ligated by two methyl-imidazoles and all heme side chains replaced with hydrogen atoms, gives rather small couplings of less than 1 meV, see Figure 3.4 C (blue solid lines). Importantly, upon inclusion of the side chains inserted in the space between the two heme groups (model 2), we obtain a significant increase in electronic coupling by a factor of 3.2. Further additions of side chains pointing away from the electron transfer partner (models 3 and 4), have only a minor effect. The coupling for the largest model investigated (model 5), comprised of all side chains including the propionates,
### Table 3.1: Computational models for electronic coupling calculations with the test heme pairs $i-j$ in STC and MtrF. $R_{6i}$ denote substituents of the porphyrin ring at positions indicated in Figure 3.3: S1: -CH$_3$, S2: -CH$_2$CH$_3$, S3: -CH$_3$CHSCH$_3$, S4: -CH$_2$CH$_2$COOH. Models marked with an asterisk contain side chains that are not covalently bound to the hemes inserted in the space between heme pairs: heme 1-2, model 5: residues Ser6, Ser17, Ala63 terminated at CB-CA bond with H; heme 3-4, model 3: residues Val48, Leu56 terminated at CB-CA bond with H, residue Pro73 terminated at CB-CA, CD-N bond with H. The axial histidines are modelled as methyl-imidazoles.

#### STC, $i,j = 1,2$ (T-shaped)

<table>
<thead>
<tr>
<th>Model</th>
<th>number of atoms in $i,j$</th>
<th>$R_{1i}$</th>
<th>$R_{2i}$</th>
<th>$R_{3i}$</th>
<th>$R_{4i}$</th>
<th>$R_{5i}$</th>
<th>$R_{6i}$</th>
<th>$R_{7i}$</th>
<th>$R_{8i}$</th>
<th>$R_{1j}$</th>
<th>$R_{2j}$</th>
<th>$R_{3j}$</th>
<th>$R_{4j}$</th>
<th>$R_{5j}$</th>
<th>$R_{6j}$</th>
<th>$R_{7j}$</th>
<th>$R_{8j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61, 61</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>73, 73</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S2</td>
<td>S1</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S2</td>
<td>S1</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S2</td>
<td>S1</td>
</tr>
<tr>
<td>3</td>
<td>77, 73</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S2</td>
<td>S1</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S2</td>
<td>S1</td>
</tr>
<tr>
<td>4</td>
<td>77, 77</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
</tr>
<tr>
<td>5*</td>
<td>94, 83</td>
<td>S4</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S4</td>
<td>S4</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S4</td>
<td>S4</td>
</tr>
<tr>
<td>6</td>
<td>111, 111</td>
<td>S4</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S4</td>
<td>S4</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S4</td>
<td>S4</td>
</tr>
</tbody>
</table>

#### STC, $i,j = 2,3$ (stacked)

<table>
<thead>
<tr>
<th>Model</th>
<th>number of atoms in $i,j$</th>
<th>$R_{1i}$</th>
<th>$R_{2i}$</th>
<th>$R_{3i}$</th>
<th>$R_{4i}$</th>
<th>$R_{5i}$</th>
<th>$R_{6i}$</th>
<th>$R_{7i}$</th>
<th>$R_{8i}$</th>
<th>$R_{1j}$</th>
<th>$R_{2j}$</th>
<th>$R_{3j}$</th>
<th>$R_{4j}$</th>
<th>$R_{5j}$</th>
<th>$R_{6j}$</th>
<th>$R_{7j}$</th>
<th>$R_{8j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61, 61</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>67, 67</td>
<td>H</td>
<td>S1</td>
<td>S1</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S1</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S1</td>
<td>H</td>
</tr>
<tr>
<td>3</td>
<td>79, 79</td>
<td>S2</td>
<td>S1</td>
<td>S1</td>
<td>S2</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S2</td>
<td>S1</td>
<td>S1</td>
<td>S2</td>
<td>H</td>
<td>H</td>
<td>S2</td>
<td>S1</td>
</tr>
<tr>
<td>4</td>
<td>86, 86</td>
<td>S4</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S4</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>H</td>
<td>H</td>
<td>S4</td>
<td>S1</td>
</tr>
<tr>
<td>5</td>
<td>107, 107</td>
<td>S4</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S4</td>
<td>S4</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S4</td>
<td>S4</td>
</tr>
</tbody>
</table>

#### STC, $i,j = 3,4$ (T-shaped)

<table>
<thead>
<tr>
<th>Model</th>
<th>number of atoms in $i,j$</th>
<th>$R_{1i}$</th>
<th>$R_{2i}$</th>
<th>$R_{3i}$</th>
<th>$R_{4i}$</th>
<th>$R_{5i}$</th>
<th>$R_{6i}$</th>
<th>$R_{7i}$</th>
<th>$R_{8i}$</th>
<th>$R_{1j}$</th>
<th>$R_{2j}$</th>
<th>$R_{3j}$</th>
<th>$R_{4j}$</th>
<th>$R_{5j}$</th>
<th>$R_{6j}$</th>
<th>$R_{7j}$</th>
<th>$R_{8j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61, 61</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>77, 70</td>
<td>S2</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S2</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S2</td>
<td>H</td>
</tr>
<tr>
<td>3</td>
<td>74, 70</td>
<td>S3</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S3</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S3</td>
<td>H</td>
</tr>
<tr>
<td>4</td>
<td>74, 74</td>
<td>S3</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S3</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S3</td>
<td>H</td>
</tr>
<tr>
<td>5</td>
<td>80, 80</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
</tr>
<tr>
<td>6</td>
<td>111, 111</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S4</td>
<td>S4</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
</tr>
<tr>
<td>7</td>
<td>111, 111</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S4</td>
<td>S4</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
</tr>
</tbody>
</table>

#### MtrF, $i,j = 1,6$ (co-planar)

<table>
<thead>
<tr>
<th>Model</th>
<th>number of atoms in $i,j$</th>
<th>$R_{1i}$</th>
<th>$R_{2i}$</th>
<th>$R_{3i}$</th>
<th>$R_{4i}$</th>
<th>$R_{5i}$</th>
<th>$R_{6i}$</th>
<th>$R_{7i}$</th>
<th>$R_{8i}$</th>
<th>$R_{1j}$</th>
<th>$R_{2j}$</th>
<th>$R_{3j}$</th>
<th>$R_{4j}$</th>
<th>$R_{5j}$</th>
<th>$R_{6j}$</th>
<th>$R_{7j}$</th>
<th>$R_{8j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61, 61</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>70, 70</td>
<td>S2</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S2</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S2</td>
<td>H</td>
</tr>
<tr>
<td>3</td>
<td>74, 70</td>
<td>S3</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S3</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S3</td>
<td>H</td>
</tr>
<tr>
<td>4</td>
<td>74, 70</td>
<td>S3</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S3</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>S1</td>
<td>S3</td>
<td>H</td>
</tr>
<tr>
<td>5</td>
<td>80, 80</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
</tr>
<tr>
<td>6</td>
<td>111, 111</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S4</td>
<td>S4</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
<td>S1</td>
<td>S3</td>
<td>S1</td>
</tr>
</tbody>
</table>

The number of connections differs by less than 2% compared to model 2 indicating that coupling is well converged with respect to system size for model 2. A similar trend is observed for heme pair 1-2 in STC, which forms a T-shaped motif at the opposite terminus of the protein (Figure 3.4 A). We find that the strong increase in electronic coupling between heme pair 3-4 (1-2) is due to Cys61 (Cys18). Figure 3.5 A shows one of the frontier orbital pair combinations that contribute
3.1. Electronic couplings

Figure 3.2: Schematic illustration of the computational models for heme-heme electronic coupling calculations of protein STC as described in Table 3.1 and Figure 3.3. The smallest models are shown in grey while atoms of the additional substituents are drawn in green (C), red (O) and yellow (S). Hydrogen atoms are not shown for clarity. This figure was prepared by Zdenek Futera. Reprint with permission.

to the electronic coupling in pair 3-4. One can clearly see from Figure 3.5 A that the sulfur 3$p$ orbital of Cys61 mixes with the Fe-heme frontier orbital of heme 3, though relatively weakly as the density of states of sulfur peaks more than 1 eV below the Fe-3$d$ states. Nonetheless,
Figure 3.3: Position of substituents $R_i, i = 1, \ldots, 8$ in the heme models defined in Table 3.1. The axial histidines (not shown) are modelled as methyl-imidazoles.

Table 3.2: Square-averaged heme-heme electronic coupling matrix element $\langle |H_{ij}|^2 \rangle^{1/2}$ for STC. The error bars are calculated from the root-mean-square fluctuations of the electronic coupling matrix elements. Models used are detailed in Table 3.1 and Figures 3.2-3.3. In model 1 all heme ring substituents including the cysteine linkages are replaced by H, in the “final model” all relevant cysteine linkages are included. “Final model” refers to model 3 for heme pair 1-2, model 2 for pair 2-3 and model 2 for 3-4. All energies are given in meV.

| heme pair $i-j$ | $\langle |H_{ij}|^2 \rangle^{1/2}$, model 1 | $\langle |H_{ij}|^2 \rangle^{1/2}$, final model |
|----------------|---------------------------------|---------------------------------|
|                | POD/PBE50 | sFODFT/PBE | POD/PBE50 | sFODFT/PBE |
| 1-2            | 0.30 ± 0.07 | 0.31 ± 0.07 | 2.17 ± 0.49 | 1.37 ± 0.25 |
| 2-3            | 2.74 ± 0.52 | 3.30 ± 0.80 | 3.08 ± 0.68 | 3.95 ± 1.08 |
| 3-4            | 0.49 ± 0.10 | 0.52 ± 0.05 | 2.08 ± 0.37 | 2.11 ± 0.34 |

the mixing is sufficiently strong to modify the tail of the Fe-heme frontier orbitals leading to an increase in overlap with the Fe-heme frontier orbitals of heme 4 and consequently to an increase in electronic coupling. By contrast, no cysteine-mediated increase in coupling is observed for heme pair 2-3 in the middle of the protein. The two heme rings form a stacked motif and approach one another at Van der Waals distance with the cysteine linkages oriented nearly perpendicular to the heme planes.

Our observation that the cysteine linkages increase the electronic couplings in heme pairs
3.1. Electronic couplings

Figure 3.4: $|H_{ab}|$ values for heme pair 1-2 (A, D, G), 2-3 (B, E, H), 3-4 (C, F, I) of protein STC. (A, B, C) Convergence with respect to the model size, as obtained with POD and FODFT methods for different density functional theory (DFT) functionals. PBE50 means 50% of PBE exchange is replaced by the exact Hartree-Fock exchange. In (A, C), the sharp increase in $|H_{ab}|$ from 2 (1) to 3 (2) is due to inclusion of the cysteine linkages in model 3 (2). The bold arrow indicates the final model chosen for presented rate calculations. (D, E, F) Thermal fluctuations of $|H_{ab}|$ at POD/PBE50 level along a MD trajectory using model 1 (green) and the final model (blue). Dashed lines indicate the accumulated average $\langle |H_{ab}|^2 \rangle^{1/2}$. (G, H, I) Corresponding thermal fluctuations and the accumulated average of $|H_{ab}|$ computed at FODFT/PBE level with the same snapshots.

3-4 and 1-2 is robust with respect to (i) the fraction of Hartree-Fock exchange (HFX) used in the exchange-correlation functional (ii) the method used for electronic coupling calculation (iii) thermal protein fluctuations. While absolute couplings are sensitive to the fraction of HFX used, a strong increase upon inclusion of cysteine linkages in the model is observed in all calculations (see Figure 3.4 A, C, blue lines). A similar trend could be found with FODFT method (Figure 3.4 A, C, red lines). In fact, the two methods give very similar results in the current application (see Figure 3.4 and Table 3.2). To take into account thermal fluctuations the $H_{ab}$ is sampled along classical molecular dynamics (MD) trajectories of the protein at 300 K. We find that the thermal average, $\langle |H_{ab}|^2 \rangle^{1/2}$, for heme pair 3-4 is a factor of 4.2 larger when the cysteine
Figure 3.5: Crystal structure of proteins STC (A) and MtrC (B) with the heme cofactors highlighted. The insets depict the enhancement of electronic couplings due to cysteine linkages (marked by arrows) inserted in the space between T-shaped heme pair 3-4 in STC (A), co-planar heme pair 6-1 in MtrC (B), and T-shaped heme pair 8-6 in MtrC (B). Isosurfaces of the Fe-heme frontier orbital pairs mediating the electron transfer are superimposed. Colour code of atoms: Fe: pink, S: yellow, O: red, N: blue, C: green. The protein secondary structure is depicted in grey.

linkages are included (see Figure 3.4 F), in line with the result obtained above for a single configuration. The corresponding increase in coupling for the other T-shaped heme pair 1-2, a factor of 7.3, is even more pronounced than for heme pair 3-4. The stronger effect is consistent with the average heme edge-to-edge separation being larger and the average sulfur-to-heme edge distance being smaller in heme pair 1-2 than in 3-4 (see Table 3.3).

Similar side chain enhancement effects have been found for the co-planar heme pair 1-6 of MtrF with POD/PBE50. Altogether, 7 QM models have been investigated for this pair, as shown in Figure 3.6 and listed in Table 3.1. Interestingly, the coupling does not change much upon
### Table 3.3: Distances (in Å) between heme pairs in STC.

$r_e$ stands for the heme edge-to-edge distance, $r_{S-e}$ stands for the distance between the sulfur atom of the intervening cysteine linkage and the edge of the neighbouring heme from the MD average and in the crystal structure, $r_h$ stands for the heavy-atom distance, which are defined as the shortest distance between any heavy atom (C, N, O, S) of the porphin ring plus side chains. “MD” values are taken from the mean value averaged over the 25 MD snapshots used for coupling calculations. “cryst.” values are calculated from the distances in the crystal structure [55].

<table>
<thead>
<tr>
<th>Heme pair</th>
<th>$r_e$ (MD)</th>
<th>$r_e$ (cryst.)</th>
<th>$r_{S-e}$ (MD)</th>
<th>$r_{S-e}$ (cryst.)</th>
<th>$r_h$ (MD)</th>
<th>$r_h$ (cryst.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>5.93</td>
<td>5.87</td>
<td>3.97</td>
<td>4.01</td>
<td>3.71</td>
<td>3.68</td>
</tr>
<tr>
<td>2-3</td>
<td>4.14</td>
<td>3.91</td>
<td>–</td>
<td>–</td>
<td>3.68</td>
<td>3.69</td>
</tr>
<tr>
<td>3-4</td>
<td>5.82</td>
<td>5.76</td>
<td>4.12</td>
<td>4.15</td>
<td>4.12</td>
<td>3.61</td>
</tr>
</tbody>
</table>

including one cysteine linkage from either the donor or the acceptor (see Figure 3.6 B, model 3, 4). However, there is a sharp increase in $|H_{ab}|$ from 3, 4 to 5 when the two cysteine linkages, one from the donor, one from the acceptor, are present. $|H_{ab}|$ slightly decreases when more side chains are included in the QM model, indicating that the QM model is not yet fully converged at model 5. For consistency, model 6, which has all side chains (except for the propionates) of the porphyrin rings included, is adopted as the final model. The electronic coupling calculations of MtrF and MtrC are carried out with the minimum model (1) and the final large model (6).

All the electronic coupling matrix elements for protein STC, MtrF, MtrC are summarised in Appendix C.

For further discussion, the rate enhancement factor from the DFT calculation is defined as the ratio

$$r_{\text{dft}} = \frac{\langle |H_{ab}^l|^2 \rangle}{\langle |H_{ab}^m|^2 \rangle},$$

(3.3)

where $H_{ab}^l$ and $H_{ab}^m$ are the coupling matrix element for heme-to-heme electron tunnelling for the large (l) and minimum (m) QM models, and $\langle \cdots \rangle$ denotes the thermal average over MD snapshots. In case of the co-planar heme pair 1-6, where the coupling enhancement effect is the greatest ($r_{\text{dft}}^{1/2} = 2500$ for MtrC (900 for MtrF), see Table 3.4), Cys189(Cys197) and Cys499(Cys476) which covalently link hemes 1 and 6 to the protein backbone, approach one another up to a S-S distance of 4.0 (3.8) Å. As in the STC case, the sulfur 3p orbital of Cys189(Cys197) weakly mixes with the Fe-heme frontier orbitals of heme 1, and a similar mixing occurs for Cys499(Cys476) and heme 6 (see Figure 3.5 B). The small delocalization of the frontier orbital over the S atoms leads to a sizable increase in orbital overlap and consequently electronic coupling. Smaller albeit notable coupling enhancements occur for the T-shaped pairs 8-6 ($r_{\text{dft}} = 36$ (64)) and 1-3 ($r_{\text{dft}} = 9$ (16)), where only one cysteine is inserted between the hemes.
3.1.2 Rate enhancement estimation with pathway-tunnelling model

Using the pathway (pw) model of Beratan and Onuchic (equation 3.2) to describe the heme-to-heme electron tunnelling enhancement by the heme side chains (including the cysteine linkages), the rate enhancement factor is defined as the ratio

\[ r_{pw} = \frac{\langle |H_{ab}^{pw}|^2 \rangle}{\langle |H_{ab}^{ts}|^2 \rangle} = \frac{\langle [\langle e^c \rangle^n e^{ts} (r_h)]^2 \rangle}{\langle [e^{ts}(r_e)]^2 \rangle}, \tag{3.4} \]
### Table 3.4: Heme-to-heme tunnelling enhancement factors in MtrC, MtrF and STC from POD/PBE50 (dft) and pathway (pw) calculations, $r_{\text{dft}}$ and $r_{\text{pw}}$, respectively.

<table>
<thead>
<tr>
<th>Heme Pair</th>
<th>Motif</th>
<th>$r_{\text{dft}}$</th>
<th>$r_{\text{pw}}$</th>
<th>$r_{\text{dft}}$</th>
<th>$r_{\text{pw}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-1</td>
<td>Co-planar</td>
<td>2500</td>
<td>2200</td>
<td>870</td>
<td>1500</td>
</tr>
<tr>
<td>1-2</td>
<td>Co-planar</td>
<td>21</td>
<td>170</td>
<td>34</td>
<td>1200</td>
</tr>
<tr>
<td>6-7</td>
<td>Co-planar</td>
<td>2.5</td>
<td>5</td>
<td>26</td>
<td>230</td>
</tr>
<tr>
<td>8-6</td>
<td>T-shaped</td>
<td>38</td>
<td>120</td>
<td>68</td>
<td>120</td>
</tr>
<tr>
<td>1-3</td>
<td>T-shaped</td>
<td>11</td>
<td>170</td>
<td>14</td>
<td>150</td>
</tr>
<tr>
<td>3-4</td>
<td>Stacked</td>
<td>1.7</td>
<td>7.5</td>
<td>8.1</td>
<td>1.8</td>
</tr>
<tr>
<td>9-8</td>
<td>Stacked</td>
<td>7.7</td>
<td>1.7</td>
<td>3.6</td>
<td>2.3</td>
</tr>
<tr>
<td>10-9</td>
<td>Stacked</td>
<td>1.3</td>
<td>1.2</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>4-5</td>
<td>Stacked</td>
<td>1.9</td>
<td>1.0</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>STC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>T-shaped</td>
<td>52</td>
<td>102</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>T-shaped</td>
<td>18</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-3</td>
<td>Stacked</td>
<td>1.3</td>
<td>–</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ equation 3.3, where $\langle |H_{ab}^h|^2 \rangle$ and $\langle |H_{ab}^m|^2 \rangle$ are taken from Appendix C.

$^b$ equation 3.4.

$^c$ $r_{S-c}$ instead of $r_h$ (see definition in Table 3.3) is used, consistent with the QM models used in the DFT calculations for STC.

where $r_h$ is the shortest distance between any heavy atom (C, N, O, S) of the two porphyrin rings plus side chains, $r_e$ is the smallest heme edge-to-heme edge distance, $n$ is the total number of covalent bonds connecting the two heavy atoms defining $r_h$ with their respective porphyrin rings, and $\langle \cdots \rangle$ denotes the thermal MD average over protein structures. The decay factor $\langle \epsilon^n_i \rangle$ on the right hand side of equation 3.4 is due to the covalent bonds from the heme edge to the heavy atom that defines $r_h$. Using the average distances from MD we obtain $r_{\text{pw}} = 42$ and 101 for heme pair 3-4 and 1-2 in STC, respectively, and $r_{\text{pw}} = 2200$ (1500) for heme pair 1-6 in MtrC (MtrF).

The heme-to-heme rate enhancement factor computed by DFT (equation 3.3) and the pathway model (equation 3.4) are listed in Table 3.4 for comparison. For most of the heme pairs, there is good agreement between the pathway model and POD/PBE50 for the enhancement factor (within one order of magnitude). The agreement is excellent for the heme pair 1-6 in the deca-heme proteins which limits the overall electron flow. This implies that the ET may be understood in terms of simple through-bond through-space mediation. A similar pathway-specific mediation effect has been found for heme-to-heme ET in cytochrome c oxidase with a methyl...
group implicated as an essential tunnelling mediator [131]. In this respect it is interesting to note that if we replace the sulfur atom by an oxygen atom or a CH$_2$ group (i.e., cysteine $\rightarrow$ serine, cysteine $\rightarrow$ methylated alanine, see Figure 3.7 B,C), the rate enhancement $r_{\text{dft}}$ reduces from 40 to 14 with DFT calculations. While the pathway model predicts a similar trend for cysteine ($r_{\text{pw}} = 134$) $\rightarrow$ serine ($r_{\text{pw}} = 55$), it fails to predict the decrease for cysteine $\rightarrow$ methylated alanine ($r_{\text{pw}} = 147$) mutation. Hence, our analysis highlights the importance of not only the atomistic but also the chemical details of the tunnelling medium. The latter may be captured by a refined pathway model with chemical specificity.

### 3.1.3 Generalized distance metric for electronic couplings

In the literature, distance separation between donor and acceptor is widely used for the quick estimation of the electronic coupling between a well-separated donor and acceptor pair [5, 125, 132, 133]. For heme pairs, a common choice is to use the edge-to-edge distance separation between donor and acceptor porphyrin rings as the distance metric [134]. The distance dependence of the coupling matrix elements, using this distance metric, are shown in Figure 3.8 A and B. The data for the minimum model (panel A) shows the expected exponential decay with respect to the heme edge-to-heme edge distance $r$, $\langle |H_{ab}|^2 \rangle^{1/2} = A \exp[-\beta (r - r_0)/2]$. The scatter around the mean is due to thermal motion of the heme rings ($T = 300$ K). The couplings decrease in the order stacked $>$ T-shaped $>$ co-planar heme-heme motif. In the stacked motif, the hemes approach one another up to Van der Waals distance (3.5–5.0 Å) resulting in couplings of several meVs, whereas in the T-shaped and co-planar motifs the edge-to-edge distances are larger (5.0–8.0 Å), and the couplings are an order of magnitude smaller, typically a few $0.1$ meV or less. The distance decay constant $\beta$ and the prefactor $A$ are determined to be $2.26$ Å$^{-1}$ and $3.49$
3.1. Electronic couplings

meV, respectively \((R^2 = 0.99, r_0 = 3.6 \text{ Å})\), in good agreement with the ones reported previously using data from MtrF only [37].

Figure 3.8: The distance dependence of heme-heme electronic coupling matrix elements, \(|H_{ab}|\), using the data of MtrC, MtrF and STC. The coupling values are computed with the minimum (m) QM model \((H_{ab} = H_{ab}^{(m)})\) in (A) and with the large (l) QM model \((H_{ab} = H_{ab}^{(l)})\) in (B, C) on structures obtained from MD simulation at room temperature. They are colour-coded according to the relative orientations of electron donating and accepting hemes: stacked motif in blue (heme pairs 10-9, 9-8, 3-4, 4-5 in MtrC and MtrF and 2-3 in STC), T-shaped in red (8-6 and 1-3 in MtrC and MtrF and 1-2, 3-4 in STC), and co-planar in green (6-1, 6-7 and 1-2 in MtrC and MtrF). Root-mean-square averages of the scattered data points were calculated for bins of width 0.4 Å (A, B) and 0.2 Å (C) and are denoted by black circles with error bars indicating the root-mean-square fluctuations. Fits to an exponential are indicated by a black line. In (A, B) the shortest heme edge-to-edge distance is used and in (C) the shortest distance between any heavy atom of heme ring and side chains.

