TECHNISCHE UNIVERSITÄT MÜNCHEN
Fakultät für Chemie
Lehrstuhl für Computergestützte Biokatalyse

DISSERTATION

Redox-coupled Proton Transfer Dynamics in Cytochrome *c* Oxidase

Shreyas Supekar

Garching b. München, 2017

TECHNISCHE UNIVERSITÄT MÜNCHEN
Fakultät für Chemie
Lehrstuhl für Computergestützte Biokatalyse

Redox-coupled Proton Transfer Dynamics in Cytochrome *c* Oxidase

Shreyas Supekar

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende(r) : Prof. Dr. Michael Sattler

Prüfer der Dissertation:

1. Prof. Dr. Ville R. I. Kaila
2. Prof. Dr. Karsten Reuter

Die Dissertation wurde am 26.09.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 06.11.2017 angenommen.

Abstract

Cytochrome *c* oxidase (C*c*O) is a trans-membrane protein in the respiratory chains of all aerobic organisms, acting as the terminal electron acceptor enzyme. C*c*O drives biological energy transduction by functioning as a redox-driven proton pump harnessing the energy released from oxygen reduction at its active site. C*c*O utilizes this energy to pump protons across a biological membrane, against an electrochemical gradient, which is used for ATP synthesis and active transport.

In the catalytic cycle of C*c*O, the enzyme sequentially receives electrons from the electron carrier protein, cytochrome *c*, and couples each reduction event to translocation of a proton across the membrane. Additionally, C*c*O employs conserved hydrophilic channels for uptake of protons from the aqueous phase to its active site. The complex coupling of electron and proton transfer reactions, and sequence of events in the catalytic cycle remain unclear, despite decades of research. Importantly, the pumped proton is transiently stored at an unknown loading site before it is ejected across the membrane, in response to redox changes in the enzyme. The identity and location of the loading site has, however, remained elusive for decades.

This dissertation aims at understanding the mechanistic coupling between the redox states of the enzyme and the proton transfer events, and to identify key functional elements in the proton pumping machinery of C*c*O by employing multi-scale quantum and classical molecular simulations. This work has provided molecular insight into the identity of the elusive proton–loading site as a structurally conserved water cluster near the active site, which is highly sensitive to the redox states of the enzyme. Furthermore, this work has provided insight into how water dynamics in the active site is sensitive to the redox states of the surrounding cofactors. A mechanism for the activation of a conserved proton conducting K-channel has also been proposed.

Zusammenfassung

Cytochrom-*c*-Oxidase (C*c*O) ist ein Transmembranprotein in der Atmungskette aller aerober Organismen, das als das terminale Elektronenakzeptor-Enzym fungiert. C*c*O steuert die biologische Energieübertragung als redoxgesteuerte Protonenpumpe, wobei es die durch Sauerstoffreduktion im aktiven Zentrum freigewordene Energie nutzt. C*c*O verwendet diese Energie, um Protonen gegen einen elektrochemischen Gradienten über eine biologische Membran zu transportieren, wobei der Gradient für die Synthese von ATP und aktiven Transport genutzt wird.

Im katalytischen Zyklus von C*c*O nimmt das Enzym nacheinander vom Elektronentransportprotein, cytochrom *c*, Elektronen auf und koppelt jeden Reduktionsvorgang an die Translokation eines Protons durch die Membran. Zusätzlich verwendet C*c*O konservierte Kanäle für die Aufnahme von Protonen aus der wässrigen Phase in das aktive Zentrum. Die komplexe Kopplung von Elektronen- und Protonentransferreaktionen sowie die Reihenfolge der einzelnen Prozesse im katalytischen Zyklus sind trotz jahrelanger Forschung immer noch unklar. Wichtig ist, dass das transferierte Proton vorübergehend an einer unbekannten Ladestelle aufbewahrt wird, bevor es in Reaktion auf Redoxveränderungen im Enzym durch die Membran transportiert wird. Die Identität und der Ort dieser Ladestelle konnten jedoch seit Jahrzehnten nicht ermittelt werden.

Das Ziel dieser Dissertation ist es, mithilfe von maßstabsübergreifenden quantenmechanischen und klassischen molekularen Simulationen die mechanistische Kopplung der Redoxzustände des Enzyms und der Protonentransferprozesse zu verstehen, sowie die entscheidenden funktionalen Elemente im Protonentransfermechanismus von C*c*O zu identifizieren. Diese Arbeit gewährt einen molekularen Einblick in die Identität der schwer identifizierbaren Protonenladestelle als einen strukturell konservierten Wassercluster in der Nähe des aktiven Zentrums, der sehr empfindlich auf die Redoxzustände des Enzyms reagiert. Des Weiteren gibt diese Arbeit einen Einblick, wie die Wasserdynamik im aktiven Zentrum auf die Redoxzustände der umgebenden Kofaktoren reagiert. Ferner wird ein Mechanismus für die Aktivierung des konservierten, protonenleitenden K-Kanals vorgeschlagen.

Contents

List of abbreviations

μVT	grand canonical ensemble
NPT	isothermal–isobaric ensemble
NVE	microcanonical ensemble
NVT	canonical ensemble
A	oxygen bound ferrous/cuprous state
ATP	adenosine triphosphate
B3LYP	Becke, three-parameter, Lee-Yang-Parr, an exchange-correlation functional
B88	Becke's functional from 1988
BNC	binuclear center
BOA	Born-Oppenheimer approximation
BP86	B88 exchange functional and PW86 correlation functional
C*c*O	cytochrome *c* oxidase
CCSD	coupled cluster with singles and doubles
CCSD(T)	coupled cluster with singles, doubles, and perturbative triples
CI	configuration interaction
cyt *c*	cytochrome *c*
DFT	density functional theory
$\mathrm{E}/\mathrm{E}_\mathrm{H}$	one-electron reduced state, non-pumping/pumping
$\mathrm{E}_\mathrm{H,R}$	one-electron reduced form of E_H state
eT	electron transfer
F	ferryl state
FAD	flavine adenine dinucleotide
FCI	full configuration interaction
FFT	fast Fourier transform
Fig.	figure
FMN	flavine mononucleotide

FTIR	fourier transformation infrared (spectroscopy)
GGA	generalised gradient approximation
GTO	gaussian-type orbitals
HCO	heme–copper oxidase
HF	Hartree-Fock
KS	Kohn-Sham
LCAO-MO	linear combination of atomic orbitals to make molecular orbitals
LDA	local-density approximation
LJ	Lennard-Jones potential
LPBE	linearized Poisson-Boltzmann equation
LSDA	local spin density approximation
LYP	Lee, Yang and Parr functional
MD	molecular dynamics
MEP	minimum energy path
MM	molecular mechanics
MP2	Møller-Plesset second order energy correction
N-side	negatively charged side of the membrane
NAD	nicotine adenine dinucleotide
NOR	nitric oxide reductase
O/O_H	fully oxidized state, non-pumping/pumping
$O_{H,R}$	one-electron reduced form of O_H state
P-side	positively charged side of the membrane
P_M	peroxy intermediate prepared from mixed valence state
P_R	peroxy intermediate prepared from fully reduced state
PB	Poisson-Boltzmann
PBE	Perdew-Burke-Ernzerhof functional
PES	potential energy surface
PLS	proton-loading site
PME	particle mesh Ewald (summation)
PMF	potential of mean force
pmf	proton motive force
pT	proton transfer

PW86	Perdew-Wang functional from 1986
PW91	Perdew-Wang functional from 1991
Q/QH2	quinone/quinol
QM	quantum mechanics
QM/MM	quantum mechanics/molecular mechanics
R	fully reduced state
RMSD	root mean square deviation
ROS	reactive oxygen species
SE	Schrödinger equation
STO	Slater-type orbitals
TDSE	time-dependent Schrödinger equation
TISE	time-independent Schrödinger equation
vdW	van der Waals forces/interactions
WHAM	weighted histogram analysis method
xc	exchange-correlation

1 Introduction

Earth came into existence about 4.5 billion years ago [1]. Fossil evidence indicates that life on earth started emerging as early as about 4.1-3.8 billion years ago [2, 3]. The extreme conditions and abundance of energy in the environment enabled synthesis of simple organic compounds like formaldehyde, acetates and phosphate anhydrides. It was originally proposed that these simple organic molecules give rise to a self-replicating RNA-like polymer [4, 5]. The lack of chemical pathways to harvest energy from the environment for sustenance would yet hinder the proliferation of life. Geological evidence suggests that early-earth had an oxygen-poor environment [6]. Recent biochemical studies postulate a prototype of metabolism may have emerged either on the earth's surface, with solar energy as the main driving force, or in ocean depths near hydrothermal vents, with alkaline geochemical gradients as an energy source [7]. These proto-metabolism pathways led to the emergence of the predecessors of first photosynthetic organisms, light harvesting green sulfur bacteria or purple bacteria like entities, possibly near hot volcanic springs or hydrothermal vents containing ample amounts of sulfur and iron. They might have used light energy and electrons from sulphides allowing reductive transformation of carbon dioxide (CO_2) into more complex carbon compounds. These were followed by cyanobacteria about 2.7 billion years ago, which first used water as an electron donor and sunlight as energy source and pumped oxygen in the atmosphere as a by-product, changing the environment from oxygen-poor to oxygen-rich [8]. This set up the stage for oxidative metabolism and further complexity in life forms explaining the massive explosion of diversity in the Cambrian era, which consequently fixed photosynthesis and respiration as the quintessentially dominant energy transduction mechanisms for emergent life on earth.