Yet, the situation is strikingly different when the heme side chains are included in the coupling calculation. While the values for the stacked motif hardly change, they increase significantly for co-planar and T-shaped motifs, to values that are just slightly below those for the stacked motif. Consequently, all couplings now fall in a rather narrow range of about 0.9-3.5 meV for MtrC (0.7-4.5 meV for MtrF). The spread of coupling values around the best fit for exponential distance decay is very large \((R^2 = 0.45, \text{ Figure 3.8 B})\), suggesting that the heme edge-to-heme edge distance is no longer a good distance metric for heme-to-heme electron tunnelling in our final model including the side chains. Instead, we extended the edge definition of the conjugated heme macrocycle and defined the generalized heavy-atom distance as the shortest distance between any heavy atom (C, N, O, S) of the porphin ring plus side chains. Using this metric, all couplings shift to shorter distances and can be fit to an exponential distance decay with parameters similar to those for the minimum model in Figure 3.8 A, \(\beta = 2.63 \text{ Å}^{-1}, A = 2.57 \text{ meV} (R^2 = 0.97, \text{ Figure 3.8 C})\), which are characteristic for through-space tunnelling. However, the scatter around the mean values is still significantly larger than for the minimum model. This is most likely because the degree of delocalization of the frontier orbitals over the side chains varies more strongly with intramolecular heme geometry than for the unsubstituted porphyrin rings and
this effect is independent on inter-heme distance. The new distance metric yields the minimum distance separation for all ET steps along the octa-heme chains in MtrC and MtrF to be around 4 Å (see Figure 3.9), with the closest separations from: ring edge-to-ring edge for the stacked pairs; one cysteine-to-ring edge for the T-shaped pairs; cysteine-to-cysteine for the co-planar pair 1-6.

Figure 3.9: MD averaged distance separation between each heme pair in MtrC (A) and MtrF (B). Blue squares utilize the conventional distance metric (edge-to-edge distance), black lines utilize the generalized distance metric defined as the shortest distance between any heavy atoms (C, N, S) of the porphin ring plus side chains. The heavy-atom distances for each heme pair along the octa-heme chain (except for the first two data points: 1-2, 6-7) are around 4 Å. $S_1$, $S_2$ represents the sulfur atom of the cysteine linkage attached to the donor and the acceptor heme, respectively. Heme pairs are indicated by their packing motifs on the x-axis: blue, stacked; red, T-shaped; green, co-planar.

It should be pointed out that the electronic couplings calculated here are for through-space tunnelling between adjacent heme cofactors including all side chains and the cysteine linkages of the heme groups. Hence, the $\beta$ values reported here are about a factor of two larger than the typical range for through-protein tunnelling, 1.0-1.5 Å$^{-1}$ [4], whilst the tunnelling distance for each consecutive hop is about a factor of 2 or more smaller than for typical through-protein tunnelling processes. As is well known, over distances of several nanometres and beyond, multi-step hopping outcompetes 1-step tunnelling due to its favourable $1/R$ scaling compared with exponential scaling for 1-step tunnelling [4]. There are amino acid side chains that bridge the gap between co-planar and T-shaped heme pairs (e.g. between hemes 1 and 6, ILE252, LEU571 in MtrC and PRO540, PRO243 in MtrF), implying that amino acid-mediated heme-to-heme tunnelling could be an alternative mechanism. However, using pathway calculations [124,135], we found that the dominant through-space tunnelling path always gave couplings at least an order of magnitude higher than any amino acid-mediated pathway, which rules out this alternative mechanism, at least at the level of pathway calculations.
3.2 Calculation of ET driving forces and reorganisation free energies

The ET driving forces and reorganisation free energies of proteins MtrF, MtrC, STC have been computed by previous members (Dr. Ehesan Ali, Dr. Marian Breuer, Dr. Carlo A. Bortolotti, Mr. Bastian Burger) in our lab.

**Simulation details.** Typically, in the equilibration MD run, the solvated protein in the fully oxidized state was initially minimized over 5000 steps and subsequently equilibrated for 500 ps with all protein atoms kept frozen. The temperature was rescaled to 300 K every 5000 steps and the Langevin barostat was applied with a target pressure of 1 bar. The protein was then slowly released by applying harmonic restraints around the crystallographic positions with force constants of 99, 25, 1.0, 0.1, and 0.001 kcal mol\(^{-1}\) Å\(^{-2}\). The duration of each of these runs was 500 ps, the MD time step was 1 fs, the temperature was rescaled as before and the volume was held constant. Eventually, all position restraints were dropped and the protein was equilibrated for 5 ns using a time step of 2 fs. The Langevin thermostat and barostat were used, with target temperature and pressure of 300 K and 1 bar, respectively.

Driving forces were obtained from heme reduction potential differences in the all-ox state of the protein, as explained in detail in Ref. [38]. For a given heme \(i\), the atomic charges of the heme, \(q_k\), were slowly changed from the values of the oxidized to the values of the reduced state in four equidistant steps, i.e. by \(\Delta q_k = (q_{k}^{\text{red}} - q_{k}^{\text{ox}})/4\), while the charges of the atoms of all other hemes \(j \neq i\) remained unchanged. After each variation in charge, the system was typically equilibrated for 3 ns followed by a production run of 5 ns. The initial structure for each thermodynamic integration window was taken from a production run of a previous window to take advantage of previous equilibration steps. Reduction potentials were obtained from the average energy gap for each window using Simpson integration. Thermodynamic integration runs were also carried out in the reverse direction from the singly reduced to the all-ox state. The difference in the reduction potentials in the forward and reverse direction was usually reasonably small (on average 25 meV) [38], showing that sampling was sufficient.

For each ET step, the reorganisation free energies are obtained from the shift of the mean vertical energy gap in initial and final ET state, using classical MD for thermal sampling and an electronically polarizable force field for calculation of the gap energy. The reorganisation free energy for ET between the heme groups was calculated from the trajectories obtained for the singly reduced hemes using the linear response approximation. For heme pair \(i - j\), the vertical
energy gap $\Delta E(N) = E_{RO}(N) - E_{OR}(N)$ was calculated along the MD trajectory with heme $i$ reduced and all other hemes oxidized, where $E_{RO}(N)$ is the total potential energy of that state (initial ET state) and $E_{OR}(N)$ the total potential energy of the state with heme $j$ reduced and all other hemes oxidized (final ET state). This gave the average vertical energy gap of the initial ET state, $\langle \Delta E \rangle_A$, and the root-mean-square fluctuations $\sigma_A = \sqrt{\langle (\Delta E - \langle \Delta E \rangle_A)^2 \rangle_A}$. Calculations of the energy gap $\Delta E$ along the trajectory with heme $j$ reduced and all other hemes oxidized gave $\langle \Delta E \rangle_B$ and $\sigma_B$. The reorganisation free energies for ET from heme $i$ to $j$ were then obtained as $\lambda^{st} = (\langle \Delta E \rangle_A - \langle \Delta E \rangle_B)/2$ and $\lambda^{var} = (\sigma_A^2 + \sigma_B^2)/(4k_BT)$. The AMBER03 classical force field [59] and TIP3P water model [136] were used for thermal sampling, and the electronically polarizable AMBER02 force field [59] and POL3 water model [75] were used in the calculation of the energy gap. The inner-sphere contribution due to the force field was replaced by a QM estimate as reported previously (50 meV) [137]. The resultant reorganisation free energies $\lambda^{st}$ (“st” for Stokes shift, summarized in Appendix C) fall in the range 0.7-1.1 eV in accord with expectations for proteins with partly solvent-exposed cofactors [4].

3.3 $H_{ab}$, $\lambda$, $\Delta A$: which is the most sensitive ET parameter?

The heme-heme ET rate constants, estimated by the three Marcus ET parameters, turn out to span a wide range from $10^4$ to $10^9$ s$^{-1}$. However, it is not clear how sensitive the computed ET rate constant is with respect to the three ET parameters – electronic coupling matrix element, reorganisation free energy, driving force. In this section I will report a detailed investigation by varying each parameter separately in the range that it spans while keeping the other two parameters fixed at their mean values.

In the three multi-heme proteins studied herein – STC, MtrC, MtrF – the electronic coupling matrix elements are between 0.70 and 4.51 meV; the reorganisation free energies are between 0.68 and 1.13 eV; the driving forces are between -0.25 and 0.19 eV. The three ET parameters are independent from one another, correlation coefficients smaller than 0.4 for any two of the three parameters. The forward ET rate constant computed with the mean of these three ET parameters (see the first row in Table 3.5), is $4 \times 10^7$ s$^{-1}$.

**Sensitivity to the electronic coupling matrix element $H_{ab}$.** The reorganisation free energy and the driving force are fixed at their mean values: $\lambda_0 = 0.87$ eV, $\Delta A_0 = -0.04$ eV. The forward ET rate constants are calculated at different electronic coupling values, as summarised in Table 3.6, columns 1,2. Since the square of $|H_{ab}|$ appears in the Marcus ET rate expression (equation 2.18), the 6-fold increase of $|H_{ab}|$ from 0.70 to 4.51 meV results in 40-fold increase in the ET rate.
3.3. $H_{ab}, \lambda, \Delta A$: which is the most sensitive ET parameter?

constants, spanning two orders of magnitude from $10^6$ to $10^8$ s$^{-1}$. Since the uncertainty of $|H_{ab}|$ is usually one order of magnitude smaller than $|H_{ab}|$ itself (see Ref. [138], Table S2), the uncertainty of the estimated ET rate constant yielding from the electronic coupling will be within one order of magnitude.

**Sensitivity to the reorganisation free energy** $\lambda$. The electronic coupling matrix element and the driving force are fixed at their averaged values: $|H_{ab}|_0 = 2.31$ meV, $\Delta A_0 = -0.04$ eV. The forward ET rate constants are calculated at different reorganisation free energies, as summarised in Table 3.6, columns 3,4. The $k_{ET} - \lambda$ dependence is hard to predict because $\lambda$ appears more than once in the rate expression. In particular, it acts on the exponential decay with the function of $\lambda + \frac{\text{const.}}{\lambda}$. It turns out that in the current regime where $\Delta A < \lambda$ is well-satisfied, $k_{ET}$ decreases with increasing $\lambda$, spanning two orders of magnitude from $10^8$ to $10^6$ s$^{-1}$. The error of $\lambda$ calculations due to inaccuracy in the classical force field is typically of the order of 0.1 eV, which means that the uncertainty of the estimated ET rate constant yielding from the reorganisation free energy is typically one order of magnitude.

**Sensitivity to the driving force** $\Delta A$. The electronic coupling matrix element and the reorganisation free energy are fixed at their averaged values: $|H_{ab}|_0 = 2.31$ meV, $\lambda_0 = 0.87$ eV. The forward ET rate constants are calculated at different driving forces, as summarised in Table 3.6, columns 5,6. In the current regime, $k_{ET}$ decreases dramatically with the increase of $\Delta A$, spanning four orders of magnitude from $10^9$ to $10^5$ s$^{-1}$. This is because $\Delta A$ first enters a quadratic function which then enters the exponential decay. In the normal Marcus regime when $\Delta A < \lambda$ is well-satisfied, the rate dependence is more prominent for $\Delta A$ than for $\lambda$ because the gradient is larger for the quadratic function than for the slowly increased branch of the asymmetric hyperbola $\lambda + \frac{\text{const.}}{\lambda}$. Unfortunately, the error of a $\Delta A$ calculation may be accumulated from several sources such as inaccuracy of force fields, choice of protonation states for the heme propionates and some charged amino acid side chains, insufficient sampling in the thermodynamic integration, so that the uncertainty is expected to be at least 0.1 eV, which means that the uncertainty of the estimated ET rate constant yielding from the uncertainty of driving force can easily exceed one order of magnitude.

To sum up, the driving force appears to be the most sensitive ET parameter in the estimation of heme-heme ET rate constants. However, to the best of my knowledge, driving force is the least accurate ET parameter among the three, if computed with conventional computational methods. Therefore, the experimental driving force is used whenever available (e.g., through difference in
Table 3.5: Summary of the mean, maximum and minimum of the computed heme-heme ET parameters in proteins STC, MtrC, MtrF (altogether 21 data points). \( |H_{ab}| \) is the electronic coupling matrix element, \( \lambda \) is the reorganisation free energy, \( \Delta A \) is the driving force. ET parameters are taken from tables in Appendix C.

|          | \( |H_{ab}| \) (meV) | \( \lambda \) (eV) | \( \Delta A \) (eV) |
|----------|-------------------|-----------------|------------------|
| Mean     | 2.31              | 0.87            | -0.04            |
| Maximum  | 4.51              | 1.13            | 0.19             |
| Minimum  | 0.70              | 0.68            | -0.25            |

Table 3.6: Sensitivity of the heme-heme ET rate constants with respect to the change of each individual Marcus ET parameter. In each set of tests, the forward ET rate constants are calculated at different values of the testing parameter varied from the minimum to the maximum in Table 3.5 while the other two ET parameters are fixed at their mean values. Mean value of the three ET parameters: \( |H_{ab}| = 2.31 \text{ meV}, \lambda_0 = 0.87 \text{ eV}, \Delta A_0 = -0.04 \text{ eV}. \)

|          | \( |H_{ab}| \) (meV) | \( \lambda \) (eV) | \( \Delta A \) (eV) | \( k_{ji} \) (s\(^{-1}\)) | \( k_{ji} \) (s\(^{-1}\)) | \( k_{ji} \) (s\(^{-1}\)) |
|----------|-------------------|-----------------|------------------|-----------------|-----------------|-----------------|
| 0.70     | 4.0 \times 10^6   | 0.68            | 3.1 \times 10^8  | -0.25           | 1.3 \times 10^9  | 1.3 \times 10^9  |
| 1.50     | 1.9 \times 10^7   | 0.77            | 1.2 \times 10^8  | -0.14           | 2.7 \times 10^8  | 2.7 \times 10^8  |
| 2.31     | 4.4 \times 10^7   | 0.87            | 4.4 \times 10^7  | -0.04           | 4.4 \times 10^7  | 4.4 \times 10^7  |
| 3.41     | 9.6 \times 10^7   | 1.00            | 1.2 \times 10^7  | 0.08            | 4.6 \times 10^6  | 4.6 \times 10^6  |
| 4.51     | 1.7 \times 10^8   | 1.13            | 3.1 \times 10^6  | 0.19            | 3.6 \times 10^5  | 3.6 \times 10^5  |

reduction potentials).

The analysis shown here would suggest that one promising strategy for tuning the ET rate constant might be to tune the heme reduction potentials. This might be achieved by, for example, changing the pH, introduction of point mutation to a nearby amino acid, and so on.

3.4 Maximum, protein-limited steady-state electron flux

Microscopically, the computed electronic couplings, reorganisation free energies and driving forces (either computed or from experiments) are used to calculate the non-adiabatic (Marcus) rate constants for all heme-to-heme ET steps. Macroscopically, these ET rate constants are then used as an input for chemical master equations, which are solved analytically to obtain the maximum, protein-limited electron flux through the multi-heme proteins.

We assume that the electron flow from an electron input site, heme \( k \), to an electron output site, heme \( l \), occurs via electron hopping along the chain of adjacent heme sites connecting \( k \) with \( l \). The electron flux between adjacent heme pairs \( i, j \) along the chain, \( J_{ji} = J_{ji-i} \), is described
3.4. Maximum, protein-limited steady-state electron flux

by a chemical master equation,

\[ J_{ji} = k_{ji}P_i(1 - P_j) - k_{ij}P_j(1 - P_i), \]  

(3.5)

where \( k_{ji} \) are the non-adiabatic (Marcus) ET rate constants between adjacent hemes \( i, j \).

\[ k_{ji} = \frac{2\pi}{\hbar} \langle |H_{ji}|^2 \rangle (4\pi\lambda_{ji}k_B T)^{-1/2} \exp \left( -\frac{(\lambda_{ji} + \Delta A_{ji})^2}{4k_B T \lambda_{ji}} \right), \]  

(3.6)

where \( H_{ji}, \lambda_{ji}, \Delta A_{ji} \) are the electronic coupling matrix element, reorganisation free energy, driving force of this ET step. \( P_i \) is the electron population of heme \( i \) (\( P_i = 0 \) for oxidized hemes and 1 for reduced hemes) [139, 140], and the terms \((1 - P_i)\) account for the fact that each heme can be occupied by only one excess electron. Electron injection into heme \( k \) with rate constant \( k_{in,k} \) and electron ejection from heme \( l \) with rate constant \( k_{out,l} \) are assumed to be irreversible, that is \( k_{l,out} = k_{in,k} = 0 \), and much faster than any of the heme-to-heme rate constants, \( k_{k,in}, k_{out,l} >> k_{ji} \quad \forall i, j \). The chemical master equation (equation 3.5) is solved analytically for the populations \( P_i \) subject to the steady-state condition

\[ J \equiv J_{k,in} = J_{ji} = J_{out,l} = \text{const} \quad \forall i, j = i + 1. \]  

(3.7)

A detailed explanation on how the coupled equations are solved analytically is given in Appendix B with protein STC as an example. The approach is originally proposed by Dr. Guido Falk von Rudorff.

In the regime where input and output rates are much faster than any of the heme-to-heme rates, the steady state flux \( J \) in equation 3.7 is limited by intra-protein ET and equals the maximum electron flux that the protein supports, \( J_{\text{max}} = J \). \( J_{\text{max}} \) has been calculated for forward and backward flow directions in STC: heme 1 \( \rightarrow \) 4, heme 4 \( \rightarrow \) 1, and all possible electron flow directions in MtrC and MtrF: \( k = 10, l = 5, 7, 2; k = 5, l = 10, 7, 2; k = 2, l = 10, 5, 7; k = 7, l = 10, 5, 2 \), using a self-consistent (SC) iteration procedure for driving force, as explained in the next subsection.

3.4.1 Influence of interactions between redox centres on ET parameters

Using the chemical master equation approach (equation 3.5) we have assumed that the ET rates \( k_{ji} \) and the rate determining ET parameters (average electronic coupling \( \langle |H_{ji}|^2 \rangle^{1/2} \), reorganisation free energy \( \lambda \), driving force \( \Delta A_{ji} \)) are independent on the instantaneous redox state of the
protein (i.e. number of reduced hemes). This is expected to be a good assumption for electronic couplings, which are determined by short-range interactions (orbital overlap). Indeed, electronic couplings have been shown to be insensitive to inclusion of the protein environment in QM/MM coupling calculations in MtrF [37]. Reorganisation free energies, primarily dependent on the protein dielectric and on the proximity of the redox groups to solvent water, are not expected to change sufficiently to have a significant effect on ET rates (very large changes in $\lambda$ of about 250 meV would be necessary to change the ET rate constant by one order of magnitude at the modest driving forces in multi-heme proteins). With regard to heme reduction potentials, NMR and redox titrations on the small tetra-heme protein STC have shown that they are sensitive to the redox state of neighbouring hemes [141, 142]. Accumulated shifts of reduction potentials of up to about 100 meV have been reported. The variations in reduction potentials due to redox state are not expected to change the steady-state electron flux greatly, which is confirmed below. Nonetheless, this effect should be investigated for quantitative modelling.

We have done so by taking into account the influence of the heme-heme interactions on the rate constants and electron flux in a “mean-field” way, by making the driving force (i.e. reduction potential differences) dependent on the heme occupation $P_i$. Alternatively, one may assume that the rate constants $k_{ji}$ depend explicitly on the redox and protonation state of the protein, i.e. on the heme occupations $P_i$. In this case, the chemical master equation (equation 3.5) can no longer be solved analytically, but other, numerical methods such as kinetic Monte Carlo would need to be used to estimate the equilibrium flux for this model.

Starting from the reduction potentials for the all-ox state of the protein (Figure 3.10 A), $\varepsilon_i(O)$, $i = 1, \ldots, N$, the steady-state flux $J_{\text{max}}$ and heme populations $P_i$ are calculated. The reduction potentials are then adjusted to the new heme occupations as follows, $\varepsilon_i = \varepsilon_i(O) - \sum_{j \neq i = 1}^{N} P_j \Delta \varepsilon_{ji}$, where $-\Delta \varepsilon_{ji}$ is the change of reduction potential of heme $i$ if the redox state of heme $j$ changes from oxidized to reduced (or, equivalently, $\Delta \varepsilon_{ji}$, is the change of reduction potential of heme $i$ if the redox state of heme $j$ changes from reduced to oxidized). To proceed, a new set of driving forces and rate constants are calculated for which a new value for $J_{\text{max}}$ and the steady state populations $P_i$ are obtained. This process is iterated until all reduction potentials change by less than a certain threshold (e.g., 0.1 meV in my calculation). If the reduction potentials of the all-red state of the protein, $\varepsilon_i(R)$, are initially used (Figure 3.10 B), as e.g. provided in Ref. [143], the adjustment of reduction potentials is $\varepsilon_i = \varepsilon_i(R) + \sum_{j \neq i = 1}^{n} (1 - P_j) \Delta \varepsilon_{ji}$, noting that $\varepsilon_i(R) = \varepsilon_i(O) - \sum_{j \neq i = 1}^{n} \Delta \varepsilon_{ji}$. 
For STC, both $\varepsilon_i$ and $\Delta \varepsilon_{ji}$ were taken from experiment [141]. Different starting redox potentials have been investigated, including all-red and all-ox states of the hemes as well as the protonation state of an unidentified ionizable centre, which are sensitive to the redox potentials as probed from the NMR and redox titrations [141]. The results from the self-consistent (SC) scheme are summarized in Table 3.7, for the flux in the forward and reverse direction. The forward flux for the state where all hemes are reduced and the ionizable site is protonated is $3.3 \times 10^6$ s$^{-1}$, before SC. Subsequent SC iterations of the heme occupations have only a small effect. The flux decreases slightly to $2.9 \times 10^6$ s$^{-1}$. The corresponding numbers for the ionizable site deprotonated are similar, $3.1 \times 10^6$ s$^{-1}$ and $3.9 \times 10^6$ s$^{-1}$ before and after SC iteration. To obtain the final electron flux we consider the total electron occupation of the four hemes, 1.66 and 1.34 for the ionizable site protonated and deprotonated, respectively. Experiments give a pKa of 7.3 and 7.0 for a total heme occupation of 2 and 1, respectively. We assume that these pKa values also apply to the ionizable site under the steady state flux condition considered herein. Hence, we find that for the steady-state heme occupations obtained, the pKa of the ionizable site is approximately 7. Therefore, we weight the flux obtained for the protonated and deprotonated ionizable site equally to obtain the final electron flux estimate for STC, $J_{\max} = 3.4 \times 10^6$ s$^{-1}$.

For MtrC and MtrF, the redox potentials are taken from computed values (from our lab for
**Table 3.7:** Computed protein-limited steady-state electron fluxes in the forward (heme $1 \rightarrow 4$) and reverse (heme $4 \rightarrow 1$) directions in STC. The reduction potentials used for calculations of $J_{\text{max}}$ are indicated as the following: R,p: reduction potentials in all-reduced state, the ionizable centre (IC) protonated; SC,p: self-consistent iteration of reduction potentials with the IC protonated; R,d: reduction potentials of all-reduced state with the IC deprotonated; SC, d: self-consistent iteration of reduction potentials with the IC deprotonated; O,d: reduction potentials of all-oxidized state with the IC deprotonated. The rate constants for ET from heme $i$ to $j$, $k_{ji}$, and $J_{\text{max}}$ are given in units of $10^{6}$ s$^{-1}$.

<table>
<thead>
<tr>
<th></th>
<th>$k_{21}$</th>
<th>$k_{32}$</th>
<th>$k_{43}$</th>
<th>$k_{12}$</th>
<th>$k_{23}$</th>
<th>$k_{34}$</th>
<th>$P_{1}$</th>
<th>$P_{2}$</th>
<th>$P_{3}$</th>
<th>$P_{4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>R,p</td>
<td>1.00</td>
<td>208.0</td>
<td>1.6</td>
<td>58.0</td>
<td>11.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
<td>0.15</td>
<td>0.00</td>
</tr>
<tr>
<td>SC,p</td>
<td>3.5</td>
<td>3.1</td>
<td>3.4</td>
<td>4.8</td>
<td>4.3</td>
<td>0.07</td>
<td>0.06</td>
<td>0.07</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>R,d</td>
<td>3.4</td>
<td>576.0</td>
<td>1.8</td>
<td>21.0</td>
<td>14.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.09</td>
<td>0.09</td>
<td>0.00</td>
</tr>
<tr>
<td>SC,d</td>
<td>4.3</td>
<td>135.0</td>
<td>1.6</td>
<td>50.0</td>
<td>16.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.09</td>
<td>0.09</td>
<td>0.00</td>
</tr>
<tr>
<td>O,d</td>
<td>8.7</td>
<td>231.0</td>
<td>1.2</td>
<td>16.0</td>
<td>13.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.31</td>
<td>0.45</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The resulting flux with different initial redox potentials and flow directions are summarized in Table 3.8.