1.1 Energy metabolism

Bioenergetics is the study of biological energy conversion machineries, which sustain all carbon based life forms [9]. Autotrophs (plants, photosynthetic bacteria) are organisms that perform photosynthesis, in which they take in CO_2 from the atmosphere and use the energy from sunlight to power the unfavorable reduction of CO_2 for the synthesis of complex organic molecules like carbohydrates, while releasing molecular oxygen (O_2) as a by-product of the photosynthesis reaction. These complex organic molecules can then be used by heterotrophs as energy sources to transduce the stored energy into ATP for endergonic cellular functions. This process is called respiration, and it involves the use of O_2 for favorable re-oxidation of carbohydrates and other complex carbon molecules back into CO_2, hence releasing the stored energy, which is subsequently transformed into ATP for performing cellular functions. The overall stoichiometry of the respiration reaction can be given as

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6H_2O + 6CO_2 + \text{energy}$$

Chemically, respiration is thus the opposite reaction to photosynthesis. This complementarity forms the basis of energy flow, sustaining the vast majority of life forms on our planet.

Cellular respiration is how life forms derive energy from the foodstuff they consume. In the scope of bioenergetics, respiration is a sequence of energy transduction reactions, which comprises the exergonic oxidation of complex organic molecules (carbohydrates, fatty acids and amino acids) and the transformation of the energy stored in these molecules to ATP, several electron carriers (NADH, NADPH, $FADH_2$), and heat. Glycolysis in eukaryotes takes place in the cytoplasm of the cell and involves breakdown of one equivalent of a 6-carbon carbohydrate into two 3-carbon pyruvate equivalents. The released free energy is employed to make two equivalents of ATP and NADH. In presence of oxygen, the pyruvate undergoes oxidative decarboxylation, wherein it is further broken down into 2-carbon acetyl-CoA, producing one equivalent of NADH. This acetyl-CoA is then fed into the citric acid cycle, which in one turnover yields three equivalents of NADH, one equivalent of ATP, one equivalent of $FADH_2$ and two equivalents of CO_2. The citric acid cycle also catabolizes proteins and fatty acids, and it takes place in the mitochondrial matrix of eukaryotes. The mitochondrion is an organelle in all eukaryotic cells, often referred to as "the Powerhouse of the Cell". Genetic evidence suggests that mitochondria were once primitive bacteria which were engulfed by primitive cells (hosts) and their evolution into an organelle is explained by endosymbiotic theory in which most of their genes of the endosymbiont are transferred to the host [10]. A mitochondrion is encapsulated by two membranes, an inner and an outer phospholipid membranes. The space inside the inner membrane is called mitochondrial matrix.

The electron carriers (NADH, $FADH_2$) generated by catabolism of complex organic molecules in the mitochondrial matrix are fed to the electron transport chain (ETC). The electrons reduce oxygen (O_2) into water, and use the free energy derived in the process to pump protons from the mitochondrial matrix to the inter-membrane space between inner and outer mitochondrial membranes [9, 11]. This accumulates a proton motive force (pmf) due to a difference in proton concentrations on the two sides of the inner membrane of mitochondria akin to a charging of a capacitor [12]. The generated pmf is used for synthesis of ATP, and for active transport.

The transmembrane protein machineries responsible for the ETC and ATP synthesis are known as respiratory complexes. The respiratory complexes (Fig 1.1), Complex I (NADH-ubiquinone-oxidoreductase), complex III (cytochrome bc_1 or quinone-cytochrome c oxidoreductase) and complex IV (cytochrome c oxidase) span the inner membrane of mitochondria and are responsible for transferring electrons along the ETC. In addition, the respiratory complexes pump protons from the mitochondrial matrix (N-side) to the inter-membrane space (P-side), except Complex II (succinate dehydrogenase), which only transfers electrons without pumping protons. The generated pmf, as a result of proton pumping, is used by ATP synthase, which drives the synthesis of ATP. ATP synthase is often referred to as (respiratory) complex V. The proton pumping is driven by the transfer of electrons through redox metal centers in the respiratory complexes in a highly controlled, stepwise fashion, gradually releasing the energy in a quasi-reversible manner. This maximizes the electrochemical work while minimizing the loss of free energy in form of heat [9].

Complex I (NADH-ubiquinone-oxidoreductase) serves as the entry-point for the electron carrier NADH, produced in the citric acid cycle. NADH supplies two electrons to the flavin mononucleotide (FMN) prosthetic group bound to Complex I and reduces it to $FMNH_2$. These electrons are then shuttled from the reduced FMN via eight iron-sulfur (Fe-S) clusters to ubiquinone (Q) reducing it to ubiquinol (QH_2) [13]. This reduction is coupled to translocation of 3-4 protons across the membrane [14]. Complex I is a huge trans-membrane L-shaped protein with distinct hydrophobic and hydrophilic parts. It is the largest of the respiratory complexes with the bovine Complex I having a molecular mass of *ca.* 1 MDa and comprising 45 subunits [15].

Complex II or succinate dehydrogenase catalyzes the oxidation of succinate to fumerate and transfers the electrons to FAD, reducing it to $FADH_2$. The electrons in $FADH_2$ are then transferred via a series of three Fe-S clusters to Q, reducing it to QH_2, and hence adding further to the quinol pool accumulated by Complex I. Complex II, however, does not translocate protons across the membrane. Complex II is the smallest of the respiratory complexes (124 kDa; 4 subunits) and also the only respiratory complex taking part in the citric acid cycle [16].

Complex III or cytochrome bc_1 (Ubiquinone-cytochrome *c* oxidoreductase) catalyzes the transfer of electrons from ubiquinol to another electron carrier, cytochrome *c*. This set of reactions is known as the Q-cycle [17]. Complex III consists of three respiratory subunits, (1) cytochrome *b*, comprising two *b*-type hemes, (2) cytochrome c_1, comprising one *c*-type heme, and (3) Rieske protein or iron-sulfur protein, comprising a two iron two sulfur (2Fe-2S) cluster. The 2Fe-2S cluster is involved in transferring electrons from QH_2 to cytochrome *c*, which is coupled to translocation of one proton across the membrane for every electron transferred to cytochrome *c* [18, 19].

Complex IV or cytochrome *c* oxidase is the terminal electron acceptor enzyme in respiratory chains of mitochondria and bacteria. It takes electrons from the reduced cytochrome *c* and shuttles them via a bimetallic copper center, Cu_A, and an A-type heme center, heme *a*, to O_2 reduction site situated at the binuclear center (BNC). The BNC constitutes of an A-type heme group, heme a_3, and a copper center called Cu_B. The (free) energy released from the reduction of O_2 to H_2O (*ca.* 37.5 kcal mol^{-1} at standard conditions) drives the complete electron transport chain. In the catalytic cycle, cytochrome *c* oxidase takes up eight protons from the mitochondrial matrix. Four of the protons are transferred to the reduced O_2 species, whereas the other four protons are translocated across the membrane into the inter-membrane space, adding to the proton gradient. The pmf generated by respiratory complexes is subsequently utilized by ATP synthase [20–23].

Complex I transfers 3-4 protons across the membrane using two electrons from the oxidation of NADH. Complex III pumps two protons across the membrane oxidizing two ubiquinol molecules and cytochrome *c* oxidase translocates four additional protons across the membrane oxidizing four molecules of cytochrome *c*. The accumulation of the protons translocated by the respiratory complexes from the mitochondrial matrix to the inter-membrane space results in the generation of an electrochemical proton gradient, $\Delta\mu_{\mathrm{H}_+}$, a combination of the pH gradient (ΔpH) as a consequence of the proton concentration gradient and a voltage component (ΔΨ) due to charge difference on either sides of the membrane,

$$\Delta\mu_{H^+} = 2.303\frac{RT}{nF}\Delta p\mathrm{H} + \Delta\Psi. \tag{1.1}$$

The electrochemical proton gradient, $\Delta\mu_{H^+}$ (or the pmf), is then used by Complex V (ATP synthase) for making ATP and for active transport via ion channels. The whole ETC ending in synthesis of ATP is known as oxidative phosphorylation. This energy transduction mechanism was first suggested by Peter Mitchell in 1961 as chemiosmotic theory [24], for which he received the Nobel prize in 1978. ATP synthase employs a remarkably distinct mechanism for catalysis involving physical rotation of its subunits, acting as a