### 3.4.2 Biological implications of ET in deca-heme proteins

In this subsection, I would like to summarize my calculations on heme-heme ET rate constants by discussing their biological implications. From the calculation of three ubiquitous multi-heme proteins binding 4 and 10 heme cofactors, we have found that the cysteine linkages inserted into the space between heme groups strongly enhance the electronic coupling. In this way, heme pairs forming motifs with larger heme edge-to-edge distances (e.g. co-planar or T-shaped) can exhibit similar ET rates as heme pairs with very short edge-to-edge distance (e.g. stacked orientations). Figure 3.11 show the heme-to-heme rate constants proportional to the width of the arrows connecting hemes, as well as the protein-limited electron flux for all 12 possible flow directions
Table 3.8: Maximum, protein-limited electron flux, $J_{\text{max}}$, through MtrC and MtrF in different flow directions. $J_{\text{max}}$ is obtained by solving the chemical master equation (equation 3.5) for heme-to-heme electron hopping. The ET driving forces were obtained by self-consistent (sc) iteration starting from the computed reduction potentials of the all-oxidized (ox-sc) or all-reduced (red-sc) redox state of the protein.

<table>
<thead>
<tr>
<th>direction</th>
<th>MtrC $\text{ox-sc}^a$</th>
<th>MtrC $\text{ox-sc}^b$</th>
<th>MtrC $\text{red-sc}^b$</th>
<th>MtrF $\text{ox-sc}^a$</th>
<th>MtrF $\text{ox-sc}^b$</th>
<th>MtrF $\text{red-sc}^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 $\rightarrow$ 5</td>
<td>$2.4 \times 10^5$</td>
<td>$1.6 \times 10^5$</td>
<td>$6.4 \times 10^4$</td>
<td>$8.9 \times 10^4$</td>
<td>$2.0 \times 10^6$</td>
<td>$8.0 \times 10^4$</td>
</tr>
<tr>
<td>5 $\rightarrow$ 10</td>
<td>$1.1 \times 10^5$</td>
<td>$5.0 \times 10^4$</td>
<td>$1.4 \times 10^6$</td>
<td>$1.2 \times 10^5$</td>
<td>$7.1 \times 10^5$</td>
<td>$2.2 \times 10^5$</td>
</tr>
<tr>
<td>10 $\rightarrow$ 7</td>
<td>$1.4 \times 10^7$</td>
<td>$6.1 \times 10^6$</td>
<td>$8.3 \times 10^4$</td>
<td>$1.0 \times 10^5$</td>
<td>$3.2 \times 10^6$</td>
<td>$9.5 \times 10^3$</td>
</tr>
<tr>
<td>7 $\rightarrow$ 10</td>
<td>$2.0 \times 10^5$</td>
<td>$2.7 \times 10^6$</td>
<td>$6.9 \times 10^5$</td>
<td>$1.4 \times 10^4$</td>
<td>$1.9 \times 10^4$</td>
<td>$3.8 \times 10^4$</td>
</tr>
<tr>
<td>10 $\rightarrow$ 2</td>
<td>$2.7 \times 10^5$</td>
<td>$2.9 \times 10^6$</td>
<td>$6.2 \times 10^4$</td>
<td>$8.3 \times 10^4$</td>
<td>$7.6 \times 10^5$</td>
<td>$7.8 \times 10^3$</td>
</tr>
<tr>
<td>2 $\rightarrow$ 10</td>
<td>$1.5 \times 10^4$</td>
<td>$1.1 \times 10^5$</td>
<td>$3.6 \times 10^5$</td>
<td>$4.8 \times 10^4$</td>
<td>$1.3 \times 10^4$</td>
<td>$3.4 \times 10^4$</td>
</tr>
<tr>
<td>2 $\rightarrow$ 5</td>
<td>$1.5 \times 10^5$</td>
<td>$7.9 \times 10^4$</td>
<td>$4.4 \times 10^5$</td>
<td>$7.2 \times 10^4$</td>
<td>$1.3 \times 10^4$</td>
<td>$3.7 \times 10^4$</td>
</tr>
<tr>
<td>5 $\rightarrow$ 2</td>
<td>$1.1 \times 10^5$</td>
<td>$5.0 \times 10^4$</td>
<td>$4.5 \times 10^6$</td>
<td>$8.0 \times 10^4$</td>
<td>$5.0 \times 10^5$</td>
<td>$2.0 \times 10^5$</td>
</tr>
<tr>
<td>7 $\rightarrow$ 5</td>
<td>$3.0 \times 10^4$</td>
<td>$1.5 \times 10^5$</td>
<td>$7.2 \times 10^5$</td>
<td>$1.6 \times 10^4$</td>
<td>$1.9 \times 10^4$</td>
<td>$3.6 \times 10^4$</td>
</tr>
<tr>
<td>5 $\rightarrow$ 7</td>
<td>$1.1 \times 10^5$</td>
<td>$5.0 \times 10^4$</td>
<td>$1.7 \times 10^6$</td>
<td>$1.5 \times 10^5$</td>
<td>$7.3 \times 10^5$</td>
<td>$2.2 \times 10^5$</td>
</tr>
<tr>
<td>2 $\rightarrow$ 7</td>
<td>$1.6 \times 10^4$</td>
<td>$1.1 \times 10^5$</td>
<td>$3.6 \times 10^5$</td>
<td>$1.2 \times 10^5$</td>
<td>$1.3 \times 10^4$</td>
<td>$3.2 \times 10^4$</td>
</tr>
<tr>
<td>7 $\rightarrow$ 2</td>
<td>$1.4 \times 10^5$</td>
<td>$2.9 \times 10^6$</td>
<td>$6.8 \times 10^5$</td>
<td>$1.6 \times 10^4$</td>
<td>$1.8 \times 10^4$</td>
<td>$3.3 \times 10^4$</td>
</tr>
</tbody>
</table>

$^a$ Reduction potentials for the all-ox redox state computed by our lab used as initial values for self-consistent iteration.

$^b$ Reduction potentials for the all-ox or all-red redox state as reported by Ref. [143] used as initial values for self-consistent iteration.

across MtrC and MtrF in the insets (in powers of 10 s$^{-1}$). The rate constants along the octa-heme chains of MtrC and MtrF span four orders of magnitude, from $\approx 10^2 - 10^9$ s$^{-1}$, and the electron flux is $\approx 10^5$ s$^{-1}$ in both the 10 $\rightarrow$ 5 and 10 $\leftarrow$ 5 directions. Electron flow from hemes 10 or 5 to the side exits 7 and 2 is similarly fast as along the octa-heme main chain, about 10$^5$ s$^{-1}$, except for 10 $\rightarrow$ 7 due to the relatively high reduction potential of heme 7. But the latter is subject to uncertainty as investigated by the previous PhD student [38, 143] and may be overestimated. The reverse flow from the side exits 7 and 2 to 10 and 5 is somewhat slower, typically about 10$^4$ s$^{-1}$, due to successive up-hill steps involving co-planar and T-shaped motifs. Similar results are obtained from the sets of reduction potentials from Barrozo et al. [143], with deviations of typically less than an order of magnitude (see Table 3.8).

We note that there is little difference in the protein-limited electron flow through MtrC and MtrF. Overall, our results indicate that MtrC and MtrF conduct electrons roughly equally well along their main axis and in perpendicular directions with little or no directional bias. The electron flow is reversible implying that both proteins not only support electron export but also electron import. While reorganisation free energies are slightly lower in MtrC than in MtrF,
no significant differences in electronic coupling are discernable. Thus, from the perspective of redox function the current characterization suggests that MtrC can be replaced by MtrF, as in fact observed experimentally [46, 47]. The higher expression levels of MtrCAB relative to MtrFDE at low O\textsubscript{2} concentrations [144] is thus more likely related to a genetic origin rather than electron transfer function.

Whilst the electron flux (with all heme side chains included) is remarkably similar for all directions, the flux enhancement due to the side chains is not the same in every direction – on the contrary, depending on the number of co-planar and T-shaped heme pairs and their free energies for a given flow direction, the flux enhancement varies from between a factor of 2 (10 \to 7 in MtrF, 1 T-shaped pair) to a factor of \approx 10^3 (10 \to 5 in MtrC, 1 co-planar and 2 T-shaped pairs). Without the cysteine-mediated coupling enhancement, the co-planar or T-shaped heme pairs in the middle of the protein limit the electron flux, whereas with coupling enhancement these ET steps become similarly fast as ET between stacked heme pairs. In this case there is no longer a clearly flux-limiting ET step; the two slowest steps are within an order of magnitude. We believe
3.5. \( I - V \) curve modelling

One of the potential applications of multi-heme proteins are bio-nanoelectronic devices for sensing, signalling, and electronic communication [4]. Although it is well established that the electron transfer (ET) in solvated multi-heme proteins happens via sequential hopping between adjacent heme pairs, the detailed mechanism of the electron transport (ETp) in bioelectronic junctions is still a matter of debate, as the protein junctions are sensitive to various factors such as the humidity, protein-electrode interface and supply of external electrons [29, 45]. In this section, I will present our studies of the electron transport (ETp) in multi-heme protein/electrode junctions. Two models were used to account for this transport – coherent electron tunnelling by model fitting and incoherent (sequential) electron hopping based on our computed ET rates. The modelled \( I - V \) curves will be compared with two different types of experiments: scanning tunnelling microscopy (STM) measurements by Rosso, El-Naggar and co-workers on single MtrC [28] and MtrF [29] proteins, and monolayer solid state ETp measurements carried out at high vacuum for STC and MtrF by Cahen et al. [31].

3.5.1 \( I - V \) from scanning tunnelling microscopy (STM)

The \( I - V \) curve of MtrF was modelled previously by Byun et al. [29] and Breuer et al. [37] with the electron hopping model. In the former study, a phenomenological form of the non-adiabatic rate equation was used for the estimation of heme-heme ET rates (with tunnelling decay factor \( \beta \approx 1 \text{Å}^{-1} \)) and a kinetic Monte Carlo algorithm was employed to solve the electron flux. The STM-measured \( I - V \) curve was fit with respect to reorganisation free energies, and it was found that \( \lambda = 0.80 \) and 0.55 eV could reproduce the curves at the high- and low-humidity conditions, respectively. In the latter study, the computed electronic couplings and reorganisation free energies were used for each heme pair along the octa-heme chain in MtrF, but this fell short of reproducing the \( \approx \) nano-Ampere (nA) currents at \( \pm 1 \) V bias voltage [29]. Even after accounting for partial protein hydration in these experiments, the computed STM currents remained underestimated by about two orders of magnitude. One possible explanation for the discrepancy is the cysteine side chains that could potentially enhance the electronic couplings in the previous rate-limited ET step(s). Since the effect of the cysteine linkages has not been included in the previous modelling by Breuer et al, following the same modelling framework I will report our
updated $I - V$ curve modelling for MtrF and MtrC.

The measured $I - V$ curves for MtrF [29] and MtrC [28] are modelled assuming the same ET mechanism as in solution, i.e., charge carrier hopping. The current is computed by $I = -e \cdot J$, where $J$ is the steady-state electron flux along the heme chain as described by equations 3.5-3.7. The heterogeneous electron input and output steps between heme 10 and the left electrode and heme 5 and the right electrode are treated as symmetric and reversible, i.e. $k_{10, in} = k_{5, out}$ and $k_{10, out} = k_{5, in}$. They are given by the electrochemical non-adiabatic ET rate equations (equations 2.19 and 2.20) [109, 145]. The heme-to-heme ET rates are obtained from equation 3.6. The ET parameters are chosen as discussed in the following.

The experiments to be modelled were carried out in air at ambient [28] and dry [29] humidity conditions. As in Ref. [37], we assume that heme-to-heme electronic couplings, $\langle |H_{ab}|^2 \rangle$, are the same as in aqueous solution because they depend only on the heme cofactor arrangement, which should not change significantly if the protein structure remains intact in the experiments. The effective electronic coupling between heme 10 (5) and the electrode was chosen to be large enough so that the current is limited by intraprotein electron transport. Defining $r = k_{10, in} / k_{ij}^{\text{min}}$ and $k_{ij}^{\text{min}}$ the slowest heme-to-heme rate constant, our choice of coupling between terminal heme and electrode gives $r = 10$. In regard to the reorganisation free energies, they are known to be sensitive to the degree of solvent exposure [146]. More specifically, Tipmanee et al. [147] found that the solvent contributed 50% to 75% of the total reorganisation free energy for a set of model ET proteins. We chose a value of $\lambda$ of about half the value computed for the fully solvated MtrC and MtrF: $\lambda = 0.4$ eV for intraprotein ET and $\lambda = 0.2$ eV at the protein-electrode interface. In regard to driving forces, no information is available on these under dry conditions, but it can be assumed that with partial or no solvation shell, ionizable groups should be either unionized or binding a counter ion so that the electrostatic potential in ambient air is more homogeneous than in water [37]. Therefore, all intrinsic driving forces were set to zero so that the total driving force for each heme-to-heme ET step is equal to the contribution from the external potential only, $\Delta A = \Delta A_{ji} = -V / (n + 1)$ with $n$ as the number of redox sites between the two electrodes ($n = 8$ for electron transport along the octa-heme chain). This assumption ignores minor differences in the potential drop due to the nonlinear spatial arrangement of hemes along the octa-heme chain, but this turns out to have a negligible effect on the current. For the electrode contacts, we assume the same potential drop as between the hemes, $E_L - E_{10} = E_5 - E_R = -V / (n + 1)$, where $E_L$, $E_R$ are the potentials of the left, right electrode, $E_{10}$, $E_5$ the reduction potentials of heme 10, heme 5.
Hence, the fraction of potential drop on the two electrodes is $\delta = \delta_L + \delta_R = 2 \times 1/(n+1) = 0.22$ ($n = 8$ for a octa-heme chain), which is close to the value used by Byun et al. [29] in their modelling of $I-V$ curves for MtrF ($\delta = 0.3$) [29].

![Figure 3.12:](image)

Figure 3.12: Experimental and computed STM $I-V$ curves for MtrF and MtrC in STM experiment. Experimental single-protein STM data are taken from Ref. [29] for MtrF (in air at low humidity conditions) and from Ref. [28] for MtrC (in air at ambient humidity). For comparison, the data for a protein monolayer junction of MtrF (in vacuum, $10^{-5}$ mbar) are also shown, data taken from Ref. [31]. The total number of MtrF molecules in the mono-layer is not known and is estimated to be roughly 10. Hence, the currents reported in Ref. [31] have been divided by this number. The computed protein-limited current for the charge carrier hopping mechanism is obtained by solving the chemical master equation (equation 3.5) using rate constants evaluated for electronic couplings obtained for the large QM model (solid lines, including the cysteine linkages) and for the small QM model (dashed lines in the inset, excluding the cysteine linkages).

Assuming the same hopping mechanism as for ET in solution [29, 37, 109, 140], we obtain currents of a few 0.1 nA at 0.5 V bias voltage for MtrC and MtrF, respectively, in good agreement with experiments [28, 29] (see Figure 3.12, “with CYS” lines). By contrast, without the cysteine-mediated electronic coupling enhancement, the currents are two orders of magnitude too low (see Figure 3.12 inset, “no CYS” lines). However, the shape of those experimental $I-V$ curves is not fully captured, especially at higher bias range ($0.5 - 1.0$ V), implying some additional mechanism, beyond incoherent hopping, might be relevant in the ETp event of protein junctions [29, 45]. In the next subsection, I will show the fitting to the $I-V$ curves with coherent electron tunnelling models.
3.5.2 \( I - V \) from suspended nanowire

In the suspended-nanowire experiment [30, 31], a closely packed protein monolayer was covalently formed on gold substrates at high vacuum environment. Three types of protein monolayers with different monolayer thicknesses were prepared using three proteins: STC (2.2±0.1 nm), MtrF (4.0±0.1 nm), and Azurin (2.0±0.2 nm). Azurin is another redox-active protein containing one single Cu ion, which shares similar dimension as STC, but has no hemes. A gold nanorod was used as a top contact, which had large contact area so that 10-1000 Au-protein-Au junctions could form at one time. Although it was confirmed from the experiment that there is no significant change in the molecular environment of the heme groups or the protein conformation upon monolayer formation, the experimental condition in the suspended-nanowire is fairly different from the STM measurements (high-vacuum vs air, protein monolayer vs single molecule, large-area suspended nanowire vs tip). The most striking feature is the temperature-independence of these \( I - V \) curves (see Figure 3.13 insets), measured at 50 mV from 80 to 300 K, which is incompatible with thermally activated incoherent hopping theory [31].

Nevertheless, the measured \( I - V \) curves are modelled for two limiting mechanistic regimes – coherent elastic tunnelling and incoherent electron hopping, as detailed below.

![Figure 3.13: I – V curves of STC (red), MtrF (blue) and Azurin (dark red) measured by suspended nanowire method (dots) and modelled by coherent tunnelling model (solid lines) and incoherent hopping model (dashed lines). Insets show the current of STC and MtrF in a wider temperature range at bias=0.05V.](image)

**Coherent electron tunnelling with Simmons model.** The coherent tunnelling theory (Simmons model) [28, 148, 149] approximates the protein matrix by a single effective barrier with barrier height \( \phi \), tunnelling length \( L \) and symmetry factor \( \alpha \) as fitting parameters. The tunnelling current
density \( i \) is given by

\[
i = \frac{M}{L^2} [(\phi - \alpha V)e^{-KL\sqrt{(\phi - \alpha V)}} - (\phi + (1 - \alpha)V)e^{-KL\sqrt{(\phi + (1 - \alpha)V)}}],
\]

(3.8)

where \( V \) is the bias between the two electrodes, \( K = \frac{4\pi}{h} \sqrt{2me} \) and \( M = \frac{e^2}{2\pi\hbar} \) are constants. Although the model yields an excellent fit (see Figure 3.13 solid lines, Table 3.9, correlation coefficient = 0.999), the tunnelling lengths obtained (\( L = 1.22, 1.57 \) and 1.21 nm for STC, MtrF and Azurin, respectively) are much shorter than the measured protein monolayer widths (i.e., distance between two electrodes). For STC and MtrF these distances are more characteristic of tunnelling from the electrode to one of the hemes in the middle of the protein, as suggested in Ref. [150].

Although three orders of magnitude higher conductance was observed via the tetra-heme STC protein junctions than via the blue Cu protein Azurin junctions, which forms monolayer junctions of similar thickness [31], the fitting only yields a slightly smaller distance decay constant \( \beta \) for STC. The coherent tunnelling model seems to be able to fit any protein almost equally well, which indicates an oversimplified model [151,152]. On the other hand, the temperature independence of the \( I - V \) curve provides strong evidence for the coherent tunnelling mechanism.

**Table 3.9:** Summary of the fitting results for the \( I - V \) curves of proteins STC, MtrF and Azurin with coherent tunnelling mechanism (Simmons model). Distance decay factor \( \beta \) is calculated by \( \beta = 1.025\phi^{1/2} \) [149].

<table>
<thead>
<tr>
<th>Fitting parameters</th>
<th>STC</th>
<th>MtrF</th>
<th>Azurin</th>
</tr>
</thead>
<tbody>
<tr>
<td>tunnelling length ( L ) (nm)</td>
<td>1.22</td>
<td>1.57</td>
<td>1.21</td>
</tr>
<tr>
<td>tunnelling barrier ( \phi ) (eV)</td>
<td>1.13</td>
<td>1.53</td>
<td>2.77</td>
</tr>
<tr>
<td>Symmetry factor ( \alpha )</td>
<td>0.42</td>
<td>0.49</td>
<td>0.51</td>
</tr>
<tr>
<td>Distance decay ( \beta ) (( \AA^{-1} ))</td>
<td>1.09</td>
<td>1.27</td>
<td>1.70</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>0.9995</td>
<td>0.9990</td>
<td>0.9999</td>
</tr>
</tbody>
</table>

**Incoherent electron hopping.** Alternatively, the ET mechanism in the cytochromes may be described by incoherent electron hopping along the heme chains of STC and MtrF, with the heme-cofactors switching between Fe(II) and Fe(III) redox states [29, 37]. The current was obtained by solving a chemical master equation for ET using non-adiabatic ET rate constants for heterogeneous ET between Au-electrodes and protein and for ET between the heme-cofactors. This is similar to the modelling of STM \( I - V \) measurements, except that here the reorganisation free energy \( \lambda \) was treated as a fitting parameter. We find that the very shallow T-dependence (Figure 3.13, insets) of the current requires small reorganisation free energies of no more than
0.10 eV. This is twice the computed value for inner-sphere reorganisation for two Fe-heme cofactors, implying that the outer-sphere contribution is very small, 0.05 eV. This is not entirely unreasonable as the protein solvation shell is removed under the experimental conditions and a significant fraction of the protein dielectric response is frozen on the time scale of electron transit across the protein (ps to ns). However, the fit of the $I - V$ spectra to the hopping model is not as good as for the coherent tunnelling model (Figure 3.13 dashed lines, correlation coefficient = 0.976, 0.980 for STC, MtrF, respectively), especially in the higher voltage regime where the current starts to saturate in contrast to experiment.

3.6 Conclusion

In this chapter, I reported detailed simulation procedure and our computed electron transfer (ET) parameters for three multi-heme proteins: STC, MtrC, MtrF. Our calculations reveal that electron hopping through these proteins is strongly enhanced by the cysteine side chains inserted in the space between heme groups. Apart from the three proteins intensively-studied herein, structural inspection over other periplasmic and outer-membrane multi-heme proteins have found that, the cysteine side chain groups also reduce significantly the porphyrin edge separations in the suboptimal heme pair packing motifs. This includes tri-heme protein PpcA [153], hexa-heme cytochrome OmcS [154] from *Geobacter sulfurreducens*, and deca-heme protein OmcA [53] from *Shewanella oneidensis*. We believe this to be a general design principle in proteins featuring c-type heme chains for acceleration of ET steps in the sub-optimal heme pair packing motifs (i.e., heme pairs with T-shaped, co-planar packing motifs), which would otherwise significantly limit the electron flux and would be too slow for biological respiration.

In general, the rate enhancement effect from our density functional theory calculations is consistent with the prediction from the empirical pathway-tunnelling model, yet the latter approach omits the chemical details of the tunnelling medium. We proposed a generalized distance metric for the donor-acceptor distance separation in the exponential decay of the electronic coupling in multi-heme proteins. The new distance metric not only takes into account the heavy atoms of the porphin ring, but also includes the side chains attached to the porphin ring, which is consistent with our QM model for the coupling calculations, and yields $\approx 4$ Å distance separation for all ET steps along the octa-heme chain in the two deca-heme proteins.

The time scales of heme-heme ET span the range 100 ps - 10 μs. The maximum, protein-limited electron flux has been calculated for all possible electron flow directions in the three proteins, using a self-consistent iteration procedure including the influence of heme-heme inter-
actions. We found that MtrC and MtrF conduct electrons roughly equally well along their main axis and in perpendicular directions with little or no directional bias. Hence, we speculate that the higher expression levels of MtrCAB relative to MtrFDE at low oxygen concentrations is more likely related to a genetic origin rather than electron transfer function.

Assuming the same hopping mechanism as for ET in solution, we modelled the $I-V$ curve for the three proteins, obtaining currents of a few 0.1 nA at 0.5 V bias voltage for MtrC and MtrF, respectively, consistent with the scanning tunnelling microscopy (STM) measurements. The favourable comparison with the STM currents should be considered with some caution, however, since a number of assumptions went into the modelling. The $I-V$ curves for monolayer solid state protein junctions were fit using both the incoherent hopping model and the coherent tunnelling model. Much better correlation was obtained using the coherent tunnelling model. The experimental conditions in this latter study are quite different to those used in the STM measurements (high-vacuum vs air, protein monolayer vs single molecule, suspended nanowire vs tip) which may tip the balance between different mechanisms.

We speculate that the true ET mechanism in the multi-heme cytochrome junctions might combine elements of both limiting pictures: slow coherent elastic tunnelling between electrode and heme cofactor that limits the experimentally measured current, followed by fast electron hopping along the heme chain. Incoherent hopping is known to have a much more favorable distance dependence ($1/L$) than tunnelling, which could provide a rational for the high conductance of multi-heme cytochrome compared to other redox proteins such as Azurin.

To properly model the $I-V$ measurements, however, one needs to obtain a reliable structural model of the protein adsorbed on the electrode, align the heme redox levels relative to the Fermi-levels of the electrodes, and incorporate partial nuclear relaxation on the time scale of electron transit between the two electrodes. These are the challenging aspects that needs to be incorporated to our modelling in the years to come [45, 155].
Chapter 4

Photo-Initiated Electron Transfer in 
Ru(bpy)$_3$-Labelled STC

Although Marcus theory, in principle, is well established for the description of thermally-mediated electron transfer (ET) processes in solvated multi-heme proteins, the computed heme-heme ET rate constants with the Marcus hopping model have not yet been verified experimentally. Existing electrochemical experiments have two limitations: i), the measured electron flux is prone to be rate-limited by the protein-electrode interface [156, 157]; ii), when sandwiched between two electrodes, the protein may adopt different electron transport mechanism from its native state in aqueous solution [45, 158]. In this chapter, I will introduce a different experimental approach, pump-probe spectroscopy, which holds promise in probing the heme-heme ET in multi-heme proteins without causing much perturbation to the protein’s native state in aqueous solution. To this end, our experimental collaborators (Prof. Julea Butt’s team at University of East Anglia and Central Laser Facility at Rutherford Appleton Laboratory) docked a Ru-dye close to a terminal heme of the tetra-heme protein STC and carried out pump-probe experiments to monitor a single electron injection in the protein and subsequent relaxation dynamics. After providing the experimental background, I will report our modelling of exactly the same experiment. In particular, I will focus on how insights from our simulation could help to extract the information on heme-heme ET from the experimental signal.

4.1 Experimental background

Pump-probe spectroscopy is often used to probe ultra-fast electron transfer processes across time scales ranging from pico- to milli-seconds, such as singlet fission [159], intersystem crossing [160, 161], and other excited state dynamics. At $t = 0$, a pump laser flash excites the system of interest, pumping a single electron to the excited state. Subsequently, at different time delays,
probe laser flashes are applied to the system, obtaining transient absorbance (TA) spectroscopy. The dynamics of the system can be studied by focusing on the time evolution of the absorbance signal at specific wavelength(s) characteristic of certain reaction process(es).

In the present work, the tetra-heme STC protein was labelled site-selectively with Ru(II)tris(bipyridine) [44]. Three STC variants with residues 10, 23, 77 at different positions mutated to cysteine were produced, two close to heme I and one close to heme IV, as illustrated in Figure 4.1 A (blue, yellow, and cyan circles). The Ru-dye was covalently docked to the protein through the sulfur atom of these mutated cysteine side chains (see Figure 4.1 B). The three Ru(II)tris(bipyridine)-labelled STC variants are denoted as Ru\textsubscript{10}:STC, Ru\textsubscript{23}:STC and Ru\textsubscript{77}:STC, respectively.

![Figure 4.1: Illustration of Ru(II)tris(bipyridine)-labelled STC protein. (A) STC protein and the microscopic reduction potentials of the four hemes in their oxidized states [141]. The four hemes are depicted in red, with protein backbone secondary structure depicted in grey. The C\textsubscript{α} atoms of residues changed to cysteines for labelling with the Ru(II)-dye photosensitizer are indicated as circles: residues 10 (blue), 23 (yellow) and 77 (cyan). Pannel A was prepared by Jessica H. van Wonderen. Adapt and reprint with permission. (B) The structural formula image of the thiol-reactive label, [Ru(II)(4-bromomethyl-4-methylbipyridine)(bipyridine)]\textsuperscript{2+} (Br atom not shown here), which is docked to the protein via a nucleophilic substitution reaction with the cysteine side chain.](image)

The hypothetical reaction scheme is summarized in Figure 4.2 A. Following photoexcitation into the Ru-dye metal-to-ligand charge transfer (MLCT) band, the triplet excited state (denoted as \textsuperscript{3}Ru:STC) was oxidatively quenched by heme reduction. Such electron transfer (reduction potential \(E_{m} \text{Ru(III)/(II*)} \approx -870 \text{ mV} [162]\)), produces a charge separated state, Ru\textsuperscript{+}:STC\textsuperscript{-}, that will return to the ground state (denoted as Ru:STC) by charge recombination (\(E_{m} \text{Ru(III)/(II)} \approx 1270 \text{ mV} [162]\)). However, heme-heme electron transfer in Ru\textsuperscript{+}:STC\textsuperscript{-} could result in each heme existing transiently as Fe(II) (denoted as CS\textsuperscript{M}, M=VI,III,II,I); the corresponding microscopic \(E_{m}\) values [141] lie between -120 and -215 mV as shown in Figure 4.1 A. Gaining direct spectroscopic evidence for electron transfer along the heme wire will be challenging, as differ-
ent hemes can not be distinguished chemically. However, we reasoned that heme-heme electron transfer might influence the dynamics of the corresponding photocycle in a manner that could be resolved by ultra-fast pump-probe spectroscopy.

The population of the three states in the reaction scheme are extracted at: 369 nm for $^3$Ru:STC, 419 nm for Ru$^+\text{-}:\text{STC}^-$, 453 nm for Ru:STC. The evolution and interconversion of these states for the Ru$_{77}$:STC are shown in Figure 4.2 B. The decay of the $^3$Ru:STC and the increase of the Ru$^+\text{-}:\text{STC}^-$ populations indicate the formation of the charge separated state. The decay of the Ru$^+\text{-}:\text{STC}^-$ and the onset of the Ru:STC populations indicate the recovery of the ground state of the Ru-dye. The sum of the three states (Figure 4.2 B, blue dots) is well conserved at any time $t$ indicating that all the relevant states have been included in the current raw data processing.

However, some features of the signals are not as expected from the reaction scheme in Figure 4.2 A: neither the $^3$Ru:STC nor the Ru:STC line exhibits mono-exponential increase or decrease, and the Ru$^+\text{-}:\text{STC}^-$ line is puzzling in that after an initial increase and slight decrease, a plateau is formed during the time window of 0.1-1 ns. To provide more insight into the TA data, the Ru$_{77}$:STC system has been modelled explicitly with molecular dynamics simulation and density functional theory.

In the following sections I will first present a detailed account of our studies of Ru$_{77}$:STC.
where the Ru-dye lies adjacent to heme IV (see Figure 4.1 A, cyan circle). An extended kinetic scheme is proposed which includes electron transfer across the heme IV ↔ heme III pair, which is validated by comparison to the predictions from our molecular simulation and corresponding electron transfer dynamics. Finally, our study of TA of proteins having the Ru-dye adjacent to heme I, \( \text{Ru}_23: \text{STC} \) and \( \text{Ru}_{10}: \text{STC} \), is presented and interpreted through an equivalent model to define rate constants for electron transfer across the heme I ↔ heme II pair.

### 4.2 Docking and molecular dynamics simulation

In order to model the Ru(II)tris(bipyridine)-labelled STC mutant, the electron transfer (ET) parameters and ET rate constants in the charge separation (CS) and charge recombination (CR) steps have to be computed. To this end, molecular dynamics simulations in the initial and final ET states for CS and CR are carried out as detailed in this section.

#### 4.2.1 Structural model for \( \text{Ru}_{77}: \text{STC} \)

Starting from the PDB structure of the native STC protein (PDB code 1M1Q [55]), all the redox-active groups were prepared in their oxidized states. All protein residues were in the standard protonation states at pH = 7 except one propionate from heme III, which was treated as protonated. The serine residue 77 (SER77) was replaced by a cysteine and the hydrogen atom of S-H was replaced by the Ru(bpy)\(_3\)(4-methylbipyridene-4'-CH\(_2\)) label (created with GaussView [163]) such that the S-C bond is 1.8 Å and the C-S-C bond angle is 98°, in accord with standard force field parameters. The orientation of the Ru(bpy)\(_3\) label with respect to the protein is determined by 3 dihedral angles, \( (\tau_1, \tau_2, \tau_3) \), defined in the inset of Figure 4.3 A. A large number of trial docking structures were generated in the conformational space spanned by the three dihedrals to obtain the most stable conformations of the Ru-dye relative to the protein (see below).

The dihedral angles \( (\tau_1, \tau_2, \tau_3) \) are defined by four consecutive atoms connected by covalent bonds, here given in terms of AMBER atom types: \( \tau_1 = \text{CYS:CA-CYS:CB-CYS:SG-BPY:C31}, \) \( \tau_2 = \text{CYS:CB-CYS:SG-BPY:C31-BPY:C30}, \) \( \tau_3 = \text{CYS:SG-BPY:C31-BPY:C30-BPY:C29}, \) where CYS is the residue name of the mutant cysteine residue 77, BPY is the residue name of Ru(bpy)\(_3\), C31 is the carbon atom in the linking methyl group, C30, C29 are two aromatic carbons. The docking was performed by sampling the three dihedral angles from 0° to 360° in increments of 10° (coarse scan), and then 5° (finer scan). In this way more than 100,000 trial structures were generated. The total energy of the system in vacuum with and without neutralizing counter ions is plotted against the three dihedral angles in Figure 4.3 B and C. We
Figure 4.3: (A) One representative docking structure of Ru77:STC with heme cofactors and the Ru-dye highlighted. The three docking dihedral angles (τ₁, τ₂, τ₃) are defined in the inset. (B, C) Energy of docking structures in vacuum without (B) and with (C) neutralizing counter ions. Each docking structure is represented by three data points placed at the same energy and at the respective values for the dihedral angles, τ₁ (black), τ₂ (red), τ₃ (green).

obtained 100 (with counter ions) and 200 (without counter ions) structures that were within 20 kcal/mol of the structure with the lowest energy. All of these structures were then energy minimized with protein (and counter ions if any) fixed for 100 steps and clustered with respect to the three dihedral angles. This resulted in a total of four unique low-energy clusters termed conformers in the following discussions, which are described by the following sets of dihedrals, (-140°, -83°, -107°), (-147°, -77°, 125°), (-111°, -60°, 103°), and (180°, 40°, 75°). These conformers were used as initial structures for molecular dynamics (MD) simulations. For MD simulations the four conformers were solvated with a shell of 15 Å with a total of 9012, 9015, 9197, 9647 water molecules (simulation box ≈ 62 × 60 × 76 Å). Na⁺ and Cl⁻ counter ions were added corresponding to an ionic strength of 0.1 M.
4.2.2 Force field parameters

The AMBER03 force field was used for docking and MD simulation together with the TIP3P water model and the monovalent ion parameters for Na\(^+\) and Cl\(^-\). The force field parameters for the hemes were the same as those used for the native STC protein. The parameters for the Ru(bpy)\(_2\)(4-methylbipyridene-4'-CH\(_2\)) label were taken from Ref. [164] for Ru(bpy)\(_3\) together with standard AMBER11 parameters for methylbenzene and methionine by chemical analogy. The total partial charges on the Ru-dye were slightly modified to sum up to +2/+3 for the reduced/oxidized state through redistribution of the extra charge onto the capping hydrogen atoms. In addition, the partial charges on the methylene group was modified to sum up to the partial charge of a hydrogen atom in the original Ru(bpy)\(_3\) force field. Parameters for bonds/angles were taken from Ref. [164] and standard AMBER force field parameters for the introduced methylene group. The RMSD of force field minimized Ru(bpy)\(_3\) relative to the DFT geometry optimized structure is small, 0.06 Å/0.05 Å in vacuum for the reduced/oxidized state, indicating that the chosen force field parameters give a good description of the Ru-dye.

4.2.3 MD simulation of Ru\(_{77}\):STC and Ru\(_{77}\):STC\(^+\) (CS\(_{IV}\))

Four MD simulations for the ground state Ru\(_{77}\):STC (Ru\(^{2+}\), all hemes Fe\(^{3+}\), abbreviated in the following as state A) were launched, starting from the four low-energy docking poses of the Ru-dye (see subsection 4.2.1). Each system was first minimized for 5000 steps to remove bad contacts in the crystallographic protein structure. Subsequently, each system was equilibrated for 500 ps (timestep 2 fs) with the protein atoms frozen and applying rescaling of the temperature to 300 K every 10 ps and a Langevin barostat with a target pressure of 1.013 bar. The protein was then slowly released by decreasing the harmonic restraints around the crystallographic positions with the force constants 99, 25, 1.0, 0.1 and 0.001 kcal·mol\(^{-1}\)·Å\(^{-2}\). For each force constant, the protein was equilibrated for 500 ps (timestep 1 fs) in the NPT ensemble applying velocity rescaling and a Langevin barostat. After all restraints were dropped, the protein was further equilibrated for 10 ns (timestep 2 fs) with a Langevin thermostat with a target temperature of 300 K and a Langevin barostat with a target pressure of 1.013 bar. Using the latter settings, subsequent production runs were carried out for 100 ns (total length 400 ns for four trajectories) in the NPT ensemble.

Analysis of the MD trajectories shows that the Ru-dye adopts four distinct conformations, see Figure 4.4 A where, for clarity, only three are shown. Heterogeneity is expected in our system as the Ru-dye is attached to the surface of the protein via the flexible protein side chain. These
conformations are stable on the nanosecond time scale and populate distinct low-energy basins characterized by dihedral angles \( \tau_2 \) and \( \tau_3 \) (see Figure 4.4 B). We note that the three dihedral angles of the equilibrated protein differ somewhat from the ones of the low-energy docking poses, possibly because solvent molecules were not present in the docking. In two conformers, 1 and 2, a bpy ligand of the Ru-dye approaches the heme IV ring very closely, with bpy-to-heme edge-edge distance = 5.2 Å (thermal average), whereas in the other two conformers 3 and 4 the distances are larger, 6.1 and 8.5 Å, respectively (see Figure 4.4 C). The corresponding distances to the second-nearest heme, heme III, are larger, by 3 Å or more, suggesting that simultaneous electron injection into heme III is significantly slower and can be neglected to a good approximation. The averaged dihedrals and Ru(bpy)_3-heme IV distances are summarized in Table 4.1, state A.

In Figure 4.5 the four trajectories are shown colour-coded indicating the instantaneous conformation of the Ru-dye (note \( t=0 \) corresponds to the start of the production run at which point conformation 2 (4) had converted to conformation 1 (3)). During the 100 ns simulations, we observe interconversion between conformers 1 and 2 on the 1 - 10 ns time scale, whereas conformers 3 and 4 appear to be stable. Hence on the time scale of current experiments, 1 ps - 100 ns, 1 and 2 interconvert and 3 and 4 are stable. We note that in reality, there might be more than 4 conformers interconverting on time scales that are different from what have been reported herein. More standard MD procedures, such as multiple replicas and enhanced sampling techniques, are required in order to draw more robust conclusions for the conformation of the Ru-dye. However, the aim of current MD simulation is to foster the understanding of the detailed reaction mechanism happening in the experiment. That is why several short MD trajectories, initialised at different local minimums of the docking structures, have been generated independently and concatenated.

Four MD simulations in the CS\(^{IV}\) state of Ru\(^{77}\):STC\(^{-}\) (Ru\(^{3+}\), heme IV Fe\(^{2+}\), all other
hemes Fe$^{3+}$, hereafter referred to as state B) were initiated from the four equilibrated conformers of state A. After equilibration to 300 K, the trajectories were run for 20 ns each. The dihedrals and equilibrium bpy-to-heme IV edge-edge distances are very similar in the states A and B, as summarized in Table 4.1. As before, conformers 1 and 2 interconvert on the time scale of 1 - 10
4.3 Estimation of ET parameters for charge separation and charge recombination

Rate constants for charge separation (CS) and charge recombination (CR) between the Ru-dye and heme IV are calculated in the framework of non-adiabatic (Marcus) ET theory for the four conformers of the Ru-dye, separately. In the following I will describe the calculations of the ET parameters required.

4.3.1 Reorganisation free energies for CS and CR

The reorganisation free energy for the CR step, $\text{CS}^\text{IV} \rightarrow \text{Ru}_{77}:\text{STC}$, was calculated for all 4 conformers $x = 1$-$4$ using the linear response approximation, $\lambda_{\text{CR, st}}^x = (\langle \Delta E \rangle_B^x - \langle \Delta E \rangle_A^x)/2$, where $\langle \Delta E \rangle_Y^x = (E_A - E_B)_Y^x$ is the vertical energy gap averaged over the trajectories run in state $Y=A,B$ that corresponds to conformer $x$, and the subscript “st” indicates that reorganisation free energy is defined via the Stokes shift. The energies $E_A$ and $E_B$ were calculated with the AMBER03 force field and averaged over all snapshots corresponding to the respective conformer. The inner-sphere contribution to reorganisation free energy is replaced by a QM estimate as reported previously [137, 164] (25 meV for heme cofactor, plus 40 meV for Ru(bpy)$_3$). The outer-sphere contribution due to the protein and solvent is divided by the Pekar factor = 1.35 to correct for the

Figure 4.5: Four MD trajectories of Ru$_{77}$:STC (in state A) starting from the four low-energy docking structures. The instantaneous conformation of the Ru-dye is colour-coded as indicated. Colour code is the same as Figure 4.4.

ns and conformer 3 and 4 are more stable. All MD simulations were carried out with NAMD software [58].
missing explicit electronic polarization of the protein and solvent in the AMBER03 force field. The above Pekar factor was obtained for a similar multi-heme protein, MtrF, by comparison to simulations with electronically polarizable force fields [4]. Similar calculations of the reorganisation free energy for CS would require the simulation of the excited triplet electronic state $^3\text{Ru}_{77}\text{STC}$. This is challenging with classical MD simulations where the force field parameters are parameterized to reproduce ground state properties. Here we assume that the altered charge distribution of $^3\text{Ru}_{77}\text{STC}$ with respect to $^3\text{Ru}_{77}\text{STC}$ has only a minor effect on the reorganisation free energy and set $\lambda_{\text{CS,ST}} = \lambda_{\text{CR,ST}}$. This should be a good approximation because in a previous work on Ru(bpy)$_3$-labelled cytochrome c it has been shown that reorganisation free energy for ET from a similar Ru(bpy)$_3$ label to Fe-heme was insensitive to the actual charge distribution on the label [165].

**Non-ergodicity correction.** In the above procedure, the reorganisation free energy was averaged over rather long trajectories. These time scales were significantly longer than the inverse of computed CR and CS rates (and the experimental CR and CS dynamics occurring on the ps-ns time scale). Nevertheless, since the experimental signal is an ensemble average over a large number of protein configurations, we decided to average over the entire lengths of the trajectories. To account for the fact that only the ps-ns response should be included, we followed the procedure from Matyushov [48] and removed from the full reorganisation free energy the frequency components that were slower than CR and CS rates. The workflow for this non-ergodicity correction [4] is illustrated in Figure 4.6 A.

For each conformer, a MD trajectory of 1 ns length was generated for state A starting from an equilibrated structure. The vertical energy gap $\Delta E$ was calculated every 5 fs and the auto-correlation function of the vertical energy gap fluctuation, $\langle \delta\Delta E(0) \cdot \delta\Delta E(t) \rangle$, where $\delta E = \Delta E - \langle \Delta E \rangle$, and its cosine transform computed. We note that in the linear response regime $\lambda_{\text{CR,ST}} = \lambda_{\text{CR,VAR}} = \frac{\sigma^2}{2k_BT}$, where $\sigma$ is the variance of the energy gap fluctuations and the subscript “var” denotes the definition of reorganisation free energy from the variance. To filter out the slow modes, we started from an initial guess $\lambda_{\text{CR,0}} = \lambda_{\text{CR,ST}}$ and obtained the ET rate $k_{\text{ET}}^{\text{CR,0}} = k_{\text{ET}}(\lambda_{\text{CR,0}})$ from the standard non-adiabatic Marcus ET formula, equation 2.18. We then updated the reorganisation free energy according to $\lambda_{\text{CR,i}} = \lambda_{\text{CR,ST}} - \frac{1}{\pi \sigma} \int_0^{2\pi/\sigma} d\omega J(\omega)$, where $\frac{J(\omega)}{\omega} = \frac{1}{\pi \sigma} \int_{-\infty}^{\infty} dt \cos(2\pi \omega \epsilon t) \int_{-\infty}^{\infty} dt' \delta\Delta E(t + t') \cdot \delta\Delta E(t')$, $J(\omega)$ is the spectral density function. In this way the contributions from modes that were slower than the rate constants were excluded from the reorganisation free energy $\lambda_{\text{CR,i}}$. After that, the CR rate was updated,
4.3. Estimation of ET parameters for charge separation and charge recombination

**Figure 4.6:** Non-ergodicity correction for outer-sphere reorganisation free energies. (A) Workflow for the self-consistent non-ergodicity correction and (B) outer-sphere reorganisation free energy for CS as a function of the inverse frequency cut-off. The self-consistently corrected reorganisation free energy is cut-off at the corresponding ET rate constant, indicated in dashed lines.

\[ k_{\text{CR}}^{\lambda} = k_{\text{ET}}(\lambda_{\text{CR}}^{\lambda}, i) \]

These steps were repeated until convergence was reached, defined by the threshold \(|k_{\text{CR},i}^{\lambda} - k_{\text{CR},i-1}^{\lambda}| < 10 \text{ fs}^{-1}\). A similar self-consistent non-ergodicity correction was carried out for the CS rates. The non-ergodicity-corrected reorganisation free energies \(\lambda_{\text{CR}}^{\lambda, n}\), \(\lambda_{\text{CS}}^{\lambda, n}\), and ET rates \(k_{\text{CR}}^{\lambda} = k_{\text{ET}}(\lambda_{\text{CR}}^{\lambda, n})\), \(k_{\text{CS}}^{\lambda} = k_{\text{ET}}(\lambda_{\text{CS}}^{\lambda, n})\), are summarized in Table 4.2, respectively. The non-ergodicity correction is significant for the fastest conformers 1 and 2 and also for 3, leading to a reduction of reorganisation free energy by 25 % for CS and by less than 10 % for CR. However, the CS rates only change by roughly a factor of 2 (uncorrected rates not shown in the table). The reason is that CS occurs in a regime where activation free energy is very small, \(\Delta A \approx -\lambda\), and in this regime the CS rate is rather insensitive to reorganisation free energy.

### 4.3.2 Electronic coupling matrix elements for CS and CR

Electronic couplings were computed with the Fragment-Orbital Density Functional Theory (FODFT) method [111, 112] as implemented in the CPMD code [91]. A systematic test on the size of the quantum mechanical (QM) model was carried out, similarly as described in section 3.1.1. The test was done on one snapshot from the trajectory and the same model was used for calculations of both charge separation and charge recombination (see Figure 4.7 A, B). For the heme IV fragment, all substituents were replaced by H except the two substituents inserting in the space between heme IV and the Ru-dye, that was, one methyl group and one cysteine.
Table 4.2: Summary of ET parameters and rate constants in CS and CR for the different conformers \(x = 1 - 4\). Reorganisation free energies (\(\lambda_{st}^{x}, \lambda_{CS}^{x}, \lambda_{CR}^{x}\)) and driving forces (\(\Delta A_{CS}, \Delta A_{CR}\)) are in units of eV, average electronic couplings \(\sqrt{\langle |H_{ab}|^2 \rangle}\) in units of meV and inverse rate constants (1/\(k_{CS}^{x}, 1/k_{CR}^{x}\)) in units of ps.

\[
\begin{array}{cccccccc}
 x & \lambda_{st}^{x} & \Delta A_{CS}^{a} & \sqrt{\langle |H_{ab}|^2 \rangle}_{B} & \lambda_{CS}^{x} & 1/k_{CS}^{x} & \Delta A_{CR}^{b} & \sqrt{\langle |H_{ab}|^2 \rangle}_{A} & \lambda_{CR}^{x} & 1/k_{CR}^{x} \\
 1 & 0.98 & 4.3 & 0.74 & 2.8 & 2.3 & 0.93 & 108 \\
 2 & 0.97 & -0.7 & 4.0 & 0.74 & 3.2 & 1.7 & 0.94 & 169 \\
 3 & 1.11 & 2.4 & 0.90 & 15.3 & -1.4 & 1.4 & 1.00 & 149 \\
 4 & 1.17 & 0.3 & 1.17 & 3900 & 0.2 & 1.17 & 1700 \\
\end{array}
\]

\(a\) Obtained from the experimental reduction potentials [141] of oxidized deprotonated heme IV in STC, \(\varepsilon^{0} = -0.129\) V vs NHE, and standard reduction potentials of Ru(bpy)\(_2\)(4,4’-dimethyl-2,2’-bipyridine) [162], \(\varepsilon^{0} = -0.83\) V vs NHE.

\(b\) Obtained from the experimental reduction potentials [141] of oxidized deprotonated heme IV in STC, \(\varepsilon^{0} = -0.129\) V vs NHE, and standard reduction potentials of Ru(bpy)\(_2\)(4,4’-dimethyl-2,2’-bipyridine) [162], \(\varepsilon^{0} = 1.27\) V vs NHE.

linkage. The side chain of the cysteine linkage was capped as -CH\(_3\)CHSCH\(_3\). For the Ru-dye fragment, all ligands as well as the side chain of the mutated cysteine residue capped with a H atom were included, resulting in Ru(bpy)\(_2\)(4-methylbipyridene-4’-CH\(_2\))-S-CH\(_3\).

DFT calculations were carried out with the Goedecker-Teter-Hutter (GTH) pseudopotential [166] and the PBE functional [79]. The orbitals of the donor and acceptor fragments were optimized so that the wave function gradient was below 10\(^{-6}\) Ha. The multiplicity of the Ru(bpy)\(_3\) fragment was set to triplet in the wave function optimization of the triplet excited state (\(^3\)MLCT state) so that the computation with conventional ground-state DFT had the correct electronic orbitals occupied. For CS, the triplet \(^3\)MLCT orbital [167] of Ru(bpy)\(_3\) was selected to couple with the two (quasi-)degenerate frontier orbitals of heme IV, which were composed of the Fe-d (t\(_{2g}\)) and heme ring orbitals (see Figure 4.7 A). For CR, the highest occupied orbital of Ru(II)(bpy)\(_3\), composed of Ru-d (t\(_{2g}\))-bpy ring orbitals, was coupled with the (quasi-)degenerate frontier orbitals of heme IV (see Figure 4.7 B). For CS and CR, the coupling matrix elements obtained for each of the two orbital pairs was square-averaged to an effective coupling matrix element, as for the heme-heme coupling calculations described in Chapter 3.