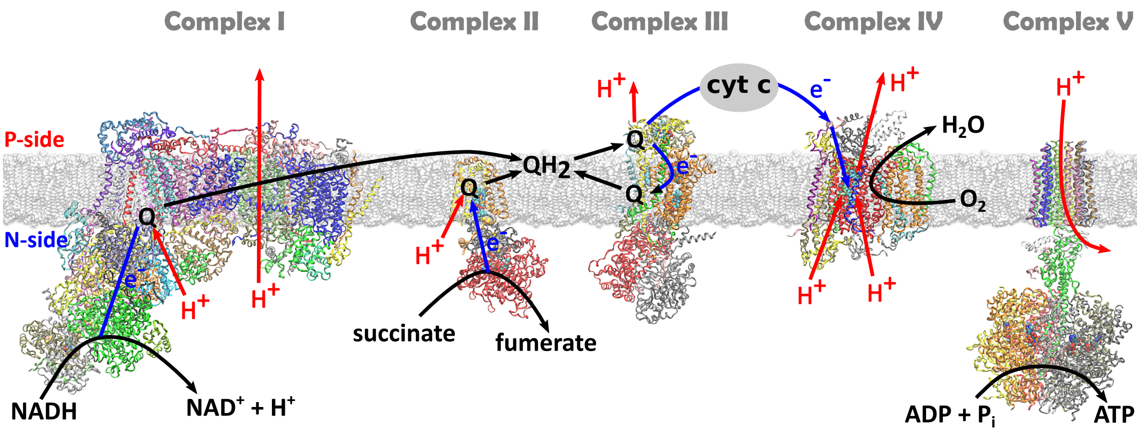

Figure 1.1: The mitochondrial respiratory chain. Respiratory complexes I-V embedded in the inner mitochondrial membrane. Blue arrows indicate electron transfer processes, black arrows indicate chemical reactions and transport, and red arrows show the pathways of proton transfer processes. Structures of the respiratory complexes were obtained from the protein data bank with PDB IDs: 5LDW, 1ZOY, 1BCC, 1V54 and 2XOK for complexes I, II, III, IV and V, respectively.

molecular dynamo transforming the electrochemical potential into ATP [25, 26].

Uncoupling proteins, like thermogenin, expressed in brown adipose tissue of newborns and hibernating animals acts as uncouplers, blocking oxidative phosphorylation. Thermogenin acts as a proton channel providing a route for the translocated protons back into the mitochondrial matrix, dissipating the electrochemical proton gradient and generating heat in the process, which is used to provide requisite warmth to the child or animal in adverse conditions [27, 28].

1.2 Proton-coupled electron transfer in bioenergetics

All the respiratory enzymes catalyze proton-coupled electron transfer (PCET) processes that involve coupled transfer of electrons and protons. During catalysis, electrons are shuttled to redox metal-centers (hemes, iron-sulfur clusters, etc.) of the respiratory enzymes, which is coupled to the transport of protons to protonable sites within the enzymes [29]. The proton transfer processes are accomplished by intra-protein water molecules and titrable sites in the protein that function as proton carriers.

Proton transfers coupled to redox events involving electron transfers were first observed in 1949 from kinetic isotope effects (KIEs) in experiments [30]. Mechanistically, PCET processes can occur in two ways, in a concerted or stepwise manner (Fig. 1.2). In the concerted (synchronous) PCET mechanism, both the electron and the proton are transferred simultaneously, for example, in hydrogen atom transfer (HAT) reactions. Although such concerted PCET reactions occur in many enzymes [31], in the scope of respiratory enzymes, stepwise (asynchronous) PCET is more relevant, wherein an electron or a proton is transferred first, followed by transfer of the second species [32].

In many respiratory enzymes, the electron and proton transport events can occur between different acceptors and donor sites, unlike HAT processes, where electrons and protons are transferred between the same species. For example, in complex I, electron transfer is

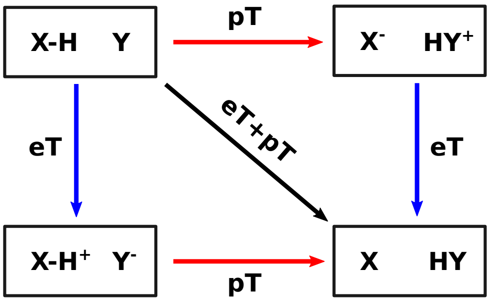

Figure 1.2: Concerted and stepwise PCET. X and Y are the species participating in the PCET reaction. Red and blue arrows indicate proton and electron transfer events, respectively. The path along the diagonal represents concerted (synchronous) mechanism, and the paths along the edges show stepwise (asynchronous) PCET mechanisms.

coupled to proton transfer taking place *ca.* 200 Å away [33]. In stepwise PCET, either the electron transfer step or the proton transfer step can be rate-limiting. In respiratory enzymes, the proton transfer step is usually rate-limiting, whereas the electron transfer is fast [20], as in cytochrome *c* oxidase (see chapter 2). The overall rate of the PCET event is thus governed by the slower proton transfer reaction.

The sequence of events and the nature of reaction intermediates in the stepwise PCET is governed by thermodynamic parameters like redox potentials of metal-centers, reorganization energies and pK_as of protonable groups involved. Typically, in respiratory enzymes the electrons and protons are transferred across large distances over tens of Ångströms. A proton is, however, *ca.* 2000 times heavier than an electron, and as a consequence, while the electrons can tunnel across distances of *ca.* 15 Å in enzymes [34, 35], protons can not. This necessitates a presence of hydrogen bonded proton carrier species, connecting the proton donor and the proton acceptor groups. The proton carriers can be mobile water molecules or protonable residues of the protein. The proton carriers are thus integral to the respiratory machineries, and understanding their function may provide valuable insight into the intricate workings of respiratory enzymes.

2 Cytochrome *c* oxidase: structure and function

Cytochrome *c* oxidase (C*c*O), is the terminal enzyme in the electron transport chains of bacteria and eukaryotes, catalyzing the four-electron reduction of molecular oxygen to water. C*c*O additionally transduces the free energy released from oxygen reduction to pump protons across the membrane against a membrane potential of up to 0.2 V [36]. The C*c*O reaction can be written as:

$$O_2 + 4e^-(\text{cyt}\,c; \text{P-side}) + 8H^+(\text{N-side}) \rightarrow 2H_2O + 4H^+(\text{P-side}).$$

C*c*O receives electrons from the electron-carrier cytochrome *c* (cyt*c*), and shuttles the electrons via transition metal-centers to the oxygen-reduction site in the interior of the enzyme. The electrons are transferred sequentially, one at a time, and each electron transfer event is coupled to the uptake of two protons from the negatively-charged side (N-side) of the mitochondrial matrix and translocation of a proton to the positively-charged side (P-side) in the inter-membrane space. The proton pumping function of C*c*O was discovered by Wikström in 1977 [37]. The initial four-electron reduction of oxygen takes place concurrently in the active site producing unprotonated water (O^{2-}; OH^-) by taking four electrons from the transition metal-centers of the enzyme and thus transferring the redox power of the O_2/H_2O redox couple to the enzyme. The subsequent pumping cycle proceeds in a stepwise fashion with each reduction event coupled to the uptake of two protons from the N-side of the membrane; one is added to reduced oxygen while the other proton is pumped to the P-side [20–23]. Understanding the chemistry of active site reduction and its coupling to proton pumping is thus essential for understanding the molecular mechanism of C*c*O catalysis.

2.1 Structure of cytochrome *c* oxidase

C*c*O belongs to the superfamily of heme–copper oxidases (HCO). All HCOs are membrane embedded enzymes that catalyze the reduction of molecular oxygen to water at a characteristic bimetallic active site comprising a copper atom, a high-spin heme group, and a tyrosine residue covalently linked to one of the histidine residues bound to the copper atom. All HCOs share subunit I, which contains the essential oxygen reduction machinery. Subunit I also has a low-spin heme which acts as a electron-queuing center before transferring electrons to the active site. C*c*O also contains a second catalytic subunit termed subunit II which houses a bimetallic copper site which receives electrons from cytochrome *c* and relays them to the low-spin heme. Based on structure and sequence data, HCOs can be classified into three types, namely, A-, B- and C-type oxidases that reduce oxygen to water and pump protons [38]. Additionally, the nitric oxide reductases (NORs) which reduce nitric oxide (NO) to nitrous oxide (N_2O), are members of the HCO family too, but lack any proton pumping function [39, 40].