For each conformer, a minimum of 20 snapshots were extracted equidistantly in time from the corresponding trajectories in states A and B for calculation of the ensemble average of the effective coupling matrix element. The couplings were square-averaged over all selected configurations. To account for the missing electronic polarization between donor and acceptor in the FODFT method, the final values reported in Table 4.2 were scaled by a factor of 1.348 as recom-
4.3. Estimation of ET parameters for charge separation and charge recombination

Figure 4.7: Pair of redox active frontier orbitals contributing to electronic coupling are shown superimposed on the fragmental structures used in the coupling calculations for charge separation (A) and charge recombination (B). Probability distributions of the electronic couplings are plotted for charge separation (C) and charge recombination (D) as obtained from MD simulations and DFT calculations. Colour code in (C) and (D) for different Ru-dye conformers: 1, black; 2, red; 3, green; 4, blue.

recommended in the benchmark study Ref. [114]. The distributions of scaled couplings, smoothed by Gaussian window averaging, is plotted in Figure 4.7 C for CS and in Figure 4.7 D for CR.

We find that electronic coupling is lower by approximately one order of magnitude for the conformer with the largest bpy-to-heme IV edge-edge distance (4, Figure 4.7 C and D, blue) compared to the conformers with shorter distances (1 and 2, Figure 4.7 C and D, black and red). By contrast, the reorganisation free energy, suitably corrected for non-ergodic effects on the present ultra-fast time scale, is higher by about 0.2 eV for 4 than for 1 and 2, which might be due to the location of conformer 4 being more solvent-exposed. The ET parameters and rate constants for CS and CR in the four conformers are summarized in Table 4.2.

Using these electron transfer parameters three distinct sets of rate constants are reported, spanning the ps-ns regime (summarized in Table 4.6). One set arises from the two conformers with the smallest distances, 1 and 2, and one set each from 3 and 4. These sets are assigned to the kinetically distinct groupings of conformers a (1 + 2), b (3) and c (4) proposed in the reaction scheme model of the TA data (see subsection 4.5 below) in order of decreasing charge separation.
rate constants (see Table 4.6). We note that additional reactive conformers may exist in addition to the ones we have found, but we would expect their electron transfer kinetics to be similar to those of either a, b or c.

4.4 Influence of the Ru-dye on heme reduction potentials

The fact that heme cofactors are rigidly-embedded in the proteins allows us to make the assumption that the heme-heme electronic couplings and reorganisation free energies, which are dominated by the geometry, remain unchanged in the Ru-dye labelled STC variants with respect to the wild type STC protein. However, reduction potentials can be fairly sensitive to the local electrostatic environment. The purpose of this section is thus to have a careful look at the effect of the positively-charged Ru-dye on the reduction potentials of negatively-charged heme cofactors. To this end, we have calculated the reduction potentials of the heme cofactors in STC and one of the Ru-dye labelled STC variants, Ru_{77}:STC. The reduction potential \( A \) is estimated from the electrostatic energy difference, \( A = E_O - E_R \), between the oxidized (O) and reduced (R) states of a certain heme cofactor. The electrostatic energies \( E_M, M = \{O,R\} \) are computed by solving the linearized Poisson-Boltzmann (P-B) equations (equation 2.38) with Delphi program [168].

**Choice of parameters for P-B equation.** The ingredients for P-B equations are positions and partial charges of all protein atoms, boundary and relative dielectric constants for the protein (internal) and the solvent (external) regions, solvent salt concentrations, and the grid spacing for the finite difference solver.

In literature, optimized dielectric constants for specific proteins are often reported to have a value larger than the default relative dielectric constant for protein, \( \varepsilon = 4.0 \) [169], and different parts of the protein might have rather different dielectric responses depending on their local environment such as charge and polar properties, hydrophobicity and cavity size [170]. Therefore, there is no universal dielectric constant for proteins due to their complex nature. In this project, the relative dielectric constant for the protein region was found to be 10.0, which best reproduced the experimental reduction potentials of native STC in solution [141]. The dielectric constant for solvent was set to 78.2, which is the dielectric constant for bulk water at \( T=25^\circ C \). The boundary for protein and solvent was automatically detected by the code with a probing radius of \( r = 1.4 \) Å. The ionic strength was set to default value (0.1 mol/l), with an ion-exclusion probe radius of 2.0 Å. The partial charges were mapped onto cubic grids with a grid spacing of 0.2 Å for the finite difference representation of the P-B equation.
4.4. Influence of the Ru-dye on heme reduction potentials

4.4.1 Heme reduction potentials for STC

The structure for the calculation was prepared as follows. Starting from the experimental crystal structure of native STC (PDB code: 1M1Q [55]), a local minimization of the experimental PDB structure was carried out in vacuum with the protein and neutralizing counter ions fixed, only allowing the heme cofactors (i.e., atoms belonging to the QM region as defined in the final model of MtrF, see section 3.1.1) to be optimized. This decision was made based on the following two considerations: i), since the present calculation was based on a static structure, the experimental structure was preserved to the best of our ability; ii), the heme cofactors, whose electronic properties were of current interest, should be in equilibrium with the surrounding electrostatic field. The pKa values of the ionizable amino-acid side chains were determined using Propka [171]. All amino acid side chains were in their standard protonation state at pH = 7, and two propionate groups, one in heme II and one in heme III, were protonated. The counter ions were removed after the local minimization.

The results are shown in Figure 4.8 (brown solid line). The trend in calculated reduction potential as a function of heme number is in good agreement with the experimental trend (brown dashed line), only slightly overestimating the driving force for heme IV → heme III, indicating reasonable choice of the protonation states and the parameters for the P-B solver.

Figure 4.8: Reduction potentials of STC and Ru77:STC computed from linearized Poisson-Boltzmann (P-B) equations. The brown solid line plots the reduction potentials of STC; black, red, green, blue solid lines are the reduction potentials of Ru77:STC colour-coded according to the four conformers of the Ru-dye. The brown dashed line plots the reduction potentials taken from NMR measurements of all-oxidized, deprotonated state from Ref. [141].
4.4.2 Heme reduction potentials for Ru$_{77}$:STC

The structures for Ru$_{77}$:STC were taken from the four typical MD snapshots (see Figure 4.4 A) while leaving all protein atoms at their crystallographic positions.

The resulting reduction potentials are plotted in Figure 4.8 (black, red, green, and blue lines). As expected, the positively-charged Ru-dye stabilised the reduced heme state, i.e. led to an increase in the reduction potential, although this effect was relatively small. The increase was 0.036 eV for heme IV (the closest heme) and 0.020 eV for the other hemes. Hence, the net effect of ET in heme IV $\rightarrow$ heme III due to the presence of the Ru-dye, resulted in a decrease in the driving force of 0.016 eV. The ET rate constant in this step was slower by a factor of 1.4, changing from 59 ns in STC to 83 ns for Ru$_{77}$:STC, which is qualitatively negligible in our studies. The small effect is expected, as the Ru-dye is fully exposed to aqueous solution, so that its charge would be significantly screened by the surrounding counter ions and solvation shell. Note that the numerical values reported above are for conformer 4. However, there is little difference in terms of Ru-dye conformations, as can be seen in Figure 4.8, where the four colour-coded lines are essentially on top of each other.

4.5 Extraction of heme-heme ET rate constant from transient absorbance data

As mentioned in section 4.1, decay of the $^3$Ru$_{77}$:STC population was expected to be mono-exponential. However, the TA reveals this was not the case (see Figure 4.2 B magenta circles). To account for the observed behaviour, fits to several modifications and extensions of the basic reaction scheme (Figure 4.2 A) were obtained by solving chemical master equations assuming that the population was initially $^3$Ru$_{77}$:STC. The only scheme to produce a reasonable fit to the $^3$Ru$_{77}$:STC decay was using a multi-exponential decay, i.e., including several kinetically distinct components in the fit. This reaction scheme could be well supported by the existence of different kinetically-separated, long-lived conformers of the Ru-dye as observed from the MD simulation.

The remaining question is whether heme-heme ET was probed in the current experiments; and if the answer is yes, what is its time scale. In the following I will first detail our fitting procedure and then answer this question.

4.5.1 CS and CR without heme-heme ET

In the simplest model as shown in Figure 4.2 A, the triplet excited state $^3$Ru:STC decays via injection of an electron into Ru$^+$:STC$^-$ state with rate constant $k_{CS}$, resulting in the formation
of the charge separated state. The latter recombines to the ground state with rate constant \( k_{CR} \). Charge separation (CS) and charge recombination (CR) are strongly exergonic (see Table 4.1), hence irreversible. Extension to include several kinetically distinct conformers of the Ru-dye, each with its own set of \( k_{CS} \) and \( k_{CR} \) (indicated by a subscript “\( x \)” in the following), leads to the following reaction scheme,

\[
\begin{align*}
3\text{Ru:STC}_x & \xrightarrow{k_{CS}^x} \text{Ru}^+ : \text{STC}^-_x \xrightarrow{k_{CR}^x} \text{Ru:STC}_x, ~ x = a, b, c, d, \ldots
\end{align*}
\]

(4.1)

described by the set of kinetic equations

\[
\begin{align*}
p(3\text{Ru:STC}_x, t) &= \sum_x p_{0,x} \cdot e^{-k_{CS}^x t} \\
p(\text{Ru}^+ : \text{STC}^-_x, t) &= \sum_x p_{0,x} \cdot (e^{-k_{CS}^x t} - e^{-k_{CR}^x t})/(k_{CR}^x - k_{CS}^x) \\
p(\text{Ru:STC}_x, t) &= \sum_x p_{0,x} \cdot (1 + (k_{CS}^x \cdot e^{-k_{CR}^x t} - k_{CR}^x \cdot e^{-k_{CS}^x t})/(k_{CR}^x - k_{CS}^x))
\end{align*}
\]

(4.2)

and \( \sum_x p_{0,x} = 1 \), where \( p_{0,x} \) is the initial triplet state population of conformer \( x \). We have systematically fit the experimental data to three such models comprised of 2, 3 and 4 distinct conformers. For each model the set of rate constants \( k_{CS}^x, k_{CR}^x \) and the initial populations of each conformer \( p_{0,x} \) are determined from a best fit of equation 4.2 to the experimental populations (i.e. 6, 9, and 12 fit parameters for the 2, 3 and 4 conformer models).

First, the populations for \( 3\text{Ru:STC} \) (denoted as A) and \( \text{Ru}^+ : \text{STC}^- \) (denoted as B) were fit separately, giving reasonably good fits with \( R^2 \geq 0.97 \). Then, the average residuals for these fits, \( \sigma_M = \sqrt{\sum_n (p(M, t_n) - f(M, t_n))^2 / N} \), were calculated for \( 3\text{Ru:STC} \) (\( \sigma_A \approx 0.05 \)) and \( \text{Ru}^+ : \text{STC}^- \) (\( \sigma_B \approx 0.005 \)), where \((t_n, p(M, t_n))\) are the experimental populations and \((t_n, f(M, t_n))\) are the fitting populations at time \( t_n \) (\( M = A, B \), \( N \) is total number of data points for each fit). They were used as weights for a simultaneous fit of A and B, in which the weighted sum of the squares of the residuals, \( \sum_n ((p(A, t_n) - f(A, t_n))/\sigma_A)^2 + \sum_n ((p(B, t_n) - f(B, t_n))/\sigma_B)^2 \) was minimized. We used this weighting procedure because the experimental data for the \( \text{Ru}^+ : \text{STC}^- \) population was statistically more certain (less noisy) than that for \( 3\text{Ru:STC} \) and should therefore have a greater weight. Since the kinetical information was fully included in the dynamics of \( 3\text{Ru:STC} \) and \( \text{Ru}^+ : \text{STC}^- \), the TA data for the ground state Ru:STC was not used in the fit. Reassuringly, our best-fit models always predict the recovery of the ground state population in very good agreement with experiment (see Figure 4.9 and Tables 4.3-4.5). The fitting protocol was originally suggested and tested by me and finally carried out by Dr. Katrin Adamczyk.

For the simultaneous fit of A and B, more than 5000 initial sets of fit parameters were
scanned, and each set was refined to minimize (locally) the weighted sum of squares of the residuals. In the case of Ru\textsubscript{77}:STC, the set of fit parameters resulting in the smallest residuals was taken for all models with different numbers of conformers. For Ru\textsubscript{23}:STC, the set of fit parameters resulting in the smallest residuals was taken for the 2-conformer and 4-conformer model. For the 3-conformer model, the set of fit parameters giving the smallest residuals resulted in a ratio of rate constants $k_{CS}/k_{CR} \approx 1000$ that was considered unphysically high. Therefore, this set was discarded and the set with the second smallest residuals was reported instead. For similar reasons, we chose the set of fit parameters with the second smallest residuals for the 2-conformer and 3-conformer model of Ru\textsubscript{10}:STC. For the 4-conformer model of this variant, all sets of fit parameters which resulted in a high-quality fit ($R^2 \approx 0.97$ for the time evolution of both A and B) showed unphysically high rate constants or ratios of rate constants $k_{CS}/k_{CR}$.

We reported one of them. We did not investigated models with more than 4 conformers as they become increasingly prone to overfitting. The fitting results are summarized in Tables 4.3-4.5 (first three lines, “eq. 4.2”) and discussed in subsection 4.5.3.

### 4.5.2 CS and CR with heme-heme ET

The multiple conformer scheme (equation 4.1) was further extended to include reversible heme-to-heme ET between the initial electron accepting heme and the adjacent heme (i.e., heme IV $\leftrightarrow$ heme III for Ru\textsubscript{77}:STC and heme I $\leftrightarrow$ heme II for Ru\textsubscript{23}:STC, Ru\textsubscript{10}:STC). The corresponding reaction scheme becomes,

$$
\begin{align*}
3\text{Ru:STC}_x & \xrightarrow{k_{CS}^x} \text{Ru}^+ : \text{STC}^-_x (CS_i^x) \xrightarrow{k_{CR}^x} \text{Ru:STC}_x, \ x = a, b, c, d, ..., \\
& \xi \xrightarrow{\xi} \xi \\
& \text{CS}_i^x
\end{align*}
$$

where the CS state $\text{Ru}^+ : \text{STC}^-_x$ is replaced by the states $CS_i^x$ and $CS_i^x$ that interconvert with ET rate constants $k_{ij}^x$ for ET from heme $i$ to $j$ and with $k_{ji}^x$ for the reverse, where $(i, j) = (IV,III)$ for Ru\textsubscript{77}:STC and (I,II) for Ru\textsubscript{23}:STC and Ru\textsubscript{10}:STC. Here we assume that the heme-heme ET rate constants are the same for each conformer, $k_{ij}^x = k_{ji}^x$ for all $x$. This is well justified as the different conformers only differ in the conformation of the Ru-dye which has a negligible effect (see discussion in section 4.4) on the heme-to-heme ET as the heme groups are rigidly embedded.
4.5. Extraction of heme-heme ET rate constant from transient absorbance data

The rate constant \( k_{i,j} \) is determined by \( k_{j,i} \) and the driving force for ET assuming detailed balance,

\[
k_{i,j} = k_{j,i} \cdot e^{-F(\epsilon_i(O) - \epsilon_j(O))/k_B T}
\]

where \( \epsilon_i(O) \) are the reduction potentials of heme \( i \) in the all-oxidized state of the protein, taken from experiment [141], and \( F \) is the Faraday constant. Hence, only one additional fitting parameter, \( k_{j,i} \), was added to the multiple conformer mod-
Chapter 4. Photo-Initiated Electron Transfer in Ru(bpy)$_3$-Labelled STC

Figure 4.10: Evolution of the experimentally defined populations of states $^3$Ru:STC (open circles), Ru$^+$:STC$^-$ (filled squares), Ru:STC (open stars) and the sum total (filled triangles) for Ru$_{77}$:STC (A), Ru$_{23}$:STC (B) and Ru$_{10}$:STC (C). Fit to multiple conformer kinetic models without heme-heme ET (eq. 4.2, $n = 2, 3, 4$) and with heme-heme ET (eq. 4.4, $n = 2, 3$) are shown in solid lines with the fit parameters summarized in Tables 4.3-4.5. (D-F) Zoom-in of the corresponding TA data for the charge separated states Ru$^+$:STC$^-$. Figures were originally prepared by Katrin Adamczyk. Adapt and reprint with permission.

Figure 4.10: Evolution of the experimentally defined populations of states $^3$Ru:STC (open circles), Ru$^+$:STC$^-$ (filled squares), Ru:STC (open stars) and the sum total (filled triangles) for Ru$_{77}$:STC (A), Ru$_{23}$:STC (B) and Ru$_{10}$:STC (C). Fit to multiple conformer kinetic models without heme-heme ET (eq. 4.2, $n = 2, 3, 4$) and with heme-heme ET (eq. 4.4, $n = 2, 3$) are shown in solid lines with the fit parameters summarized in Tables 4.3-4.5. (D-F) Zoom-in of the corresponding TA data for the charge separated states Ru$^+$:STC$^-$. Figures were originally prepared by Katrin Adamczyk. Adapt and reprint with permission.

els investigated in subsection 4.5.1: $k_{IV,III} = k_{III,IV} \cdot 0.79$ for Ru$_{77}$:STC and $k_{I,II} = k_{II,I} \cdot 0.07$ for Ru$_{23}$:STC and Ru$_{10}$:STC, at $T = 300$K.

We have adopted the same fitting procedure as described in the previous subsection. Again, we have systematically fit the experimental data comprised of 2 and 3 distinct conformers, and report the set of physical fit parameters that result in the smallest residuals. The fitting results are summarized in Tables 4.3-4.5 (last two lines, “eq. 4.4”) and discussed in the following subsection.

4.5.3 Discussion of model fits

The final rate constants, conformer population and $R^2$ values are summarized in Tables 4.3-4.5 and the corresponding fits to the experimental TA data are shown in Figure 4.10. We found that at least 3 conformers are necessary to model the experimental TA data.

Ru$_{77}$:STC and Ru$_{23}$:STC. For Ru$_{77}$:STC and Ru$_{23}$:STC (see Figure 4.10 A,B), the 4-conformer model only slightly improves the fit. The model with 3 conformers and heme-heme ET gives the best fit of the statistically most certain (least noisy) experimental populations, Ru$^+$:STC$^-$. In particular, the plateau between 0.1 and 1 ns for Ru$_{77}$:STC$^-$ (between 2 and 20 ns for Ru$_{23}$:STC$^-$) and the decay at longer times is very well described compared to the models without the inclusion of heme-heme ET (see Figure 4.10 D,E).
4.5. Extraction of heme-heme ET rate constant from transient absorbance data

Table 4.3: Summary of fitting results for the experimental TA data of Ru77:STC. Equation 4.2 refers to the multiple-conformer model without invoking heme-heme ET, whereas equation 4.4 refers to the multiple-conformer model with heme IV ↔ heme III ET invoked. \( n \) is the number of conformers included in the kinetic model. A, B, C stand for the states \(^1\)Ru:STC, \(^{+}\)Ru:STC, Ru:STC, respectively. The inverse rate constants are given in the unit of ps. This table was originally prepared by Katrin Adamczyk. Adapt and reprint with permission.

<table>
<thead>
<tr>
<th>Model</th>
<th>( R_A^2 )</th>
<th>( R_B^2 )</th>
<th>( R_C^2 )</th>
<th>contribution</th>
<th>1/( k_{CS} )</th>
<th>1/( k_{CR} )</th>
<th>1/( k_{III,IV} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n=2 ), eq. 4.2</td>
<td>0.9749</td>
<td>0.8848</td>
<td>0.9781</td>
<td>( 41% )</td>
<td>48</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>( n=3 ), eq. 4.2</td>
<td>0.9806</td>
<td>0.9663</td>
<td>0.9837</td>
<td>( 38% )</td>
<td>442</td>
<td>113</td>
<td></td>
</tr>
<tr>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>( n=4 ), eq. 4.2</td>
<td>0.9811</td>
<td>0.9706</td>
<td>0.9833</td>
<td>( 34% )</td>
<td>305</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>( n=2 ), eq. 4.4</td>
<td>0.9746</td>
<td>0.9322</td>
<td>0.9760</td>
<td>( 36% )</td>
<td>40</td>
<td>35</td>
<td>3500</td>
</tr>
<tr>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>( n=3 ), eq. 4.4</td>
<td>0.9804</td>
<td>0.9706</td>
<td>0.9831</td>
<td>( 33% )</td>
<td>310</td>
<td>88</td>
<td>11500</td>
</tr>
<tr>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
</tbody>
</table>

In addition to the models described in this section, we also investigated other more complex kinetic models, e.g., including ET along the entire tetra-heme chain, simultaneous electron injection in two hemes, allowing conformer interconversion, but none of these more complicated models improved on the model of equation 4.4 with \( n=3 \). Hence, we conclude that this is our best working model in the sense that it invokes the smallest number of conformers required to explain the experimental TA data for these two Ru-dye labelled STC variants. However, we do not rule out electron transfer further along the STC heme chain, e.g., from heme III to heme II for Ru77:STC. Such process was not included in our kinetic model partly because our predicted ET rate constants between the stacked pair heme II ↔ heme III are approximately an order of magnitude greater than for the T-shaped heme pairs (see Appendix C).

While our main focus is the heme-heme ET, it is apparent that contributions from multiple conformers are required to account for the complete photocycle that extends from 0.1 pico- to 1 milli-second after excitation of the Ru-dye. For conformer \( a \) where charge separation and recombination are fastest and have rate constants much greater than heme-heme ET time scale, the model predicts \(^{+}\)Ru\(_{77}^+\)-STC\(^-\) (\(^{+}\)Ru\(_{23}^+\)-STC\(^-\)) existing only as CS\(^{IV}\) (CS\(^I\)) (see Figure 4.9 C and D, red lines). By contrast, in conformer \( c \) where charge separation and recombination are slowest,
Table 4.4: Summary of fitting results for the experimental TA data of Ru23:STC. Equation 4.2 refers to the multiple-conformer model without invoking heme-heme ET, whereas equation 4.4 refers to the multiple-conformer model with heme I $\leftrightarrow$ heme II ET invoked. $n$ is the number of conformers included in the kinetic model. A, B, C stand for the states $^{3}$Ru:STC, $^{5}$Ru:STC$^{-}$, Ru:STC, respectively. The inverse rate constants are given in the unit of ps. This table was originally prepared by Katrin Adamczyk. Adapt and reprint with permission.

<table>
<thead>
<tr>
<th>Model</th>
<th>$R_A^2$</th>
<th>$R_B^2$</th>
<th>$R_C^2$</th>
<th>contribution</th>
<th>$1/k_{CS}$</th>
<th>$1/k_{CR}$</th>
<th>$1/k_{II,I}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n=2$, eq. 4.2</td>
<td>0.9656</td>
<td>0.8077</td>
<td>0.9787</td>
<td>57%</td>
<td>49</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>43%</td>
<td>12764</td>
<td>1157</td>
<td></td>
</tr>
<tr>
<td>$n=3$, eq. 4.2</td>
<td>0.9688</td>
<td>0.9279</td>
<td>0.9800</td>
<td>27%</td>
<td>17</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32%</td>
<td>129</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>41%</td>
<td>22740</td>
<td>2317</td>
<td></td>
</tr>
<tr>
<td>$n=4$, eq. 4.2</td>
<td>0.9765</td>
<td>0.9726</td>
<td>0.9870</td>
<td>28%</td>
<td>17</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23%</td>
<td>77</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>47%</td>
<td>7509</td>
<td>268</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2%</td>
<td>3509</td>
<td>17100</td>
<td></td>
</tr>
<tr>
<td>$n=2$, eq. 4.4</td>
<td>0.9695</td>
<td>0.8774</td>
<td>0.9829</td>
<td>52%</td>
<td>43</td>
<td>11</td>
<td>5200</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>48%</td>
<td>6469</td>
<td>328</td>
<td></td>
</tr>
<tr>
<td>$n=3$, eq. 4.4</td>
<td>0.9762</td>
<td>0.9715</td>
<td>0.9869</td>
<td>27%</td>
<td>16</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26%</td>
<td>84</td>
<td>32</td>
<td>8000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>47%</td>
<td>8845</td>
<td>462</td>
<td></td>
</tr>
</tbody>
</table>

Heme IV (I) $\rightarrow$ heme III (II) electron transfer is competitive with charge recombination and a notable amount of CS$^{III}$ (CS$^{II}$) is predicted in addition to CS$^{IV}$ (CS$^{I}$) (see Figure 4.9 C and D, blue shaded and blue open respectively).

**Ru$_{10}$:STC.** For Ru$_{10}$:STC, the model with 4 conformers but no heme-to-heme electron transfer significantly improves the fit for all three populations (see Figure 4.10 C). However, with four reacting conformers, the variations in $k_{CS}/k_{CR}$ are $> 10^3$ and, while it is difficult to place an upper limit on this ratio, we consider values exceeding $10^2$ as unlikely. With three reactive conformers and heme-heme ET in the slowest, all conformers have $k_{CS}/k_{CR} < 11$. For this model, the predicted heme I $\rightarrow$ heme II electron transfer rate constant of $143 \times 10^6$ s$^{-1}$ is very similar to the value of $125 \times 10^6$ s$^{-1}$ fit from the TA of Ru$_{23}$:STC. Therefore, we consider these two models produced equally good fits to the transient populations for Ru$_{10}$:STC: equation 4.4 with $n=3$, and equation 4.2 with $n=4$. The corresponding fits and parameters are summarized in Figure 4.10 C and F and Table 4.5.

We note that a previous study [33] reported ultra-fast TA of a tri-heme containing cytochrome, PpcA, photosensitized for light-driven electron transfer in the same way as the STC proteins described here. Charge separated states were generated from photoexcitation of the
Table 4.5: Summary of fitting results for the experimental TA data of Ru\textsubscript{10}:STC. Equation 4.2 refers to the multiple-conformer model without invoking heme-heme ET, whereas equation 4.4 refers to the multiple-conformer model with heme I $\leftrightarrow$ heme II ET invoked. $n$ is the number of conformers included in the kinetic model. A, B, C stand for the states \textsuperscript{3}Ru:STC, Ru\textsuperscript{+}:STC$^-$, Ru:STC, respectively. The inverse rate constants are given in the unit of ps. This table was originally prepared by Katrin Adamczyk. Adapt and reprint with permission.

<table>
<thead>
<tr>
<th>Model</th>
<th>$R_A^2$</th>
<th>$R_B^2$</th>
<th>$R_C^2$</th>
<th>contribution</th>
<th>$1/k_{CS}$</th>
<th>$1/k_{CR}$</th>
<th>$1/k_{II}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n=2$, eq. 4.2</td>
<td>0.9119</td>
<td>0.5074</td>
<td>0.8749</td>
<td>55%</td>
<td>114</td>
<td>13</td>
<td>683</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>45%</td>
<td>4199</td>
<td>65%</td>
<td>23</td>
</tr>
<tr>
<td>$n=3$, eq. 4.2</td>
<td>0.9124</td>
<td>0.9792</td>
<td>0.9108</td>
<td>65%</td>
<td>50</td>
<td>1.5</td>
<td>6184</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5%</td>
<td>27</td>
<td>65%</td>
<td>4199</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30%</td>
<td>81997</td>
<td>65%</td>
<td>162980</td>
</tr>
<tr>
<td>$n=4$, eq. 4.2</td>
<td>0.9709</td>
<td>0.9917</td>
<td>0.9571</td>
<td>46%</td>
<td>23</td>
<td>0.8</td>
<td>439</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5%</td>
<td>18</td>
<td>46%</td>
<td>432</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>47%</td>
<td>4830</td>
<td>47%</td>
<td>432</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2%</td>
<td>3920</td>
<td>2%</td>
<td>162980</td>
</tr>
<tr>
<td>$n=2$, eq. 4.4</td>
<td>0.8459</td>
<td>0.9288</td>
<td>0.8927</td>
<td>44%</td>
<td>66</td>
<td>4%</td>
<td>4700</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>51%</td>
<td>2355</td>
<td>51%</td>
<td>291</td>
</tr>
<tr>
<td>$n=3$, eq. 4.4</td>
<td>0.9648</td>
<td>0.9722</td>
<td>0.9417</td>
<td>28%</td>
<td>18</td>
<td>1.6</td>
<td>6900</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>27%</td>
<td>170</td>
<td>27%</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>45%</td>
<td>3504</td>
<td>45%</td>
<td>440</td>
</tr>
</tbody>
</table>

fully oxidized, and fully reduced, PpcA. Nevertheless, the authors were unable to find evidence for heme-to-heme electron transfer. In view of the results presented here we suggest that this does not preclude light-driven heme-to-heme electron transfer in Ru:PpcA. Rather this process may have gone undetected due to contributions from a number of reactive conformers and/or the relative rates of electron transfer in the corresponding photocycles. As illustrated here, the evidence for heme-to-heme electron transfer is compelling for Ru\textsubscript{77}:STC and Ru\textsubscript{23}:STC but less clear cut for Ru\textsubscript{10}:STC.

**Comparison to simulation.** Comparison of measured and calculated rates for charge separation and charge recombination for Ru\textsubscript{77}:STC (Table 4.6), shows they are in reasonable agreement, typically within an order of magnitude, and gives confidence in our kinetic model for the photochemistry with 3 conformers and heme-to-heme electron transfer. A better agreement between computation and experiment can hardly be expected given that model structures of the Ru(II)tris(bipyridine)-labelled STC are used for the calculations and that charge separation and charge recombination rates are very sensitive to small changes in donor/acceptor distance and orientation.

We find that the computed rate constants are about an order of magnitude smaller than the
Table 4.6: Electron transfer rate constants\(^1\) for the Ru\(_{77}\):STC photocycle. Values are derived from fit to the observed transients and from MD/DFT calculations as described in the text. This table was originally prepared by Julea Butt. Adapt and reprint with permission.

<table>
<thead>
<tr>
<th>contribution %</th>
<th>conformer</th>
<th>a</th>
<th>b</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k_{CS} (10^6 s^{-1}))</td>
<td>Fit</td>
<td>41700</td>
<td>3300</td>
<td>261</td>
</tr>
<tr>
<td></td>
<td>MD/DFT(^3)</td>
<td>3330</td>
<td>66700</td>
<td>256</td>
</tr>
<tr>
<td>(k_{CR} (10^6 s^{-1}))</td>
<td>Fit</td>
<td>30300</td>
<td>11400</td>
<td>1230</td>
</tr>
<tr>
<td></td>
<td>MD/DFT(^3)</td>
<td>7190</td>
<td>6710</td>
<td>588</td>
</tr>
<tr>
<td>(k_{III,IV} (10^6 s^{-1}))</td>
<td>Fit</td>
<td>87</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MD/DFT(^2,4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) \(k_{CS}\), \(k_{CR}\) and \(k_{III,IV}\) are defined by equation 4.4 and fit with \(n=3\). The parameters used for MD/DFT calculations are summarized in Tables 4.1 and 4.2.

\(^2\) Assumed to be the same for all conformers.

\(^3\) Values for \(a\) averaged over two conformers (1 and 2) with the shortest bpy-to-heme V edge-edge distances.

\(^4\) Electronic coupling and reorganisation free energy are taken from Appendix C, driving force \(\Delta A = -10\) meV instead of 6 meV has been used (i.e., shifted up by +0.016 eV, see section 4.4 for details).

Experimental fits, a factor of 7 for heme IV \(\rightarrow\) heme III and a factor of 14 for heme I \(\rightarrow\) heme II ET. A possible reason for the greater discrepancy in the latter is that the reorganisation free energy, 1.08 eV for heme pair I/II, might be overestimated in the calculations. For comparison, the predicted value for heme pair IV/III is 0.88 eV. The error of these estimates due to inaccuracies in the protein and water force field is typically of the order of 0.1 eV [4]. If the reorganisation free energy for heme pair I/II is lowered by 0.1 eV, to 0.98 eV, the deviation with respect to experiment decreases from a factor of 14 to 5. An additional source of the discrepancy could be the presence of the Ru-dye, which could give rise to a slight reduction in the reorganisation free energy for the adjacent heme pair IV/III and heme pair I/II. The terminal hemes, in particular heme I, have a very high solvent accessible surface area in unlabelled STC and the attachment of the bulky and hydrophobic Ru-dye in their vicinity will lower solvent access to the terminal hemes and therefore, most likely, lower reorganisation free energy and increase the ET rate. We do not think this effect is very large but possibly large enough to explain at least part of the remaining difference between computed and experimental rate constants. Taking into account all of these uncertainties, our computed electron hopping rates give fairly good support for the interpretation of the TA data in terms of light-initiated electron transfer across the T-shaped heme pairs heme
4.6. Summary

In this chapter I have presented our joint study with Butt et al. who are experts in protein engineering and ultra-fast pump-probe spectroscopy. The study aims to establish whether ultra-fast spectroscopy could probe heme-heme electron transfer (ET) dynamics in STC and, in turn, inform discussions surrounding the ET mechanism in solvated multi-heme proteins.

A Ru-dye was docked close to either heme I or heme IV and ultra-fast pump-probe spectroscopy was used to monitor electron injection in the protein and subsequent relaxation dynamics. Our molecular dynamics simulations and density functional theory calculations provide a microscopic view of the contributing processes, allowing us to present rate constants for heme I ↔ heme II and heme IV ↔ heme III electron transfer that are indicative of heme-to-heme electron transfer on the 10 ns time scale. The extracted heme-heme ET rate constant is in good agreement with our prediction using the Marcus hopping model.

The evidence for heme-to-heme ET is compelling for two of the three protein variants studied herein but less clear cut for the third one. One reason for this is the highly exothermic charge recombination to the Ru-dye, strongly competing with heme-heme ET. Future experiments to obtain more direct evidence for heme-heme ET through pump-probe spectroscopies of multi-heme cytochromes are highly desired. Possible directions include: engineering the protein to contain a spectroscopically distinct, e.g. HIS/MET coordinated or high-spin heme, and engineering the charge recombination to be less exothermic by, e.g. engineering the reduction potential of the Ru-dye, or putting the Ru-dye further away from the hemes.
Chapter 5

Effect of Electronic Polarizability on
Reorganisation Free Energy

5.1 Electronic polarizability claimed to cause ergodicity-breaking effect

One central assumption in Marcus theory is that the thermal fluctuations of the ET reaction coordinate, vertical energy gap \( \Delta E \) [172,173], are Gaussian distributed and that the phase space is sampled according to Boltzmann distribution on the time scale of the ET reaction. An important consequence is that the horizontal offset of the two free energy parabolas, defined as twice the Stokes reorganisation free energy, \( \lambda^{st} \), is not independent of their curvatures (or force constant \( k \propto 1/\lambda^{var} \), “var” for variance of the corresponding Gaussian fluctuations), but in fact related to it since \( \lambda^{st} \equiv \lambda^{var} \equiv \lambda \) is exact for Gaussian fluctuations (see Figure 5.1 A, pair of parabolas in blue). Hence, only one reorganisation free energy (\( \lambda \)) appears in Marcus theory. In the majority of cases this assumption is fulfilled, which may be rationalized by the central limit theorem even though the energy gap fluctuations due to the molecules of a condensed phase system are not strictly independent.

Reorganisation free energies of heme cofactors span a relatively narrow range, because the reorganisation comes mainly from protein and solvent and the local environment around the hemes only have fairly small contribution. Experimentally, the majority of reorganisation free energies for tunnelling within proteins with cofactors excluded from the aqueous phase are in the range of 0.6-0.9 eV [174], and the reorganisation free energy of the oxidation of solvated horse-heart cytochrome c was measured to be 0.58 eV [175]. However, classical MD usually report reorganisation free energies that strongly overestimate the experimental values [4]. This overestimation is usually attributed to the use of standard, electronically nonpolarizable protein force
114 Chapter 5. Effect of Electronic Polarizability on Reorganisation Free Energy

Figure 5.1: (A) Parabolic free energy profiles for ET in Marcus theory ($\lambda_{\text{st}} = \lambda_{\text{var}}$, blue) and when ergodicity is broken ($\lambda_{\text{st}} < \lambda_{\text{var}}$, red). In the latter case the protein conformational transition between the equilibrium configurations in reduced and oxidized states is too slow to occur on the ET time scale leading to a reduced Stokes shift as indicated by the block arrows. Consequently, the ET activation free energy at zero driving force is reduced by a factor $\kappa^{-2} = (\lambda_{\text{st}}/\lambda_{\text{var}})^2$. (B) ET coupled to slow conformational change of the protein. The initial state is denoted D-A and the final state D$^+$-$A^-$. The reaction is described by a fast collective coordinate $P$ for the protein and solvent modes coupling to ET and a coordinate $q$ for the slow conformational change. ET along $P$ occurs on a faster time scale than protein conformational change along $q$, resulting in the formation of a local minimum B’ rather than the equilibrium state B on the product surface. Adapted with permission from Ref. [48]. Copyright 2013, AIP Publishing LLC.

fields and water models, with perhaps a minor effect from nuclear tunnelling or anharmonic effect [112,176]. However, it has been recently suggested that alternatively, proteins might be able to achieve higher reaction rates by means of ergodicity breaking so that a reaction reorganisation free energy $\lambda^r$ could be defined which lowers the reaction barrier [48, 49].

Naturally, non-ergodicity is common in ultra-fast (picosecond) photoexcited ET reactions, e.g. in Photosystem II (PS II) [48, 118–120]. Reorganisation and hence activation free energy are strongly overestimated if the energy gap is sampled on time scales longer than the actual ET event [118, 119]. To ease this problem, a self-consistent non-ergodicity correction has been devised [48] where frequency components faster than the ET event are simply removed from the
5.1. Electronic polarizability claimed to cause ergodicity-breaking effect

reorganisation free energy computed with ergodic trajectories (please refer to subsection 4.3.1 for an example applying this correction). This method has been successfully combined with Marcus-Sumi theory [121] to explain the non-exponential population decay observed experimentally in PSII [118–120].

In a series of recent papers, Matyushov et al. claims that ergodicity-breaking effects extend well into the regime of thermal biological electron transfer, typically occurring on the microsecond or slower time scales [48–52,120]. In analogy with glass-forming materials, the often highly charged protein-water interface creates a rugged energy landscape that is not explored ergodically on the time scale of the ET event (see Figure 5.1 B). While the energy gap fluctuations may still be Gaussian and arise, amongst others, from protein motions necessary to bring the protein from the equilibrium structure of the initial to the equilibrium structure of the final ET state, the rugged energy landscape of the protein-water interface may not allow for this transition to actually occur on the ET time scale (dotted lines in Figure 5.1 B) keeping the solvated protein trapped in some local minimum on the final ET state surface (arrows in Figure 5.1 B). Due to incomplete relaxation, the Stokes reorganisation free energy $\lambda^{st}$ is now no longer equal to but smaller than the variance reorganisation free energy $\lambda^{var}$, $\lambda^{st} < \lambda^{var}$, amounting to a horizontal shift of the parabolas towards the origin concomitant with a reduction in activation free energy (pair of red parabolas in Figure 5.1 A). Requiring the condition of zero energy gap at the point where the free energy curves cross, it is straightforward to show that in this scenario the activation free energy is still given by the usual Marcus expression but with $\lambda$ replaced by the smaller reaction reorganisation free energy (superscript “r”) [48–52],

$$\Delta A^{\pm,r} = \frac{(\lambda^r + \Delta A^0)^2}{4\lambda^r}$$

$$\lambda^r = \frac{(\lambda^{st})^2}{\lambda^{var}} = \frac{\lambda^{var}}{\kappa_G^2}$$

The ratio $\kappa_G = \lambda^{var}/\lambda^{st}$ characterizes the barrier-lowering effect at zero driving force due to the mismatch. The existence of ergodicity-breaking effects would profoundly change our traditional understanding of thermally activated biological ET reactions. It would mean that certain biological ET reactions occur on physiological time scales because the protein does not have sufficient time to convert between the (global) equilibrium structures of initial and final ET states, and if it had, ET would be too slow in these proteins to support biological function.

Using long molecular dynamics (MD) simulations, Matyushov et al. reported a strong
mismatch between $\lambda^{st}$ and $\lambda^{var}$, with ratios $\kappa_G = 7.9$ for the Cu-containing protein plastocyanin [120], and $\kappa_G = 2.3$ for oxidation of cytochrome c (cyt c) at room temperature [50–52]. For plastocyanine the large values for $\lambda^{var}$ were traced back to an incomplete compensation of the gap fluctuations due to the charged/dipolar protein residues at the protein surface by the gap fluctuations due to the dipoles of the water molecules solvating these residues [49]. In cyt c the origin of ergodicity-breaking was of a more subtle nature: it was traced back to the strong electronic polarizability of the heme-c cofactor in oxidized and reduced states and it disappeared when the heme c cofactor was treated non-polarizable [50].

In their demonstration, the electronic polarization of the inner-sphere QM centre has been described via dipole-field interactions. Electronic structure calculations are carried out for the QM centre in vacuum and the interaction with the electrostatic field generated by the MM atoms is calculated perturbatively via a multipole expansion of the QM centre, truncated at second order,

$$H_{ij}^{(M)} = (\varepsilon_i^{(M)} + Q_{Fe}^{(M)} V_{Fe}) \delta_{ij} - \mu_{ij}^{(M)} \cdot E_{Fe}$$

$$E_M = \text{Min}[EV_i(H^{(M)})], \quad M = R, O,$$

where $H^{(M)}$ is the perturbed Hamiltonian matrix for oxidation state $M$ with matrix elements $H_{ij}^{(M)}$, $\varepsilon_i^{(M)}$ are the unperturbed ground and excited electronic state energies of the isolated cofactor, $Q_{Fe}^{(M)}$ is the total charge of the QM region in redox state $M$ assumed to be located on the Fe atom (i.e., $Q_{Fe}^{(O)} = -1e$, $Q_{Fe}^{(R)} = -2e$, position of the Fe atom is approximated as the QM centre of a heme cofactor), $V_{Fe}$ is the Coulomb potential and $E_{Fe} = -\nabla V_{Fe}$ the corresponding electric field on the Fe atom due to the MM atoms, and $\mu_{ij}^{(M)} = -e \langle \phi_i | r | \phi_j \rangle$ are the transition-dipole-moment matrix elements obtained from QM calculations in gas phase. The potential energy of redox state $M$, $E_M$ in equation 5.4, is the lowest eigenvalue (EV) of $H^{(M)}$.

However, their calculations have been done with a sub-optimal QM model which might lead to over-polarization at the QM-MM boundary. In addition, the electronic structures are computed with semi-empirical method ZINDO/S which is not a robust method in calculating the transition dipoles. In order to thoroughly test the impact of inner- and outer-sphere electronic polarisation on the reorganisation free energies, $\lambda^{st}$ and $\lambda^{var}$ are calculated for oxidation of native horse-heart cyt c from MD simulations run over hundreds of nanoseconds with a variety of different levels of theories. The sensitivity of results has been investigated with regard to: (i) the protein and water force field used (CHARMM vs AMBER), (ii) the inclusion of electronic polarizability for the
redox active heme c cofactor using QM/MM calculations with the QM region treated at the level of density functional theory (DFT) or perturbed matrix method (PMM) [177–181], and (iii) the inclusion of electronic polarizability for protein and water using a polarizable force field with induced atomic dipoles. Cytochrome c is chosen for the study because experimental reorganisation free energies for oxidation in solution are well known and agree among different electrochemical measurements [175,182] and because simulation data from the group of Matyushov are available for comparison [50–52].

As expected, no evidence is found for ergodicity-breaking effects in cyt c on the hundred nanosecond time scale, in contrast to Refs. [50–52]. Both reorganisation free energies are about equal, $\lambda_{st} \approx \lambda_{var} \approx 1.0$ eV without or with inclusion of electronic polarizability of the heme centre, yet significantly overestimate experimental values obtained from electrochemistry, 0.58 eV [175, 182]. Treating the protein and solvent electronically polarizable reduces both reorganisation free energies by $\approx 40\%$ aligning the computed estimate with experiment. Variance reorganisation free energies as large as the ones reported in Refs. [50–52] (close to 3 eV) are only obtained if the electrostatic field at the heme site is scaled by a factor of about 5. Hence, while ergodicity-breaking effects may occur for other biological ET reactions, present simulations suggest that this is not the case in cyt c. In the remainder of this chapter I will present in detail the simulation results followed by some concluding remarks.

5.2 Results and discussion

5.2.1 Reorganisation free energies from vertical energy gaps

The Stokes and variance reorganisation free energies, $\lambda_{st}$ and $\lambda_{var}$, respectively, are defined by equations 2.29 and 2.30. Here the initial and final states are oxidized ($M = O$) and reduced ($M = R$) state of cytochrome c (cyt c). In the limit of linear response (i.e., Gaussian gap fluctuations) and ergodic sampling, $\lambda_{st} = \lambda_{R}^{var} = \lambda_{O}^{var}$ [4]. The thermal averages are obtained by sampling the energy gap $\Delta E$ (equation 2.31) along MD trajectories for oxidized and reduced solvated cyt c. The numerical results are summarized in Table 5.3. The simulation results and the computational procedures will be discussed in detail in the following subsections.

5.2.2 Force fields: CHARMM27 vs AMBER03

The same simulation system is set up following Refs. [50–52] to ensure a fair comparison. The simulation box contains 101441 atoms in total. Simulations were carried out with two protein force fields, CHARMM27 [183], with the same charges and bond parameters [184] as in
Ref. [52], and AMBER03 [59], with heme atomic charges and bond parameters taken from previous simulations in our lab [38, 147]. The topologies for CHARMM27 and AMBER03 were generated with the psfgen tool in VMD [185] and LeaP in AMBER Tools 16 [59], respectively.

MD simulations were carried out with the SHAKE algorithm to constrain O-H bonds in water molecules. Particle mesh Ewald was used for the electrostatics with a real space cutoff of 12 Å and the same cutoff was applied for Van der Waals interactions. The solvated protein in the reduced oxidation state was initially minimized for 5000 steps and subsequently equilibrated for 500 ps with all protein atoms kept frozen. The temperature was rescaled to 300 K every 5000 steps and a Langevin barostat was applied with a target pressure of 1 bar, piston period of 100 fs, and piston decay time of 50 fs. The protein was then slowly released by applying harmonic restraints around the crystallographic positions with decreasing force constants of 99, 25, 1.0, 0.1, and 0.001 kcal mol$^{-1}$ Å$^{-2}$. For each restraining force 500 ps MD simulations in the NPT ensemble were carried out using a 1 fs MD time step. The thermostat damping coefficient was 1 ps$^{-1}$ and the barostat parameters were the same as before. Then all restraints were removed and the protein equilibrated for 10 ns in the NPT ensemble, followed by equilibration of 10 ns in the NVT ensemble using a 2 fs time step. After equilibration, the size of the simulation box was 101.08 Å × 101.08 Å × 101.08 Å. Finally the temperature was decreased to 290 K with a cooling rate of 1K/ns. A production run trajectory of 288 ns (250 ns) was generated in the NVT ensemble at $T = 290$ K using the Langevin thermostat for CHARMM27 (AMBER03). Simulations of oxidized cyt c were initiated from an equilibrated snapshot of the reduced protein. The system was equilibrated for 20 ns in the NVT ensemble at 290 K, followed by a production run of 288.5 ns for CHARMM27 and 250.0 ns for AMBER03, see Figure 5.2 A for a snapshot of the solvated protein. Snapshots were saved with the frequency of 10 ps in the production runs. The energy gap (equation 2.31) was calculated for all the 28,000 (25,000) equidistantly spaced snapshots sampled along each of the CHARMM27 (AMBER03) production trajectories in reduced and oxidized states (sampling frequency = 100 snapshots/NS), as well as for a sub-ensemble of 513 equidistantly spaced snapshots along each trajectory (sampling frequency = 2 snapshots/NS). Reorganisation free energies obtained with the lower sampling frequency reproduced the values obtained for the higher sampling frequency to within 0.05 eV. The sub-ensemble was used for QM(DFT)/MM calculations so as to reduce the computational cost, as will be detailed in subsection 5.2.3. All classical MD simulations were carried out with the NAMD code [58].

The vertical energy gaps along trajectories for reduced and oxidized cyt c are shown in Fig-
Figure 5.2: (A) Snapshot of the simulation box containing cytochrome c in aqueous solution (101441 atoms in total, 1745 protein atoms) and a zoom-in on the protein with the heme cofactor shown in stick representation (Fe, pink; S, yellow; O, red; N, blue; C, cyan) and secondary structure elements in cartoon representation. (B) QM model 1 and (C) QM model 2 in QM/MM calculations. In model 1, the two cysteine linkages and the two axial ligands were capped at the \( \alpha \)-C and \( \beta \)-C bond and saturated with hydrogens (cyan and white spheres, respectively); heme propionates were not included in the QM region. In model 2, deprotonated propionates were included and the peptide bond was capped and saturated with hydrogens. The carbon atoms that were saturated with hydrogens in model 1 are shown as cyan shaded spheres in panel (C), for comparison. Model 2 was used in Ref. [52].