A-type oxidases have both catalytic subunits (I and II) and employ two proton channels, the D- and K-channels, named after conserved aspertate and lysine residues, for conducting protons from the N-side to the interior of the enzyme. In the B- and C-type HCOs, the D-pathway is absent, and they employ a single proton channel analogous to the K-channel for uptake of all the protons [41–43]. X-ray structures of several bacterial and eukaryotic HCOs are now available from the extensive work in the past two decades [41, 44–48]. Most X-ray structures of C*c*O from organisms including *Bos taurus, Paracoccus denitrificans* and *Rhodobacter sphaeroides* are members of the A-type HCO subfamily [47, 49–54]. The B-type enzymes, found in bacteria, have been solved for *Thermus thermophilus* [55, 56]. The C-type oxidases, also found in bacteria, are only expressed at low oxygen levels, and structures are available from *Paracoccus denitrificans* and *Rhodobacter sphaeroides* [42, 57]. A-type C*c*Os have been extensively studied over the past few years shedding light on various mechanistic and structural features of the enzyme. The X-ray structure of C*c*O from Bovine heart (*Bos taurus*) was used to model the enzyme in this study, and hence the structure and residue numbering used in the subsequent text will pertain to the stated bovine enzyme. Bovine C*c*O consists of 13 subunits. Subunits I, II and III are encoded by mitochondrial DNA, whereas the remaining ten subunits are encoded by nuclear DNA [58]. Subunits I, II and III are critical for catalytic activity and are also conserved in bacterial C*c*Os. The remaining ten subunits are likely to be involved in assembly and stabilization of the catalytic subunits [23, 59–62].

2.1.1 Subunit I

Subunit I of bovine C*c*O is composed of twelve membrane-spanning α-helices and contains a six-coordinated low-spin heme *a* and the oxygen reducing BNC comprising a five-coordinated high-spin heme a_3 and $\mathrm{Cu_B}$ site. The metal centers are located in the interior of the protein, approximately a third of the membrane thickness from the P-side of the membrane, and are anchored in their positions through coordinating bonds with conserved residues (Fig. 2.5).

Heme *a*

Hemes are prosthetic groups containing an iron atom coordinated by a tetradentate porphyrin molecule as a chelating ligand. The low-spin heme *a* site consists of a heme A group, which has a formyl and a hydroxyethylfarnesyl group attached to the porphyrin ring. The heme A group is also ligated to the conserved His-61 and His-378 residues completing the octahedral coordination sphere. The vinyl group of heme *a* forms a hydrogen bond with Arg-38. The A-propionate group of heme *a* makes a hydrogen bond with Tyr-54, while the D-propionate group makes ion pairs with Arg-438 and Arg-439 residues. The hydroxyl group of the hydroxyethylfarnesyl chain is

bonded to Ser-382 (Fig. 2.1). In the oxidized form, heme *a* exists in a doublet state ($\mathrm{Fe^{III}}$; S=1/2), whereas in the reduced form, it exists as a singlet state ($\mathrm{Fe^{II}}$; S=0) [63, 64].

Binuclear center

The BNC is the active site of the enzyme catalyzing the four-electron oxygen reduction chemistry. The BNC is made of two redox-active metal centers, heme a_3 and $\mathrm{Cu_B}$, which are separated by *ca.* 5 Å and comprise the oxygen-binding site.

Heme a_3 is a penta-coordinated high-spin heme A group with His-376 acting as coordination ligand on the distal side of the BNC. The sixth coordination on the proximal side is provided by the oxygen molecule to be reduced upon binding at the BNC. The A-propionate of heme a_3 is hydrogen bonded to the protonated Asp-364 and His-368

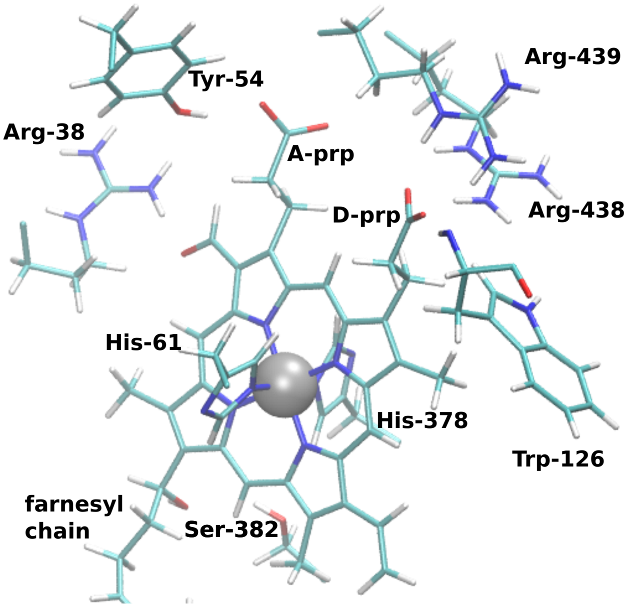

Figure 2.1: The structure of heme *a* site. The iron atom (in silver) coordinated by a porphyrin ring and conserved histidine residues His-61 and His-378. The vinyl group of heme A group is hydrogen bonded to Arg-38. The A-propionate is hydrogen bonded to Tyr-54. The D-propionate group forms ion pairs with Arg-438 and Arg-439, and a hydrogen bond with the backbone of Trp-126. The hydroxyethylfarnesyl chain of heme A group forms a hydrogen bond with Ser-382.

residues. The D-propionate group is hydrogen bonded to Trp-126 and forms an ion pair with Arg-438. The hydroxyl group of the hydroxyethylfarnesyl chain is hydrogen bonded to Tyr-244 (Fig. 2.2).

The $\mathrm{Cu_B}$ site consists of a copper atom ligated by three conserved histidine residues, His-240, His-290 and His-291. Additionally, a conserved tyrosine residue, Tyr-244, forms a covalent bond with His-240 (Fig. 2.2). The histidine-tyrosine cross-link is a post-translational modification and has been shown conserved across all HCOs based on structural and mass-spectrometric studies [51, 65]. Mutation of the tyrosine to phenylalanine renders C*c*O inactive, emphasizing its functional importance [66]. The cross-linkage has been suggested to modulate the redox potential of the $\mathrm{Cu_B}$ site and the proton affinity of protonable groups in the $\mathrm{Cu_B}$ site for proton and electron transfer processes in the catalytic cycle [67, 68].

In its oxidized form, heme a_3 exists in a sextet state ($\mathrm{Fe^{III}}$; S=5/2), whereas in the reduced form it exists in a quintet state ($\mathrm{Fe^{II}}$; S=2). Additionally, heme a_3 also exhibits a high-valent triplet state ($\mathrm{Fe^{IV}}$; S=1) coordinated by an oxo ligand ($\mathrm{O^{2-}}$) [64, 69–71]. Heme a_3 can also bind several other ligands like CO, $\mathrm{CN^-}$, hydroxide, etc. High field ligands like $\mathrm{CN^-}$ can turn heme a_3 into a low-spin heme [72, 73]. The $\mathrm{Cu_B}$ site exists in singlet (S=0), doublet (S=1/2), or triplet states (S=1), depending on the oxidation state of the copper atom ($\mathrm{Cu^I}$ or $\mathrm{Cu^{II}}$) and the character of the cross-linked Tyr-244 (protonated, deprotonated or radical). A hydroxyl ($\mathrm{OH^-}$) or a water group acts as the fourth ligand completing its tetragonal coordination sphere. The close spatial proximity of heme a_3 and $\mathrm{Cu_B}$ leads to antiferromagnetic spin coupling, rendering the BNC silent to EPR experiments [68, 70, 74].