Figure 5.3 A (CHARMM27) and B (AMBER03). They fluctuate relatively stably around their mean values on the simulated time scale of \( \approx 250 \) ns, and the distribution of energy gaps fit Gaussian functions almost perfectly \( (R^2 > 0.999) \) for both force fields. The convergence of reorganisation free energies \( \lambda^{st} \) and \( \lambda^{var} \) with respect to simulation time is shown in Figure 5.3 C and D. While \( \lambda^{st} \) is converged after a few ns, it takes significantly longer (50 ns for CHARMM27, 150 ns for AMBER03) to obtain values for \( \lambda^{var} \) that are close to the final result. This is not surprising as it is well known that mean square fluctuations take longer to converge than the mean. The somewhat abrupt changes in the accumulated average of \( \lambda^{var} \) for AMBER03 are related to short drifts in the energy gap due to rare protein fluctuations.

The RMSD with respect to the crystal structure was stable and reasonably small for AMBER03 trajectories, averaging to about 2.0 Å in reduced and oxidized states. The protein
Figure 5.3: Fluctuations and convergence of the vertical energy gap $\Delta E$ (equation 2.31) for oxidation of solvated cytochrome c. MD trajectories were run with the non-polarizable (np) CHARMM27 ((A), (C), (E)) and AMBER03 force fields ((B), (D), (F)). In (A), (B) the fluctuations and distributions are shown for MD trajectories in the reduced (black) and oxidized (red) states. They are used for the calculation of the accumulated averages of Stokes and variance reorganisation free energies, $\lambda_{st}$ (equation 2.29), $\lambda_{R}^{var}$, $\lambda_{O}^{var}$ (equation 2.30), respectively (panels (C) and (D)). The convergence of the reorganisation free energies with respect to the sampling frequency (SF) of snapshots at constant trajectory length ($\approx 250$-$300$ ns) is shown in panels (E) and (F). Convergence to within 0.05 eV is reached at 2 snapshots/ns (SF2 in Table 5.3).

backbone seems to be more flexible with CHARMM27, the RMSD averaging to about 3.0 Å, which might be due to some slow protein motions during the MD simulation. Similar to Ref. [178], the reversible backbone open-closures are observed from residue THR19 to GLY29 in both CHARMM27 and AMBER03 trajectories, and from residue THR40 to LYS55 only in CHARMM27 trajectories. The heme cofactor was fairly rigid, with RMSD less than 1.0 Å in all trajectories.

The two force fields give almost identical values for $\lambda_{st}$ to within the statistical error bar, 0.95 and 0.93 eV for CHARMM27 and AMBER03, respectively. For CHARMM27 this value is almost perfectly matched by $\lambda_{R}^{var}$ and $\lambda_{O}^{var}$, 0.96 eV for both oxidation states. For AMBER03 the two variance reorganisation free energies differ by 0.2 eV but this difference is equal to the statistical error bar for $\lambda_{O}^{var}$ implying that present simulations on the 100 ns timescale are still
insufficient to fully converge the fluctuations of the energy gap for this force field. However, the average $\lambda_{\text{var}} = (\lambda_{R_{\text{var}}} + \lambda_{O_{\text{var}}})/2 = 0.93$ eV matches perfectly $\lambda_{st}$. Disregarding this issue, it can be concluded that for both force fields the ratio $\kappa_G = \lambda_{\text{var}} / \lambda_{st} = 1.0$ as in Marcus theory and that there is no signature of non-ergodic effects at the level of non-polarizable force field simulations on the time scale of a few 100 ns. This result and the numerical values for reorganisation free energy are similar to the ones reported by Matyushov and co-workers for the CHARMM27 force field [50].

**Figure 5.4:** Histograms of vertical energy gap $\Delta E$ for oxidation of solvated cytochrome c for different computational models and force fields. The histograms generated with the reduced (oxidized) state trajectories are plotted with (without) drop lines. In (A) energy gap calculations at the MM(np), QM(DFT)/MM(np) and QM(PMM)/MM(np) levels were carried out for MD trajectories obtained with the CHARMM27 protein force field and TIP3P water model. In QM/MM calculations QM model 2 was used. In (B) energy gap calculations at the MM(np), MM(pol/np) and QM(DFT)/MM(np) levels were carried out for configurations from MD trajectories obtained with the AMBER03 protein force field and TIP3P water model. In QM/MM calculations QM model 1 was used. In MM(pol/np) calculations the energy gap was calculated with the polarizable AMBER02 force field and POL3 water model. “np” stands for electronically non-polarizable, “pol” stands for electronically polarizable. All of the histograms are generated from the 513-snapshot sub-ensembles to ensure a fair comparison. They can be fit to Gaussian distributions nicely, with $R^2 > 0.97$ for all of the histograms. The fits are shown with thin lines on top of corresponding histograms.

### 5.2.3 Electronic polarizability of heme

**QM(DFT)/MM.** As briefly mentioned in subsection 5.2.2, to reduce the computational effort, the largest equidistant spacing at MM level, or equivalently, the minimum sampling frequency of snapshots along the $\approx 250$ ns MD trajectories is first determined to reproduce the reorganisation free energies obtained for the originally chosen sampling frequency (100 snapshots/ns) to within 0.05 eV. I obtain a minimum sampling frequency of 2 snapshots/ns for both force fields (see Figure 5.3 E and F), which corresponds to 513 snapshots along the $\approx 250$ ns MD trajectories. The distribution of gap energies with 513 snapshots can still be fit to Gaussian functions nicely (see
Figure 5.4, dark red). QM(DFT)/MM calculations of the energy gap were carried out on the 513 equidistantly spaced snapshots taken from each of the classical MD simulations in reduced and oxidized states to probe the effect of electronic polarizability of the QM region on reorganisation free energy.

Two different QM models in QM(DFT)/MM calculations were employed for the CHARMM27 trajectories (models 1 and 2) and only one QM/MM model for AMBER03 trajectories (model 1). Model 1 comprises of the heme ring, axial ligands and side chains excluding the propionates (Figure 5.2 B). The latter were saturated with hydrogen atoms at the positions CAA-CBA and CAD-CBD to CAA-H and CAD-H, respectively. The CB-CA bonds of the cysteine linkages, axial histidine and methionine were cut and saturated with a hydrogen atom to CB-H. A small amount of charge was redistributed from Fe to the -CH$_2$-COO propionate side chains in the MM region to enforce an integer total charge of the QM region. In model 2 (Figure 5.2 C), used in Ref. [50–52], the two propionates were included in their deprotonated form. The bonds of the axial ligands HIS18 and MET80 and of cysteine linkage CYS14 were capped at the backbone C-N position and saturated to C-H. The total charge of the QM region was $0^+1e$ for reduced/oxidized state in model 1 and $-2e/-1e$ in model 2. QM model 2 is not an ideal set-up for QM/MM calculations because the propionates form a salt bridge with neighbouring positively charged residues and both should be either included or excluded from the QM region rather than treated at separate levels of theory. Moreover, the polar amide bonds of the amino acid residues were cut and saturated with hydrogens, instead of the apolar carbon-carbon bonds as is best practice. QM model 1 was designed to remove these issues. However, the reorganisation free energies were not very sensitive to these different set-ups (see Table 5.3), which might be partially rationalized by looking at the frontier orbitals from the DFT calculations in QM/MM. In model 2, although some electron density moves to the two propionates in HOMO-1 and HOMO-2 (see Figure 5.5 E and F) due to the cut of the salt bridges, the main character of the frontier orbitals is still correctly captured. Similar as a histidine-histidine coordinated heme, HOMO-1 and HOMO-2 of this histidine-methionine coordinated heme are quasi-degenerate $d_{\pi}$ orbitals when computed at PBE level.

The QM/MM energy gaps were calculated with the CP2K package [90]. For each structure the electronic potential energies for oxidized and reduced states, $E_O$ and $E_R$ in equation 2.31, were obtained by self-consistent iteration of the Kohn-Sham equations for the QM region in the electrostatic field generated by the fixed point charges of the MM region applying periodic
5.2. Results and discussion

Figure 5.5: Frontier orbitals of a test snapshot in QM/MM calculations with model 1 and model 2. Blue, red correspond to isosurface values of ± 0.025. The upper panel (A,B,C) shows the HOMO, HOMO-1, HOMO-2, respectively, computed with QM model 1, and the lower panel (D,E,F) shows the corresponding orbitals computed with QM model 2. The heme cofactors are shown with stick representation (Fe, pink; S, yellow; O, red; N, blue; C, cyan). See Figure 5.2 and text for model explanation.

boundary conditions. The QM region was described at DFT level with PBE [79] functional, DZVP basis set and GTH atomic pseudopotentials [126]. The spin multiplicities for the reduced and oxidized states were singlet and doublet, respectively. For each snapshot, the QM part was centered in a 30 Å × 30 Å × 30 Å box which guaranteed at least 7.0 Å vacuum padding in each direction. The generalized hybrid orbital method [97] was used to link QM and MM atoms at their boundaries. The MM settings were the same as used in the classical MD simulations above. In both models, the charge distribution for the (non-polarizable) MM atoms was the same for reduced and oxidized states. The wave function gradient was converged to $10^{-5}$ a.u.

The accumulated average of the reorganisation free energies at the QM(DFT)/MM level is shown in Figure 5.6 A (CHARMM27 trajectories) and B (AMBER03 trajectories), red lines. We find that the values are very similar to the ones from classical MD calculations. $\lambda^\text{st}$ slightly decreases and $\lambda^\text{val}$ slightly increases, but the difference between the two reorganisation free energies remains within two (CHARMM27) and one (AMBER03) error bars, giving $\kappa_0 = 1.14$ (QM model 1, CHARMM27 trajectories) and 1.20 (QM model 2, CHARMM27 trajectories) and 1.07 (QM model 1, AMBER03 trajectories). The gap energies still satisfy Gaussian distributions.
Figure 5.6: Accumulating average of reorganisation free energies for oxidation of solvated cytochrome c for different computational models and force fields. The Stokes reorganisation free energy, $\lambda_{st}$ (equation 2.29), is shown in solid lines, and the variance reorganisation free energies, $\lambda_{var_R}$ and $\lambda_{var_O}$ (equation 2.30), in dotted and dashed lines, respectively. In (A) energy gap calculations at the MM(np), QM(DFT)/MM(np) and QM(PMM)/MM(np) levels were carried out for MD trajectories obtained with the CHARMM27 protein force field and TIP3P water model. In QM/MM calculations QM model 2 was used. In (B) energy gap calculations at the MM(np) and QM(DFT)/MM(np) levels were carried out for configurations from MD trajectories obtained with the AMBER03 protein force field and TIP3P water model. In QM/MM calculations QM model 1 was used. In (C) MM(pol/np) calculations are presented where the energy gap was calculated with the polarizable AMBER02 force field and POL3 water on configurations obtained from MD simulations with the AMBER03 protein force field and TIP3P water model. In MM(pol/pol) both energy gap and MD simulations were carried out with the AMBER02 force field and POL3 water model. “np” stands for electronically non-polarizable, “pol” stands for electronically polarizable. Experimental reorganisation free energy from electrochemistry is shown in green [175]. See Table 5.3 for numerical values.
(see Figure 5.4, red). Hence, we do not observe a truly significant increase in the variance reorganisation free energies upon inclusion of heme electronic polarizability at the DFT level of theory. This is in line with simulation data from Ref. [178] ($\kappa_G = 1.26$), but in contrast to the results of Ref. [52], where $\kappa_G$ values as large as 2.3 were reported. In the latter two studies electronic polarizability of the QM region was included by the perturbed matrix method (PMM), which differs in some important aspects from the QM(DFT)/MM method as will be explained in the following.

**QM(PMM)/MM.** Following the protocol of Ref. [50, 52], we use ZINDO/S semi-empirical electronic structure calculations [186] on QM model 2 of the heme cofactor to obtain $\epsilon_i^{(M)}$ and $\mu_{ij}^{(M)}$. QM(PMM)/MM calculations of the energy gap (equation 2.31) with potential energies from equations 5.3-5.4 were carried out on the 28,000 equidistantly spaced snapshots sampled along each of the CHARMM27 production trajectories and the sub-ensemble of 513 snapshots as described previously. For evaluation of the energy gap along the reduced trajectory, the unperturbed ground state and excitation energies $\epsilon_i^{(M)}$ and the transition-dipole-moment matrix elements $\mu_{ij}^{(M)}$ were calculated for reduced ($M = R$) and oxidized ($M = O$) QM model 2 on the reduced crystal structure geometry using ZINDO/S semi-empirical electronic structure calculations, as implemented in Gaussian 16 [186]. These parameters remained unchanged for all the snapshots. For evaluation of the energy gap along the oxidized trajectory, similar calculations for $\epsilon_i^{(M)}$ and $\mu_{ij}^{(M)}$ were carried out on the oxidized crystal structure geometry. The electrostatic potential and field on the Fe atom, $V_{Fe}$ and $E_{Fe}$, were obtained from force field calculations within periodic boundary condition using NAMD [58]. Although ZINDO/S is not expected to be the most reliable method for the calculation of transition dipoles, and potentially better DFT-based approaches are available for large molecules, the goal here is to reproduce the results of Ref. [50, 52], which is why ZINDO/S was used in the present calculations. All the PMM calculations presented herein were carried out by Zdenek Futera, who collaborated on this project.

The reorganisation free energies obtained from QM(PMM)/MM are shown in Figure 5.6 A (brown lines). The energy gap distributions are again well described by Gaussians (see Figure 5.4 A, brown) in agreement with previous QM(PMM)/MM calculations of horse-heart cyt c [180]. Both $\lambda^{st}$ and $\lambda^{var}$ match very well the values obtained from present MM and QM(DFT)/MM calculations; we obtain $\kappa_G = 1.20$ (QM model 2, CHARMM27 trj) which is virtually identical with the result from QM(DFT)/MM calculations. A large increase in $\lambda^{var}$ as reported in Ref. [52] is not observed even though the same system set-up is now used. Similarly
Figure 5.7: Gap energy auto-correlation function (ACF, red curve) and its time integral (blue curve) as obtained for oxidation of cytochrome c in reduced (Red) and oxidized (Ox) states. The gap energies were evaluated at QM(PMM)/MM level of theory on 280 ns trajectories with sampling frequency 10 ps. ACFs decay to zero within 20 ns (panels A and C) and their values stay negligibly small beyond that time (not shown for clarity). The correlation time obtained by ACF integration within the first 20 ns is 57.4 (61.8) ns for reduced (oxidized) system, respectively. Initial decay of ACFs, shown in panels B and D, proceeds rapidly within 500 (150) ps for reduced (oxidized) system. This figure was prepared by Zdenek Futera. Reprint with permission.

large electronic polarizabilities are obtained for the heme cofactor as in Ref. [50, 52] (54 Å³ and 27 Å³ for the reduced and oxidized states when including 100 electronic states for construction of $H^{(M)}$), but they do not lead to a significant enhancement of the energy gap fluctuations. The reason is that the electrostatic field at the Fe site is relatively modest (0.05-0.5 V/Å) in our simulations so that the off-diagonal terms $\mu_{ij}^{(M)} \cdot E_{Fe}$ in equation 5.3 are small compared to the diagonal energy differences $\epsilon_i^{(M)} - \epsilon_j^{(M)}$ resulting in little mixing between the states. Indeed, if polarizability is entirely neglected and the second term in equation 5.3 is set to zero, $\lambda^{var}$ decreases by no more than 0.05 eV. To obtain values for $\lambda^{var}$ as large as the ones reported in Ref. [50, 52] (2-3 eV), the electrostatic field due to protein and water would need to be scaled by a factor of about 5 in our simulations. While the heme polarizability converges very slowly with the number of
5.2. Results and discussion

Table 5.1: Convergence of electronic polarizability $\alpha$ of the electronic ground state (index 0) of reduced (R) and oxidized (O) heme c cofactor with number of excited states $N$. The molecular polarizability tensor was computed using perturbative formula $\alpha_{ab} = 2 \sum_{j>0} \mu_{0j} \mu_{bj}/\Delta \epsilon_j$, where transition dipoles $\mu_{0j}$ and excitation energies $\Delta \epsilon_j = \epsilon_j - \epsilon_0$ were obtained from ZINDO/S electronic-structure calculations performed on model 2 using the Gaussian 16 software [186]. Isotropic polarizabilities $\alpha = \frac{1}{3} \sum_a \alpha_{aa}$ are shown together with their differences $\Delta \alpha = \alpha_R - \alpha_O$, both on reduced and oxidized crystal structures. All values are in Å$^3$. This table was prepared by Zdenek Futera. Reprint with permission.

<table>
<thead>
<tr>
<th>$N$</th>
<th>Reduced structure</th>
<th>Oxidized structure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_R$</td>
<td>$\Delta \alpha$</td>
</tr>
<tr>
<td>10</td>
<td>3.99</td>
<td>2.31</td>
</tr>
<tr>
<td>25</td>
<td>38.48</td>
<td>10.77</td>
</tr>
<tr>
<td>50</td>
<td>50.35</td>
<td>20.15</td>
</tr>
<tr>
<td>75</td>
<td>52.68</td>
<td>23.47</td>
</tr>
<tr>
<td>100</td>
<td>53.91</td>
<td>27.28</td>
</tr>
<tr>
<td>250</td>
<td>62.36</td>
<td>35.89</td>
</tr>
<tr>
<td>500</td>
<td>67.57</td>
<td>40.48</td>
</tr>
</tbody>
</table>

Table 5.2: Convergence of reorganisation free energies $\lambda$ with number of excited states $N$ used in QM(PMM)/MM(np) approach. The values were obtained from energy gap calculations on 513 MD samples (sampling frequency 2) of reduced/oxidized cytochrome c with CHARMM27 and TIP3P water model. The effect of neglecting excited-to-excited state transition dipoles ("$\mu_{0k}$ only") is shown for comparison. All values are in eV. This table was prepared by Zdenek Futera. Reprint with permission.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\lambda_{\text{var}}^\text{R}$</th>
<th>$\mu_{0k}$ only</th>
<th>$\lambda^\text{var}_O$</th>
<th>$\lambda^\text{st}$</th>
<th>$\lambda_{\text{var}}^\text{R}$</th>
<th>$\lambda^\text{var}_O$</th>
<th>$\lambda^\text{st}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.07</td>
<td>1.24</td>
<td>0.99</td>
<td>1.07</td>
<td>1.24</td>
<td>0.99</td>
<td>1.24</td>
</tr>
<tr>
<td>25</td>
<td>1.17</td>
<td>1.35</td>
<td>1.02</td>
<td>1.15</td>
<td>1.34</td>
<td>1.01</td>
<td>1.34</td>
</tr>
<tr>
<td>50</td>
<td>1.16</td>
<td>1.33</td>
<td>1.02</td>
<td>1.12</td>
<td>1.31</td>
<td>1.01</td>
<td>1.31</td>
</tr>
<tr>
<td>75</td>
<td>1.18</td>
<td>1.33</td>
<td>1.03</td>
<td>1.13</td>
<td>1.31</td>
<td>1.01</td>
<td>1.31</td>
</tr>
<tr>
<td>100</td>
<td>1.17</td>
<td>1.33</td>
<td>1.02</td>
<td>1.12</td>
<td>1.31</td>
<td>1.01</td>
<td>1.31</td>
</tr>
<tr>
<td>250</td>
<td>1.16</td>
<td>1.31</td>
<td>1.02</td>
<td>1.12</td>
<td>1.30</td>
<td>1.01</td>
<td>1.30</td>
</tr>
</tbody>
</table>

electronic states used to construct $H^{(M)}$ (see Table 5.1), $\lambda_{\text{var}}$ is converged after including only the ten lowest states (see Table 5.2). This gives credence to the QM(PMM)/MM calculations of Ref. [178] where 13 states were included. The energy gap auto-correlation function shows the typical signatures for solvated redox proteins (see Figure 5.7), a sharp initial decay on the 100 ps time scale due to relaxation of bulk solvent and amino acid side chains, and a longer decay to zero on the 10 ns time scale characteristic of the slower backbone motions. The calculated correlation time is around 60 ns.
5.2.4 Electronic polarizability of protein and water

Finally, the effect of electronic polarizability of the MM region (protein and water) on reorganisation free energies has been investigated. To this end, the energy gap was calculated with the polarizable AMBER02 force field \[59\] and POL3 water model \[75\] on the same set of structures for which QM(DFT)/MM calculations were carried out (i.e. AMBER03 + TIP3P water trajectories). The partial charges of the heme cofactors were scaled back from their effective (larger) values in the condensed phase to the smaller values in the gas phase. Electronic polarizability of the MM atoms was modelled by atomic and isotropic induced dipoles, while the polarizability for the atoms of the QM region was set to zero. This choice was made to probe the effect of the polarizability of the MM region (outer-sphere) only. The induced dipoles were iterated until successive estimates agreed to within \(10^{-4}\) Debye in the root-mean-square sense. All calculations with polarizable force fields including the MD simulations below were carried out with the Sander program in AMBER Tools 16 \[59\].

The distributions of the energy gap fluctuations are again well-described by Gaussians (see Figure 5.4 B, purple). The reorganisation free energies are shown in Figure 5.6 C (purple lines), which decrease dramatically from 0.93 eV for non-polarizable MM calculations to \(\lambda^\text{st} = 0.69\) and \(\lambda^\text{var} = 0.55\) eV. These values are in good agreement with the experimental estimate from electrochemistry, 0.58 eV \[175\].

In previous work on electron self-exchange between Ru-aqua ions a similar reduction in reorganisation free energy was obtained upon treating the solvent electronically polarizable, but this reduction was partly offset when both MD and energy gap calculations were consistently carried out with the same polarizable force field due to differences in the solvation structure between non-polarizable and polarizable water models \[187\]. Therefore, the effect of outer-sphere electronic polarizability on reorganisation free energies were tested further with both MD simulations and energy gap calculations carried out with the polarizable AMBER02 force field and POL3 water model.

For the MD simulation with explicitly polarizable force fields, the initial structure for the oxidized state was extracted from a snapshot of the previous AMBER03 production run trajectory in the oxidized state. Due to increased computational cost in the Ewald summation in the polarizable MD, the convergence of reorganisation free energy with size of the solvation shell was checked at the non-polarizable MD level and the system size was reduced by decreasing the number of water molecules to 6544. The system with reduced number of water molecules
was initially equilibrated for 10 ns in the NPT ensemble with the non-polarizable AMBER03 force field and TIP3P water and then equilibrated with the polarizable AMBER02 force field and POL3 water: 500 ps in the NVT ensemble, 2.5 ns in the NPT ensemble to 290 K and 1.013 bar using a Langevin barostat with piston period 100 fs, piston decay time and thermostat damping coefficient $15.0 \text{ ps}^{-1}$, and finally 5 ns in the NVT ensemble at a fixed cell size of $59.5 \AA \times 59.4 \AA \times 59.4 \AA$. Instead of solving the induced dipoles iteratively at each step, the Car-Parinello scheme was adopted for efficient MD propagation, wherein each dipole was assigned a fictitious mass of 0.33 a.u. Production runs were carried out for 50 ns with a MD time step of 1 fs. Polarizable MD simulation of the reduced protein was carried out similarly with a starting structure extracted from a snapshot of the previous AMBER03 production run trajectory in the reduced state. The energy gap (equation 2.31) was calculated with the polarizable AMBER02 force field [59] and POL3 water model [75] on 10 ps equidistantly spaced snapshots taken from each of the 50 ns MD trajectories run for reduced and oxidized states with the AMBER02 force field [59] and POL3 water.

The reorganisation free energies decreased slightly further to $\lambda_{\text{st}} = 0.53 \text{ eV}$ and $\lambda_{\text{var}} = 0.46 \text{ eV}$, corresponding to a decrease of 43% and 50% compared to the values from non-polarizable MM calculations. This means that in the current simulations, the reduction in reorganisation free energies due to the smaller partial charges used in polarizable force field is not compensated for by the induced dipoles and/or additional structural relaxation in response to the introduction of induced dipoles, resulting in a significant net reduction in reorganisation free energy. If the inner-sphere contribution of the heme cofactor, here described at the MM level, is replaced by a QM description ($\lambda_i = 0.025 \text{ eV}$ for isolated heme c cofactor in vacuum [165]), the final result is very close to experiment, $\lambda_{\text{st}} = 0.56 \text{ eV}$ and $\lambda_{\text{var}} = 0.49 \text{ eV}$, while $\kappa_G$ remains close to unity.

Although switching from a non-polarizable to a polarizable force field is not equivalent to simply changing the dielectric constant of the medium, the importance of outer-sphere electronic polarizability may be explained by continuum theory, which predicts that outer-sphere reorganisation free energy is proportional to the inverse of the Pekar factor, $\lambda_o \propto 1/\epsilon_{\text{op}} - 1/\epsilon_s$. For aqueous cyt c, the optical dielectric constant $\epsilon_{\text{op}}$ may be estimated from experimental data to be 1.84 [165] while for non-polarizable force fields $\epsilon_{\text{op}} = 1$. Hence, assuming $\epsilon_s \gg \epsilon_{\text{op}}$ (which should be fulfilled for solvated proteins), the reorganisation free energy is predicted to be overestimated by a factor $\approx 1/1.84$ or 46% if protein and water are not treated electronically
polarizable. This agrees very well with present simulation results, even though for other proteins a somewhat smaller effect of 30-40% was typically found [4].

The reduction in reorganisation free energy upon inclusion of electronic polarizability of the outer sphere is in line with a bulk of existing simulation data for electron transfer and oxidation reactions in proteins [4, 137, 147, 165, 188] and aqueous transition metal ions [111, 164, 187, 189, 190]. It has been shown previously that MD reorganisation free energies in Ru-modified proteins are strongly overestimated compared to the experimental values by Gray and Winkler if electronic polarizability of protein and water is not explicitly accounted for but in much better agreement if included [4, 147, 187, 188]. Along the same lines, the reorganisation free energies for oxidation of simple transition metal ions in liquid water, such as aqueous Ru(bpy)$_2^{2+}$ [164] and Mn$^{2+}$ [189], were obtained in excellent agreement with experimental values from liquid jet photo-emission spectroscopy by Winter and co-workers [164, 189] when the same polarizable POL3 water was used as in present protein simulations.

5.3 Conclusion

In this chapter, the effect of electronic polarizability on reorganisation free energies has been thoroughly investigated in the example of oxidation of cytochrome c, with a variety of state-of-the-art methodologies ranging from non-polarizable MD simulations, QM/MM calculations with polarizable QM centre at DFT and PMM level and polarizable MD simulation. According to our calculations, the overestimation (40-50 %) of experimental reorganisation free energies, typically obtained with non-polarizable force fields, is due to missing explicit electronic polarisation in the protein and water force fields used, rather than the ergodicity-breaking hypothesis advocated by Matyushov and co-workers. Similar values for Stokes and variance reorganisation free energies are obtained, $\lambda_{st} \approx \lambda_{var}$, with different levels of theory on the 250 ns time scale.

This conclusion is in line with all the MD simulations carried out previously in our lab of similar heme and Cu-containing redox proteins on shorter time scales (typically on 10-50 ns): close agreement between $\lambda_{st}$ and $\lambda_{var}$ ($\kappa_G \approx 1$) was obtained for Ru-modified cyt c [147, 165] and b$_5$ [147], multi-heme cytochromes STC [138], NtrfB [146], MtrF [137] and MtrC [146], cytochrome c oxidase [188], the blue Cu-protein azurin [4] as well as a porphyrine-binding four-helix bundle protein [147] (for the latter, $\kappa_G \approx 3.7$ in the early study [191] reversed to $\kappa_G \approx 1.1$ when protein and water were treated electronically polarizable [147]). However, one may argue that those 10-50 ns simulations could only capture nanosecond or faster protein motions, and ergodicity-breaking protein motions might occur at longer times that are closer to the time scale...
of ET in these proteins, typically microseconds. The current simulations at 250 ns time scale indicate that ET in solvated cyt c is an ergodic process and well described by linear response, i.e., Gaussian solvation theory. In other words, cytochrome c is a classic Marcus system.

Table 5.3: Computed reorganisation free energies for oxidation of solvated horse-heart cytochrome c (all values in eV).$^a$

<table>
<thead>
<tr>
<th>CHARMM FF</th>
<th>MM(np)$^b$</th>
<th>QM(DFT)/MM(np)$^c$</th>
<th>QM(PMM)/MM(np)$^d$</th>
<th>Ref [178]$^f$</th>
<th>Ref [52]$^i$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SF1$^j$</td>
<td>SF2$^j$</td>
<td>SF1$^j$</td>
<td>SF2$^j$</td>
<td>SF1$^j$</td>
</tr>
<tr>
<td>$\lambda^{ab}$</td>
<td>0.95 ± 0.01</td>
<td>0.95 ± 0.01</td>
<td>0.89 ± 0.01</td>
<td>0.86 ± 0.01</td>
<td>1.00 ± 0.01</td>
</tr>
<tr>
<td>$\lambda_{\text{var}}^{R}$</td>
<td>0.92 ± 0.06</td>
<td>0.96 ± 0.08</td>
<td>0.90 ± 0.08</td>
<td>0.93 ± 0.07</td>
<td>1.23 ± 0.08</td>
</tr>
<tr>
<td>$\lambda_{\text{var}}^{O}$</td>
<td>0.92 ± 0.03</td>
<td>0.96 ± 0.06</td>
<td>1.03 ± 0.06</td>
<td>1.03 ± 0.08</td>
<td>1.24 ± 0.07</td>
</tr>
<tr>
<td>$\kappa_G^{c}$</td>
<td>0.97</td>
<td>1.01</td>
<td>1.14</td>
<td>1.20</td>
<td>1.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AMBER FF</th>
<th>MM(np)$^b$</th>
<th>QM(DFT)/MM(np)$^c$</th>
<th>MM(pol/np)$^d$</th>
<th>MM(pol/pol)$^d$</th>
<th>SF1$^j$</th>
<th>SF2$^j$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SF1$^j$</td>
<td>SF2$^j$</td>
<td>SF1$^j$</td>
<td>SF2$^j$</td>
<td>SF1$^j$</td>
<td></td>
</tr>
<tr>
<td>$\lambda^{ab}$</td>
<td>0.92 ± 0.01</td>
<td>0.93 ± 0.01</td>
<td>0.89 ± 0.01</td>
<td>0.69 ± 0.01</td>
<td>0.53 ± 0.01</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{\text{var}}^{R}$</td>
<td>0.80 ± 0.04</td>
<td>0.80 ± 0.05</td>
<td>0.92 ± 0.06</td>
<td>0.46 ± 0.03</td>
<td>0.39 ± 0.03</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{\text{var}}^{O}$</td>
<td>1.02 ± 0.16</td>
<td>1.06 ± 0.17</td>
<td>0.98 ± 0.13</td>
<td>0.63 ± 0.10</td>
<td>0.52 ± 0.04</td>
<td></td>
</tr>
<tr>
<td>$\kappa_G^{c}$</td>
<td>0.99</td>
<td>1.00</td>
<td>1.07</td>
<td>0.79</td>
<td>0.86</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ The experimental reorganisation free energy is 0.58 eV [175]. The statistical uncertainties of mean and variance of the computed energy gap (equation 2.31) are estimated by the uniformly minimum variance unbiased estimator [192], $u(\langle \Delta E \rangle) = \sigma / \sqrt{n/s}, u(\delta\Delta E^2) = \sigma^2 \sqrt{2/(n/s - 1)},$ where $n$ is the number of data points, $\sigma$ is the standard deviation and $s$ the statistical inefficiency. The statistical uncertainties for the reorganisation free energies are estimated by Gaussian error propagation.

$^b$ Stokes reorganisation free energy (equation 2.29).

$^c$ Variance reorganisation free energies (equation 2.30).

$^d$ $\kappa_G = (\lambda_{\text{var}}^R + \lambda_{\text{var}}^O) / (2\lambda^{ab})$.

$^e$ Energy gap and MD calculations with electronically non-polarizable (np) force fields, CHARMM27 or AMBER03 and TIP3P water.

$^f$ Sampling frequency (SF) for energy gap calculations = 100 snapshots/ns.

$^g$ SF for energy gap calculations: 2 snapshots/ns.

$^h$ QM(PBE)/MM calculations on trajectories sampled with CHARMM27 or AMBER03 and TIP3P water, respectively.

$^i$ See Figure 5.2 B for a description of QM model 1.

$^j$ See Figure 5.2 C for a description of QM model 2.

$^k$ QM(PMM)/MM calculations on trajectories sampled with CHARMM27 and TIP3P water at $T = 290K$, respectively.

$^l$ QM(PMM)/MM calculation for yeast cytochrome c using trajectories sampled with gromos96 and SPC water at $T = 300K$, respectively. Numerical values were taken as reported in Ref. [50]. For horse-heart cytochrome c a value of $\lambda^{ab} = 0.89$ eV was reported in Ref. [180], but no values for variance reorganisation free energy.

$^m$ Energy gap calculations with electronically polarizable AMBER02 force field and POL3 water on trajectories sampled with non-polarizable AMBER03 force field and TIP3P water.

$^n$ Electronically polarizable AMBER02 / POL3 force-field MD and $\Delta E$ calculations.
Chapter 6

Conclusion and Outlook

To conclude, in my PhD work I have used density functional theory and molecular dynamics simulation to calculate all heme-to-heme electron transfer (ET) rate constants in three ubiquitous multi-heme proteins from bacteria *Shewanella* binding 4 and 10 heme cofactors. The electronic coupling matrix elements, computed with our extended QM models, are typically a few meV. Reorganisation free energies for the heme pairs, properly corrected for the missing explicit polarisation in the force fields used, are usually slightly less than 1 eV. According to literature, Marcus theory is well-established to describe electron transfer in this regime, where reorganisation free energies are much greater than the electronic couplings. Indeed, the heme-heme ET rate constants extracted from the pump-probe measurements are in good agreement with our predicted time scale using Marcus theory.

The results presented in my PhD work reveal that electron hopping through these multi-heme proteins is strongly enhanced by cysteine side chains inserted in the space between heme groups. We believe this to be a general design principle in this family of proteins for acceleration of ET steps that would otherwise be too slow for biological respiration. With the inclusion of these side chains in our calculations, the heme-heme ET rate constants in these three multi-heme proteins are predicted to fall in the range $2 \times 10^4 - 5 \times 10^9$ s$^{-1}$ (on the time scales of 100 ps - 10 µs). Using those ET rate constants, the maximum, protein-limited electron flux for all possible electron flow directions in these three proteins has been estimated. We found that MtrC and MtrF conduct electrons roughly equally well along their main axis and in perpendicular directions with little or no directional bias. Hence, we speculate that the higher expression levels of MtrCAB relative to MtrFDE at low oxygen concentrations is more likely related to a genetic origin rather than electron transfer function.

I would like to conclude by commenting on some directions for future work.
Direct evidence for heme-heme ET through pump-probe spectroscopies. Our collaborating study reported in Chapter 4 has demonstrated the ability of ultra-fast spectroscopy to probe heme-heme electron transfer dynamics in multi-heme proteins and, in turn, inform discussions surrounding the ET mechanism in solvated multi-heme proteins. However, the probed population that undergoes heme-heme ET and forms the second charge separated state is very small, due to highly exergonic nature of direct charge recombination to the Ru-dye. Moving forward, experiments exploring the opportunities to obtain direct evidence for heme-heme electron transfer through pump-probe spectroscopies of multi-heme cytochromes containing spectroscopically distinct, e.g. histidine-methionine coordinated or high-spin heme, at a defined location, are ongoing in our collaborating laboratories.

Electron flux of MtrCAB in realistic biological context. As briefly mentioned in the introduction, the protein complex MtrCAB, which consists of two deca-heme proteins MtrC and MtrA, forms the central extracellular electron transfer component in bacteria *Shewanella*. Once the crystal structures of MtrA and MtrB are available, it would be possible to estimate the electron flux along the entire MtrCAB complex, in the most realistic context possible, i.e., embedded in the membrane bilayer. Foreseeing the size of this system, atomistic molecular dynamics simulation might become infeasible. Instead, ET parameters would have to be estimated from empirical formulae and continuum models. Potential challenges in this direction might be the system set-up, and the choice of reliable modelling methods.

Electron transfer between multi-heme protein and insoluble substrate. This would be the last step of the anaerobic respiration in *Shewanella*, where the electron transfers from multi-heme protein to insoluble substrate, e.g., iron oxide. In the proteoliposome experiment of MtrCAB, different ET rates were measured for the electron transport across the lipid bilayer to different externally located mineral substrates [41], which indicates that the interfacing ET step might be slower than heme-heme ET steps. Obtaining an upper limit for this ET step would be insightful, both biologically and mechanistically. However, properly describing the interaction between protein and solid substrates will be challenging, both for the stable binding of the protein on the substrate, and the transient interaction.

I-V curve modelling. Although the STM-measured a few 0.1 nA at 0.5 V bias voltage could be achieved with the extended electronic coupling model, the shapes of the experimental $I - V$ curves were not fully captured, especially at high bias. On the other hand, the coherent tunnelling model, which assumes that the heme states are not involved in the conductive channel, produced
a perfect fit of the temperature independent $I - V$ curves measured in protein junctions, but can not explain the three orders of magnitude higher conductance measured in multi-heme protein STC compared to Azurin, another redox protein of similar size. Apart from the lack of a reliable structural model of the protein adsorbed on the electrode in the current study, additional mechanism, possibly between these two limits, might be relevant in the electron transport events of protein junctions. The mechanistic difference between electron transfer in solvated multi-heme proteins and electron transport in ambient protein junctions, remains to be understood from an atomistic point of view. This may also pose challenges to existing simulation methods.

**Reorganisation free energy of cytochrome c computed with QM/MM(pol).** In Chapter 5 we have systematically studied the effect of electronic polarizability on the calculation of reorganisation free energies, by first testing the electronic polarizability of the inner sphere (heme cofactor) via QM/MM, and then testing the electronic polarizability of the outer sphere (protein and water) via polarizable MD. However, the case of mutual polarisation, i.e., simultaneous inclusion of inner-sphere and outer-sphere electronic polarizability, has not been investigated. Since QM/MM with the polarizable embedding has been an area of active development in recent years, there might be software package that can handle our system already out in this community, although a further investigation does not fit in the timeline of my PhD. Since the inner-sphere contribution to the reorganisation free energy is small, we do not expect this effect to be large enough to change the conclusion reached herein. Nevertheless, this is certainly worth revisiting.
Appendix A

Derivation of non-adiabatic Marcus rate expression

Marcus electron transfer (ET) rate expression is a generalization of transition state theory in non-adiabatic situation incorporating Landau-Zener (LZ) transition probability at the transition point [104].

Landau-Zener transition probability states that: for two diabatic states $a$ and $b$, assuming the system is in state $a$ at $t = -\infty$, then the probability of finding the system in state $b$ at $t = \infty$ is:

\[
P_{LZ} = 1 - \exp\left(-\frac{2\pi |H_{ab}|^2}{\hbar |d\frac{d(E_a - E_b)}{dt}| R^*}\right) \quad (A.1)
\]

where $H_{ab} = \langle \psi_a | H | \psi_b \rangle$ is the electronic coupling of the two states, $R^*$ is the nuclear coordinate at crossing point, $E_a$ and $E_b$ are energies in state $a$ and $b$.

Marcus considers ET rate as an integral of three terms:

\[
k_{ET} = \int_{0}^{\infty} dp P(x^*, p) \delta x \cdot \frac{p}{m\delta x} \cdot P_{LZ}(p) \quad (A.2)
\]

where $x$ denotes the reaction coordinate, $P(x^*, p) \delta x$ is the probability of being in the transition state (TS) region $\delta x$ with a given momentum $p > 0$, $\frac{p}{m\delta x}$ is the flux through TS region, $P_{LZ}(p)$ is the LZ probability of transition from state $a$ to state $b$ in the TS region.

Assuming quadratic initial and final state free energy surfaces (i.e., harmonic approximation with oscillating frequency $\omega_0$) with minima shifted by $\Delta x$ and minimum energy shifted by $\Delta E_m$,

\[
E_a = \frac{1}{2} m \omega_0 x^2, \quad E_b = \frac{1}{2} m \omega_0 (x + \Delta x)^2 + \Delta E_m, \quad (A.3)
\]
Appendix A. Derivation of non-adiabatic Marcus rate expression

The reorganisation free energy $\lambda$ is defined as the energy needed to make a displacement $\Delta \vec{x}$ from the minimum of one state to the minimum of the other state while staying on the same free energy surface:

$$\lambda = \frac{1}{2} m \omega_0 \Delta \vec{x}^2 \quad \text{(A.4)}$$

In the harmonic approximation LZ probability is given by:

$$P_{LZ} = 1 - \exp\left(-\frac{2 \pi H_{ab}^2}{\hbar \omega_0 (2 \lambda m^{-1})^{1/2} |p|}\right) \quad \text{(A.5)}$$

Replacing $|p|$ in equation A.5 by the average positive momentum $\langle p_+ \rangle = (2 \pi k_B T/m)^{-1/2}$ and assuming non-adiabatic limit (i.e., weak coupling limit, $2 \pi^{3/2} H_{ab}^2 \ll \hbar \omega_0 (\lambda k_B T)^{1/2}$), the exponent in the LZ factor can be Taylor expanded to first order giving the following expression:

$$P_{LZ} \approx \frac{2 \pi^{3/2} H_{ab}^2}{\hbar \omega_0 (\lambda k_B T)^{1/2}} \quad \text{(A.6)}$$

Inserting equation A.6 to equation A.2 and integrating the expression analytically gives the rate expression equation 2.18.
Appendix B

Derivation of analytic solution to chemical master equation: example of STC

The analytic solution was proposed by Dr. Guido Falk von Rudorff. Here is an example for solving the maximum, steady-state flux (i.e., equation 3.5) of the tetra-heme protein STC with this analytic approach.

To proceed with, assuming that electron injection into the protein at heme 1 with rate constant $k_{\text{in},1}$ and electron ejection from the protein at heme 4 with rate constant $k_{\text{out},4}$ are irreversible, that is

$$k_{\text{in},1} = k_{\text{out},4} = 0 \quad (B.1)$$

Under these conditions the electron flow is in the forward direction from heme 1 to heme 4. Requiring steady state (equation 3.7) gives the following recursive relationship for the steady-state populations,

$$P_{i+1} = [k_{i+1,i} P_i - k_{1,in}(1 - P_1)]/[k_{i+1,i} + P_i (k_{i+1,i} - k_{i,i+1})] \quad (B.2)$$

Starting from the initial equation $J \equiv J_{1,in} = k_{1,in}(1 - P_1)$, the population $P_1$ is expressed as,

$$P_1 = 1 - \frac{J}{k_{1,in}}. \quad (B.3)$$

$P_1$ is then inserted into equation B.2 to obtain

$$P_2 = \frac{-Jk_{21} - Jk_{1,in} + k_{21}k_{1,in}}{Jk_{12} - Jk_{21} + k_{21}k_{1,in}} \quad (B.4)$$

Sequentially, $P_3$ and $P_4$ could be expressed as a function of $J$ by inserting $P_1$ to equation B.2.
Substituting $P_4$ in the equation $P_4(J)$ by the last equation for steady state flux, $J = k_{out,4}P_4$, a polynomial equation in $J$ of degree 3 is obtained,

$$
0 = J^3 (k_{12}k_{34} - k_{12}k_{43} - k_{21}k_{34} + k_{21}k_{43}) + J^2 (k_{12}k_{23}k_{34} + k_{12}k_{23}k_{out,4}) \\
+ k_{12}k_{43}k_{out,4} - k_{21}k_{32}k_{43} - k_{21}k_{32}k_{out,4} + k_{21}k_{34}k_{1,in} - k_{21}k_{43}k_{1,in} - k_{21}k_{43}k_{out,4} \ \\
+ k_{23}k_{34}k_{1,in} + k_{23}k_{1,in}k_{out,4} - k_{32}k_{43}k_{1,in} - k_{32}k_{1,in}k_{out,4} \ \\
+ J (k_{21}k_{32}k_{43}k_{1,in} + k_{21}k_{32}k_{43}k_{out,4} + k_{21}k_{32}k_{1,in}k_{out,4} + k_{21}k_{43}k_{1,in}k_{out,4} + k_{32}k_{43}k_{1,in}k_{out,4} - k_{21}k_{32}k_{43}k_{1,in}k_{out,4}.
$$

(B.5)

A python script has been written that generates and solves the polynomial equation in $J$ for an arbitrary long chain of heme sites. Any solution for $J$ is discarded if $J < 0$ or if it yields any of the $P_i$ outside the physical range (0 ≤ $P_i$ ≤ 1). There is no guarantee on a unique physical solution, but in my calculations I always obtain exactly one physical solution. Here, the polynomial equation B.5 is solved for the case where $k_{1,in}$ and $k_{out,4}$ are much faster than any ET rate between heme pairs, $k_{1,in},k_{out,4} >> k_{ji} \quad \forall i, j$. This gives the maximum, protein limited electron flux through STC, $J = J_{max}$, in the forward direction, from heme 1 to heme 4. A similar polynomial is obtained and solved for the flux in the reverse direction, from heme 4 to heme 1, assuming $k_{1,in} = k_{out,4} = 0$ and $k_{4,out},k_{in,1} >> k_{ji} \quad \forall i, j$. 

140 Appendix B. Derivation of analytic solution to chemical master equation: example of STC
Appendix C

ET parameters of STC, MtrC, MtrF

Table C.1: Summary of ET parameters and rate constants for heme-to-heme ET in STC. Electronic coupling matrix elements, $\langle |H_{ab}|^2 \rangle^{1/2}$, computed from POD calculations and averaged over MD trajectories are given for the large QM model. Reorganisation free energies, $\lambda$, in the all-oxidized (all-ox) redox state of the proteins are obtained from MD simulations. The driving forces are obtained by heme reduction potential difference, $\Delta A_{ji} = A_i - A_j$, with the reduction potentials taken from NMR measurements of all-oxidized, deprotonated state from Ref. [141]. The non-adiabatic (Marcus) rate constants for ET from heme $i$ to heme $j$, $k_{ji}$, and for the reverse direction, $k_{ij}$, are given for the all-ox state of the protein.

| Heme pair $i$-$j$ | $\langle |H_{ab}|^2 \rangle^{1/2}$ (meV) | $\lambda$ (eV) | $\Delta A_{ji}$ (meV) | $k_{ji}$ (s$^{-1}$) | $k_{ij}$ (s$^{-1}$) |
|-------------------|-------------------------------------|---------------|----------------------|-------------------|-------------------|
| 1-2               | 2.17                                | 1.08          | -70                  | $8.7 \times 10^6$ | $5.8 \times 10^5$ |
| 2-3               | 3.08                                | 0.76          | -21                  | $1.7 \times 10^8$ | $7.4 \times 10^7$ |
| 3-4               | 2.08                                | 0.88          | 6                    | $1.3 \times 10^7$ | $1.7 \times 10^7$ |
### Table C.2: Summary of computed ET parameters and rate constants for heme-to-heme ET in MtrC and MtrF.

| Heme pair | Protein | \(|\langle|H_{ab}|^2\rangle^{1/2}|\) (meV) | \(\lambda\) (eV) | \(\Delta A_{ji}\) (eV) | \(k_{ji}\) (s\(^{-1}\)) | \(k_{ij}\) (s\(^{-1}\)) |
|-----------|---------|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| 10-9      | MtrC    | 2.82                            | 0.70            | -0.02           | \(2.8 \times 10^8\) | \(1.2 \times 10^8\) |
| 9-8       | MtrC    | 3.46                            | 0.73            | 0.12            | \(1.6 \times 10^7\) | \(1.8 \times 10^9\) |
| 9-8       | MtrF    | 3.03                            | 0.93            | -0.13           | \(1.9 \times 10^8\) | \(1.5 \times 10^6\) |
| 8-6       | MtrF    | 2.22                            | 0.68            | -0.21           | \(4.1 \times 10^9\) | \(1.5 \times 10^8\) |
| 8-6       |         | 2.14                            | 0.87            | -0.10           | \(1.2 \times 10^8\) | \(2.2 \times 10^6\) |
| 6-1       |         | 2.01                            | 0.93            | 0.18            | \(2.1 \times 10^7\) | \(1.9 \times 10^8\) |
| 6-1       |         | 2.95                            | 0.94            | -0.01           | \(2.1 \times 10^7\) | \(1.4 \times 10^7\) |
| 1-3       |         | 0.91                            | 0.77            | 0.11            | \(9.6 \times 10^5\) | \(6.5 \times 10^7\) |
| 1-3       |         | 1.25                            | 0.96            | 0.12            | \(2.0 \times 10^5\) | \(2.3 \times 10^7\) |
| 3-4       |         | 1.50                            | 0.71            | -0.25           | \(2.4 \times 10^9\) | \(1.8 \times 10^5\) |
| 3-4       |         | 4.51                            | 0.75            | 0.10            | \(3.3 \times 10^7\) | \(1.8 \times 10^9\) |
| 4-5       |         | 3.21                            | 0.85            | -0.11           | \(3.9 \times 10^8\) | \(4.8 \times 10^6\) |
| 4-5       |         | 3.37                            | 0.84            | -0.22           | \(2.6 \times 10^9\) | \(4.6 \times 10^5\) |
| 6-7       |         | 1.49                            | 0.81            | -0.17           | \(2.9 \times 10^8\) | \(4.7 \times 10^5\) |
| 6-7       |         | 0.81                            | 1.06            | -0.13           | \(3.7 \times 10^6\) | \(2.9 \times 10^4\) |
| 1-2       |         | 1.01                            | 0.90            | -0.22           | \(1.2 \times 10^8\) | \(2.7 \times 10^4\) |
| 1-2       |         | 0.70                            | 1.13            | 0.02            | \(1.0 \times 10^5\) | \(1.9 \times 10^5\) |
Bibliography