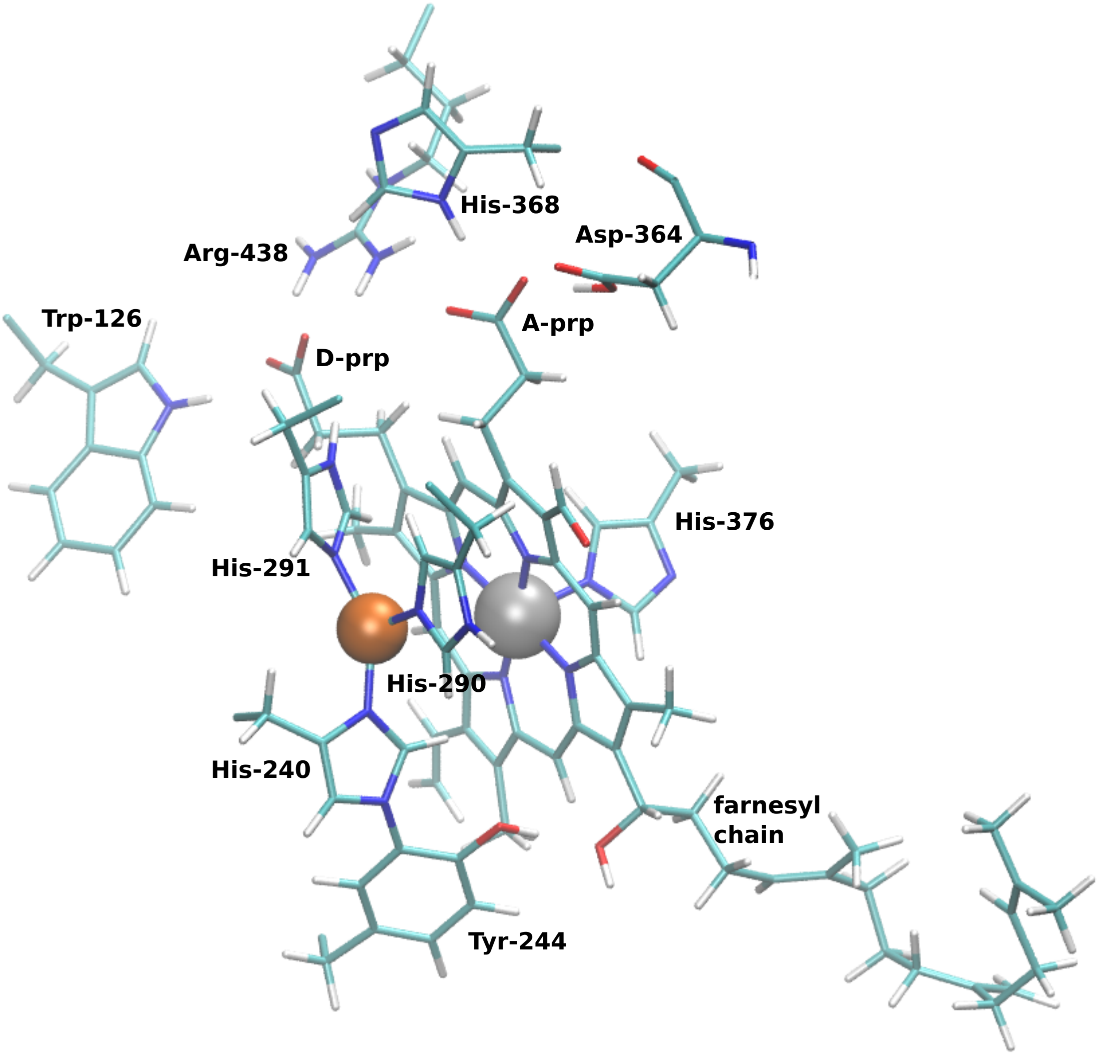

Figure 2.2: The binuclear center (BNC) composed of heme a_3 and $\mathrm{Cu_B}$ cofactors with surrounding protein residues. Heme a_3, with iron atom (in silver) is coordinated by a porphyrin ring and a histidine residue, His-376. The D-propionate group forms an ion pair with Arg-438 and is hydrogen bonded to Trp-126. The A-propionate group forms hydrogen bonds with Asp-364 and His-368 residues. The $\mathrm{Cu_B}$ site with the copper atom (in orange) is coordinated by histidine residues, His-290, His-291, and His-240 (cross-linked to Tyr-244). Tyr-244 also forms a hydrogen bond with the farnesyl chain of heme a_3.

2.1.2 Subunit II

Subunit II of bovine C*c*O is composed of two transmembrane α-helices and a globular β-sheet domain capping the α-helices of subunit I on the P-side of the membrane. Subunit II provides the docking site for the reduced cytochrome *c*. The β-sheet domain also houses the final redox-active center, the bimetallic $\mathrm{Cu_A}$ site, consisting of two copper atoms. The two copper atoms are bridged by two cysteine residues, Cys-196 and Cys-200. In addition, His-161, His-204, Met-207, and the backbone carbonyl of Glu-198 ligate the copper atoms, such that each copper atom has a tetragonal coordination sphere (Fig. 2.3). In the reduced form, $\mathrm{Cu_A}$ exists in a singlet state ($\mathrm{Cu}^{I}\mathrm{Cu}^{I}$; S=0), whereas in the oxidized form it exists in a doublet state ($\mathrm{Cu}^{I}\mathrm{Cu}^{II}$; S=1/2). However, the oxidized state is valence delocalized, and hence each copper is assigned an formal oxidation state of +1.5 [75, 76].

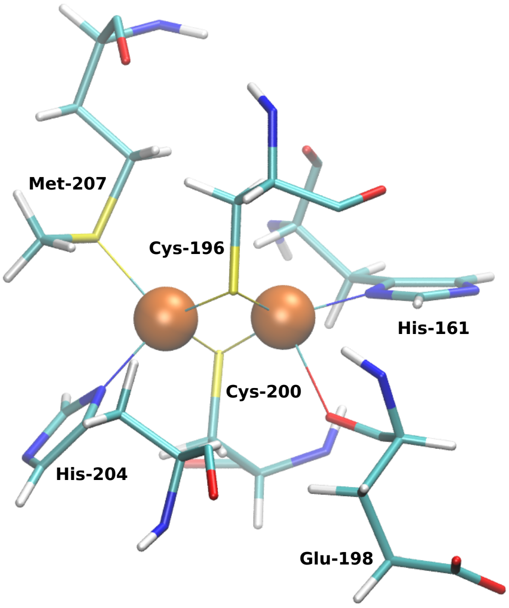

Figure 2.3: The structure of Cu_A center and its ligands. The bimetallic Cu_A center with two copper atoms (in orange) which are ligated by bridging Cys-196 and Cys-200 residues. The copper atoms are also ligated by His-161, His-204, Met-207, and the backbone of Glu-198 residues.

2.1.3 Subunit III

Subunit III consists of seven transmembrane α-helices. In addition, it contains three tightly bound phospholipids which have been suggested to be involved in delivery of molecular oxygen to the oxygen reduction site (BNC). Subunit III is also suggested to be involved in the folding process of C*c*O [77]. Knockout of subunit III leads to suicide inhibition linked to the Cu_B site, causing a decrease in the turnover rate of C*c*O and eventual inactivity, further emphasizing its physiological significance in efficient and safe delivery of molecular oxygen, while avoiding the formation of detrimental reactive oxygen species [78–80]. Mutations near the N-terminal domain of subunit III have been reported to slowdown the proton uptake through the D-channel, and the mutations are suggested to upshift the pK_a of the initial proton acceptor in the D-channel. [79–82]

2.1.4 Mg^{2+} center

The magnesium center is a redox-inactive metal center containing a Mg^{2+} cation situated at the interface of subunits I and II, and is located between the Cu_A center and heme a_3. The Mg^{2+} center is only present in A-type C*c*Os [23]. The octahedral coordination sphere of Mg^{2+} consists of Asp-369 and His-368 from subunit I, Glu-198 from subunit II, and three water molecules (Fig. 2.4). The Mg^{2+} cation can be replaced by an EPR-active (paramagnetic) Mn^{2+} cation, which can be used to probe the coordination geometry and the nature of coordinating ligands in EPR experiments. This led to indirect evidence that the Mg^{2+} center could be involved in expulsion of water molecules produced in C*c*O turnover. Also, computational studies suggest involvement of the Mg^{2+} center in the proton exit route to the P-side of the membrane [83–86]. B- and C-type C*c*Os, which lack the Mg^{2+} center, are known to pump protons with half the efficiency of A-type C*c*Os, suggesting a possibility that the Mg^{2+} center is actively involved in the proton pumping mechanism of C*c*O [87] (but cf. Ref. [88]).

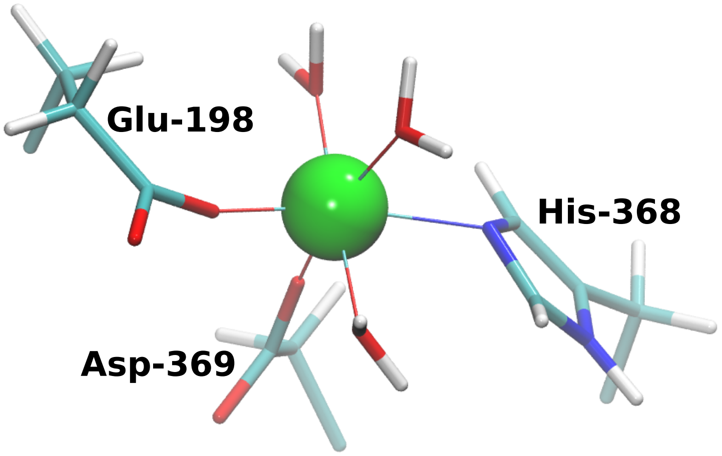

Figure 2.4: The structure of Mg^{2+} center and its ligands. The magnesium atom (in green) is coordinated by residues, His-368, Asp-369 and Glu-198 (subunit II). Three water molecules complete the octahedral coordination sphere of the Mg^{2+} cation.

2.1.5 Proton delivery channels in C*c*O

C*c*O adds protons and electrons from the opposite sides of the membrane to molecular oxygen at the BNC located in the interior of the enzyme, and couples this reduction to translocation protons across the membrane. To transport the protons to and from the N- and P-sides of the membrane to the interior of enzyme, C*c*O employs channels comprising a strategically arranged set of hydrophilic and protonable residues, along with intra-protein water molecules, which can act as proton conduits. Mechanistically, such a proton conduction generally occurs by Grotthuss mechanism, wherein an excess proton moves along a hydrogen bonded network through concomitant bond breakage and bond formation to reach the destination. In Grotthuss mechanism, the actual proton transfer occurs only among consecutive hydrogen bonded pairs progressively along the network, with a different proton transferred among consecutive pairs [89–91]. Computational studies of intra-protein proton transfers by Grotthuss mechanism suggest these proton transfer processes to be the rate-limiting steps in enzyme catalysis [92, 93]. C*c*O employs proton channels which (1) conduct protons from the negatively charged N-side of the membrane to the interior of the enzyme, and (2) conduct protons from the interior of the enzyme to the positively charged P-side of the membrane. In either cases, the transport is unfavorable due to the transmembrane electrochemical proton gradient (pmf). Such an active proton transport process needs extra thermodynamic work from the enzyme which is provided by coupling it to an exergonic reaction, namely, the oxygen reduction reaction in C*c*O [37]. Additionally, such vectorial and unfavorable proton transport must be mechanistically gated to prevent the thermodynamically favorable back-flow of protons to the proton source.

Structural studies of C*c*O suggest three proton delivery channels, all of which are primarily composed of subunit I residues: (1) the D-channel, named after the Asp-91 residue, located near the N-side, (2) the K-channel, named after the Lys-319 residue, situated about halfway between N-side and BNC, and (3) the H-channel, named after the His-413 residue, situated at a similar membrane depth as Lys-319 [21–23, 94].

D-channel

The D-channel connects the N-side via Asp-91 to the buried Glu-242, the terminus of D-channel, located at about half of the membrane width and about 10 Å from both heme

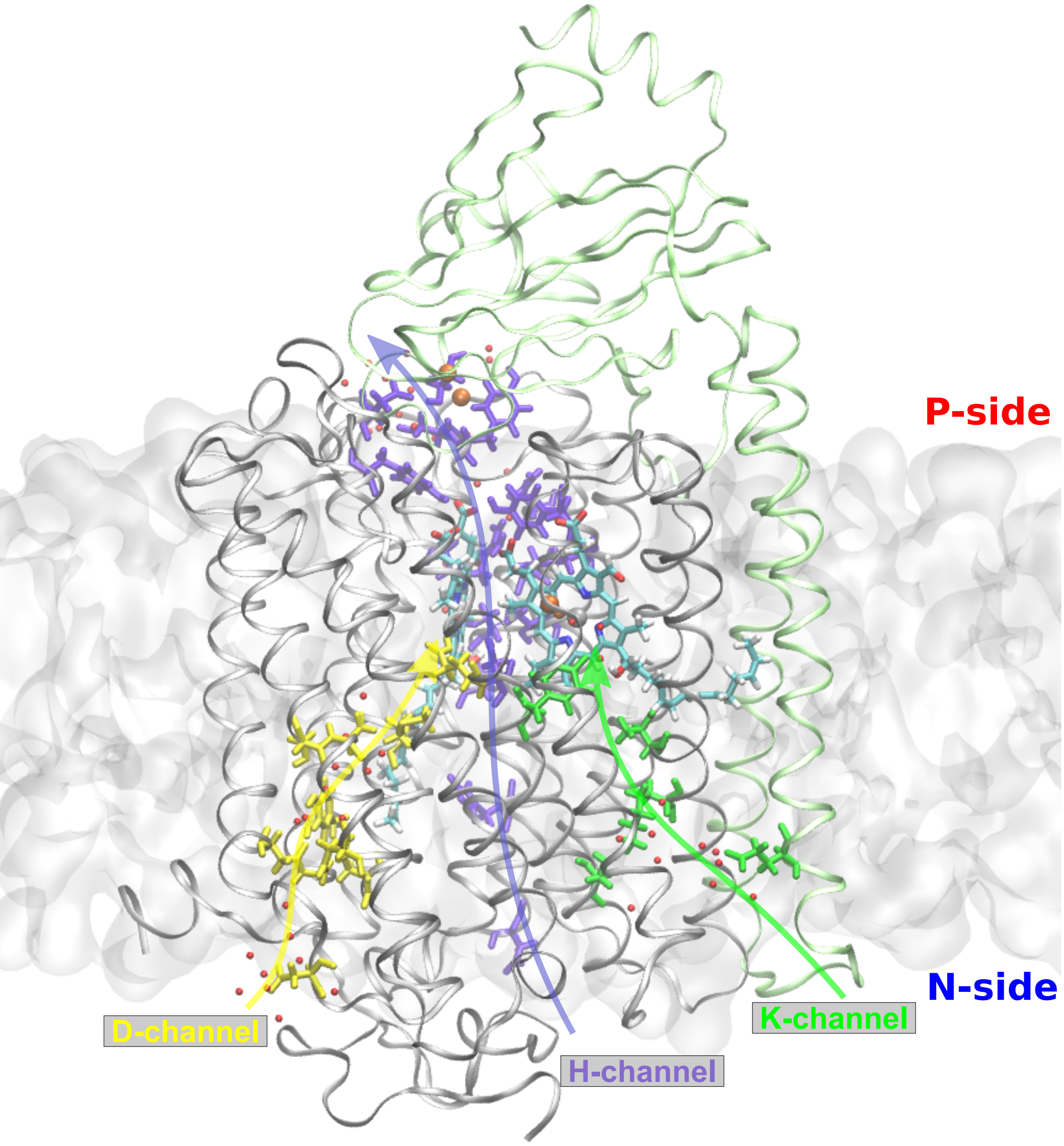

Figure 2.5: Proton conducting D-, K- and H-channels in C*c*O. The D-channel (shown in yellow) is an array of polar residues starting at Asp-91 (N-side) leading to Glu-242 (near the active site). The K-channel (shown in green) starts at Glu-62 on the N-side and proceeds via Lys-319 to Tyr-244 ($\mathrm{Cu_B}$ site). The H-channel (shown in purple) starts at His-413 on the N-side and proceeds via residues along heme *a* and $\mathrm{Cu_A}$ to the P-side. Subunit I is illustrated in gray ribbons and subunit II is depicted in green ribbon representation.

a and heme a_3 sites. Crystallographic studies have resolved at least ten water molecules, which are hydrogen bonded to the asparagine, serine, methionine, tyrosine and Glu-242 residue lining the D-channel [50, 95] (Fig. 2.5). The D-channel is conserved in all A-type C*c*Os. Experimental studies of several residues in the D-channel have helped show that in A-type C*c*Os, all four protons pumped in the reaction cycle are taken up through the D-channel, whereas only two out of four protons for oxygen reduction originate from the D-channel. The Glu-242 residue at the end of the D-channel thus shuttles protons to both, the BNC for oxygen chemistry, and the proton-loading site (PLS), located in the vicinity of the BNC, which transiently stores the pumped protons, before eventual ejection to the P-side [22, 96] (see 2.3.1). Combining experimental data and computational studies, it has been proposed that Glu-242 acts as a valve, with its side-chain conformation depending on its protonation state [96–105]. In the protonated form, Glu-242 breaks contact with the D-channel pointing away from the N-side and towards the BNC and the PLS, ready to supply a proton for oxygen reduction chemistry or pumping. Upon transferring its proton, the deprotonated glutamate readily flips "down" to pick up a proton from the D-channel followed by an "up" flip of the protonated form [99, 106]. Interestingly, several mutants of the residues lining the D-channel abolish pumping function without significantly affecting the reduction chemistry at the BNC [107–109]. Recent studies have found that the proton uptake from the D-channel occurs through a "proton-collecting antenna", comprising Asp-91 and several negatively charged residues of subunit III [78, 81, 110].

K-channel

The K-channel is conserved in all C*c*Os [50]. The proton uptake from the K-channel starts at Glu-62 (subunit II) located on the N-side and is fairly close in proximity to the lipid membrane enveloping C*c*O. The K-channel proceeds through Ser-255, Lys-319, and Thr-316 leading to the conserved Tyr-244 residue covalently linked to the His-240 ligand of $\mathrm{Cu_B}$ center [95, 111] (Fig. 2.5). The K-channel is much less structured than the D-channel and much less is known about its uptake mechanism. Crystallographic studies have identified only one water molecule in the K-channel, present between Ser-255 and Lys-319 [23]. From structural analysis, it was originally proposed that the K-channel starts at the Ser-255 residue [47]. However, computational studies predict Glu-62 (subunit II) to be the proton entry point, with Glu-62 strongly interacting with Lys-319 in order to facilitate proton uptake [112]. A recent study suggests a rather extended range of hydrophilic sites, including Glu-62 and Ser-255 residues, might be involved in the proton uptake via K-channel [113]. Mutagenesis studies of several residues lining the K-channel indicate massive inhibition of the reduction of BNC [114–116]. In A-type C*c*Os, the K-channel does not supply any pumped protons but supplies two protons to the BNC, and plays a key role in all instances of BNC reduction [94, 96, 101, 117, 118]. Interestingly, the B- and C-type oxidases have no D-channel, and the K-channel or its structural analogue supplies all the protons needed for pumping and BNC chemistry [41–43, 119].

H-channel

The H-channel starts near the N-side at His-413 residue and proceeds along a series of water molecules and hydrophilic residues like serines, and a threonine, leading to the hydroxyethylfarnesyl group of the low-spin heme *a*. Further on, it proceeds via Arg-38 (hydrogen bonded to the formyl group of heme *a*) to Tyr-371 via a water molecule. Several water molecules connect Tyr-371 to the backbone of Tyr-440 and Ser-441. This backbone was proposed to be the gating element for proton transfer leading to Ser-205 (subunit II) and eventually to Asp-51, near the $\mathrm{Cu_A}$ center (Fig. 2.5), and is exposed to the aqueous phase on the P-side [23, 50, 62, 120, 121]. Mutagenesis studies of several residues in the H-channel resulted in a loss of the pumping function while still maintaining normal turnover activity [50, 120]. This effect is similar to several D-channel mutations, but surprisingly the same D-channel mutations in bacterial A-type C*c*O systems under study resulted in no effect on the pumping function or BNC reduction [122]. Though most of the H-pathway is conserved in A-type C*c*Os, key elements like the peptide gate and Asp-51 are not present in bacterial A-type C*c*Os. Also, unlike mammalian C*c*Os, mutation of H-channel residues in bacterial C*c*Os have no effect on pumping and BNC chemistry. In light of the structural and mutation studies, it has been proposed that the H-channel rather acts like a dielectric well, facilitating electron transfer from cytochrome *c* via Cu_A to heme *a* [94].

2.1.6 Oxygen delivery and water exit pathways

The non-polar oxygen molecule can easily permeate membranes. However, to direct the molecule to the catalytic site of C*c*O, the enzyme has to employ specific channels to direct the oxygen transport. Structural and computational studies predict three highly hydrophobic pathways directing O_2 transport from the membrane, where oxygen is in abundance, towards the active site [50, 123]. Mutation of residues forming the pathways have shown to lead to severe slowdown or complete inhibition of catalytic activity of C*c*O [124, 125].

During catalysis, C*c*O produces large amounts of water at the BNC (ca. 1000 s^{-1}) [22]. The enzyme needs to constantly remove the water produced as a by-product of the reaction at the active site to ensure optimum turnover rates. EPR studies suggest that water molecules generated at the BNC exit via the water filled hydrophilic cavity above heme *a*, through the Mg^{2+}/Mn^{2+} site [85]. A recent molecular dynamics study by Sugitani *et al.* [86] also suggests several possible water exit pathways emanating in the vicinity of the Mg^{2+} site, leading to the P-side of the membrane.

2.2 Catalytic cycle

In living organisms, atmospheric molecular oxygen is a favorable substrate for the pool of electrons generated in the process of respiration for two key reasons: (1) high redox potential of molecular oxygen, which enables pmf generation, and (2) high barrier for reduction, which makes it almost inert in most environments to formation of often toxic, reactive oxygen species (ROS) formed as a consequence of uncontrolled reduction. Nevertheless, molecular oxygen can be subjected to controlled reduction by an enzyme like C*c*O [22, 126].

Molecular oxygen binds to the reduced form of the enzyme, known as the R state, to form an oxygen adduct intermediate, called the A state (Fig. 2.6), which was spectroscopically observed by Chance *et al.* in 1975 [127, 128]. The oxygen binding occurs at a timescale of *ca.* 10 μs [129]. The A state is very unstable, and decays into the so-called mixed-valence peroxy state, P_M, in 300 μs, and involves splitting of the O=O bond and its rapid four-electron reduction with the electrons derived from the reduced BNC(Fe[II]; Cu[I]; Tyr-OH). Two electrons are donated by heme a_3 forming a ferryl intermediate, Fe[IV]=O, one electron from Cu_B forming a Cu[II]-OH intermediate, and one electron from Tyr-244. Tyr-244 also internally transfers a proton to the oxygen ligand of copper forming a hydroxy ligand and leading to the formation of a neutral tyrosyl radical ($TyrO^{\bullet}$) [22]. This rapid reduction of molecular oxygen transfers the energy stored in the molecular oxygen to the enzyme and prevents the formation of undesirable ROS [126]. One-electron reduction of the BNC by an electron transferred from heme *a* gives the P_R state, where the electron is transferred to Tyr-244 making a tyrosinate anion ($TyrO^-$). The P_R state is quite unique as it is the only experimentally observable catalytic state of C*c*O which includes an extra electron at the BNC uncompensated by a proton. The P_R state is obtained in experiments when heme *a* is in its reduced state, before binding of molecular oxygen to the R state occurs [22, 98]. A proton transfer from the D-channel to the Cu_B center leads to the oxo-ferryl species (F state) with a water molecule coordinated to the copper atom at the BNC. Formation of the F-state is coupled to a translocation of a proton across the membrane [22, 96]. The steps described above complete one of the

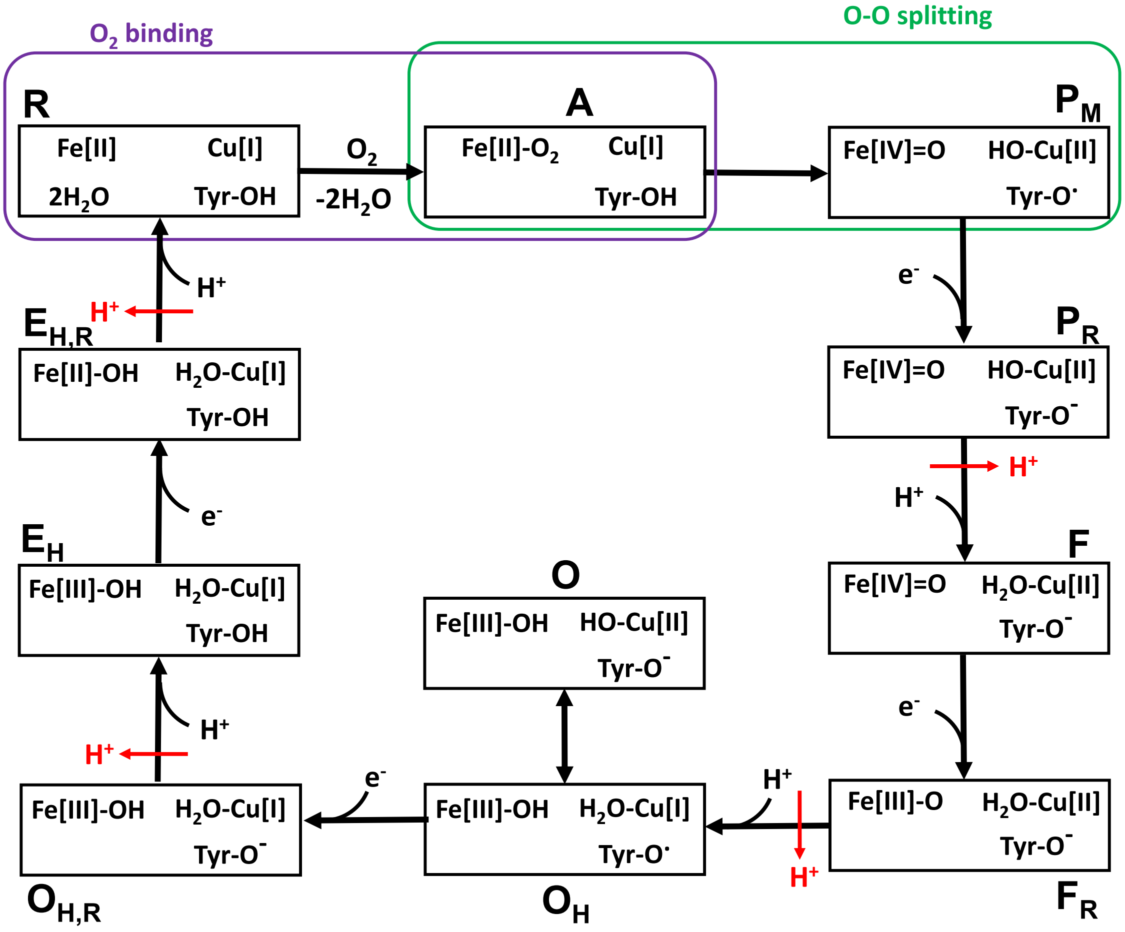

Figure 2.6: Catalytic cycle of C*c*O with intermediate redox-states at the BNC. Electrons and protons in black are transferred to the BNC for oxygen reduction chemistry. Protons in red indicate a proton pumping event. The catalysis proceeds with changes in the oxidation states of heme a_3 (Fe) and Cu_B (Cu) with the cross-linked Tyr-244 (TyrO) in protonated, deprotonated or radical forms. The reaction cycle has been adapted from Kaila *et al.* [22].

four PCET reactions in the catalytic cycle. One-electron reduction of the F state leads to the F_R state with the electron transferred to the iron of heme a_3. Protonation of the BNC through the D-channel and eventual pumping of a proton leads to the oxidized, O_H state. The reaction sequence R → F → O_H is known as the oxidative half of the catalytic cycle. The O_H state is a high energy state, with Cu_B in a strained geometry, which likely amounts to the free energy required to pump two more protons in the eventual catalytic steps [22, 130]. C*c*O can also exist in the spectroscopically indistinct, lower energy, non-pumping O state, with copper in a [II] oxidation state with a hydroxo ligand (Tyr-244 in anionic form) and the heme a_3 iron bound to a water ligand [22, 67, 82, 131]. Due to its non-pumping nature, the O state has less physiological significance [68, 131].

In the reductive half of the catalytic cycle ($O_H \rightarrow R$), following a similar sequence of PCET events, one-electron reduction of O_H leads to the $O_{H,R}$ state with the electron arriving at the Cu_B site. Transfer of a proton from the K-channel to the anionic Tyr-244 along with pumping of a third proton leads to the E_H state. One-electron reduction leads to the $E_{H,R}$ state, which upon protonation through the K-channel and translocation of a proton across the membrane gives back the reduced, R state of the enzyme [22, 68].

2.3 Proton pumping in C*c*O

The function of C*c*O as a true proton pump was shown by Wikström in 1977 [37]. The protons are pumped across the membrane against an electrochemical gradient, which is utilized universally by lifeforms for ATP production and active transport. C*c*O adds four electrons to molecular oxygen, reducing it to water. For each electron transferred to the reduction, a proton is translocated across the membrane and a proton is used for oxygen chemistry. A net charge separation of eight units takes place; four due to the electrons and protons obtained from opposite sides of the membrane, which are consumed in the BNC chemistry, and four due to the protons pumped across the membrane. While the reduction chemistry at the BNC produces different intermediates along the cycle, the proton pumping mechanism employed by C*c*O, based on mechanistic studies by Fadda *et al.* [132] can be assumed to be essentially the same in each of the four pumping events in the cycle. Though the mechanism of redox-coupled proton transfer in C*c*O has been extensively studied over the years, several aspects of the proton pumping machinery, like, the location of the PLS, proton exit pathway, gating of protons in the active site and the mechanism of proton loading, remain elusive. [21, 22, 82].

2.3.1 The proton-loading site

The pumped protons, taken up from the N-side through the D-channel are known to be transiently stored at a proton-loading site (PLS) before being expelled out to the P-side of the membrane [98, 103]. Mechanistically, the PLS demands to be spatially distinct from the BNC with a lack of a possible protonic connectivity to the BNC to prevent the consumption of pumped protons for oxygen chemistry, and essentially short-circuiting the C*c*O reaction. Additionally, the transient nature of proton storage at the PLS necessitates the pK_a of PLS to be sensitive to redox changes in the active site of C*c*O. Indirect evidence from mutagenesis experiments indicate that deprotonation of Glu-242 induces conformational changes near the D-propionate group of heme a_3, suggesting protonation of the PLS and hence asserting that the D-propionate of heme a_3 is in the vicinity of the PLS [54, 96, 133–135]. IR studies showed that reduction of heme a leads to spectral changes that were assigned to the guanidinium groups of Arg-438, which forms an ion-pair with the D-propionate group of heme a_3, indicating protonation of the PLS [136, 137]. Computational studies using Continuum Electrostatics and Molecular Dynamics methods suggested that the reduction of heme a causes dissociation of the hydrogen bonded network around the A-propionate group of heme a_3, which led to the proposition that the A-propionate group of heme a_3 could serve as the PLS [138]. Quantum chemical (DFT) calculations also suggested that the A-propionate group of heme a_3 is likely to serve as the PLS [139]. Several groups have also suggested histidine of the Cu_B site to be the PLS, based on continuum electrostatics and DFT studies [140, 141]. A recent study combining molecular mechanics and continuum electrostatics studies also suggested that the nature of the PLS is delocalized and it could constitute of several protonable sites above heme a_3 [142]. More recently, a structural study based on crystallographic structures suggested a proton uptake via H-channel with Mg^{2+} center serving as the PLS [87].

2.3.2 Mechanism of proton pumping

One of the earliest suggestions of proton pumping mechanisms was the histidine cycle by Morgan *et al.* [143] in 1994, wherein, the His-291 residue of the Cu_B site was proposed to be the PLS. His-291 was proposed to exist bound and dissociated to the copper atom of the Cu_B site as it cycles between the protonated and deprotonated (imidazolium/imidazolate) forms, respectively. However, the histidine-cycle model for proton-pumping failed to generate any experimental support. Experimental studies by Rich *et al.* [135, 144] suggested coupling of BNC reduction to the uptake of two protons from the N-side in accordance with the charge neutrality principle in the active site of C*c*O. Based on the available experimental data, in 1998, Michel proposed a pumping model based on electrostatic interactions, in which the pumped proton was taken up from the D-channel, which was coupled to the reduction of heme *a* and the proton was stored at the PLS in the proximity of the propionate groups of the hemes [145]. Thereafter, following the addition of another proton at the BNC, the pumped proton at the PLS is pumped out due to electrostatic repulsion from the incoming proton at BNC. However, there is no consensus yet on issues like: How does the enzyme direct protons to different protonation sites based on its redox state? What prevents the proton loaded at the PLS from leaking back towards the thermodynamically favorable N-side?

Based on MD simulations and free energy calculations, in 2003, Wikström *et al.* [146] proposed a water-gated model of proton delivery dependent upon the redox state of the enzyme. In the proposal, the proton transfer from Glu-242 to the PLS or to the BNC is governed by spatial orientations of water molecules in the non-polar cavity above Glu-242. The redox states of the enzyme were proposed to select for the proton delivery path to different site. With heme *a* reduced and the BNC in oxidized state, the water molecules orient to enable proton delivery to the PLS for pumping function, whereas upon electron transfer from heme *a* to the BNC, the water molecules reorganize to break the connectivity with the PLS, while establishing a proton delivery path to the BNC for oxygen reduction chemistry. Furthermore, the reorganization of water networks in the non-polar cavity were proposed to be a consequence of the electric field experienced by the water dipoles as a result of the redox states of heme *a* and the BNC.

Another model for proton pumping was suggested by Brzezinski *et al.* [38, 54], in which initially a proton is transferred to the BNC from the protonated Glu-242. The deprotonated Glu-242 then causes a conformational change around the D-propionate of heme a_3, as seen in mutagenesis studies. The structural changes were attributed to raise the pK_a of the putative PLS in the vicinity of the D-propionate group. The increased pK_a leads to protonation of the PLS from the D-channel. Upon reprotonation of Glu-242, the structural alterations around the D-propionate relax to their original state, reducing the pK_a of the PLS, which leads to eventual expulsion of protons to the P-side of the membrane. However, electrometric measurements have shown that electron transfer to the BNC is coupled to a proton transfer event in the enzyme suggesting PLS protonation being coupled to reduction of the BNC. Computational studies [106, 147], however suggest that reprotonation of Glu-242 occurs on faster timescales, making this mechanism of pumping less feasible.

Kim *et al.* [147–149] studied the proton pumping mechanism of C*c*O by developing minimalistic kinetic models based on available structural, energetic and kinetic data in literature and solving the kinetic master equation for efficient proton pumping scenarios. The models validate the role of Glu-242 acting as a proton valve [99] and electrostatic coupling of the proton and electron transfer processes. Additionally, the endergonic proton transfer from Glu-242 to the PLS is proposed to be facilitated by a reduced heme *a*, as proposed in the water-gate model [146] of proton pumping. The authors further suggested that the gating of protons in C*c*O can be achieved in a purely kinetic manner by modulation of barriers for electron and proton transfer processes among sites. Moreover, the model also suggests that loading of the PLS precedes the proton transfer to the BNC, and a loaded PLS facilitates the reduction and protonation of the BNC.

Warshel *et al.* [150–154] used various simulation approaches, like Monte-Carlo, semi-empiricial QM/MM, and coarse-grained methodologies to examine the electron and proton transfer processes in the wild-type and non-pumping mutants in C*c*O. The semi-macroscopic treatment enabled them to sample much longer timescales than those accessible to full atomistic simulations. Proton transfer energetics from their calculations suggest that the D-propionate group of heme a_3 might not be the PLS. The authors argue that Glu-242 side-chain rotation might not be enough to serve as a valve preventing proton leaks, as the apparent free energy barrier for rotation of Glu-242 is lower than that for other events in the PCET reaction. Another argument presented against the Glu-242 valve mechanism is the principle of microscopic reversibility which would apparently hamper Glu-242 side-chain orientation from preventing back-flow of the loaded protons from the PLS. However, these propositions have been countered by other computational and theoretical studies, which strongly suggest the involvement of Glu-242 in gating and preventing proton leaks [82, 106, 149, 155]. Moreover, the calculations were performed by constraining the side-chain conformations of key residues, which are likely to have a big influence on the proton transfer dynamics and energetics (see chapter 6).

Semi-empirical reactive molecular dynamics simulations were performed by Voth *et al.* [155–158] by introducing an excess proton in the computational setup to evaluate the energetics of proton transfer reactions in C*c*O. The simulations suggested the possibility of a water-filled D-channel to hold a hydrated proton which could be used to rapidly reprotonate Glu-242 after transferring a proton to the PLS [157]. The authors suggested that reduction of heme a is a prerequisite to proton loading at the PLS, which was modeled as the D-propionate of heme a_3 [156], in agreement with the pumping model suggested by Wikström and coworkers [146]. The simulations also support the valve mechanism of Glu-242 [99], where, in presence of reduced heme a, Glu-242 transfers its proton to the PLS and flips down to pick up another proton. The proton was transferred via the D-propionate group to the A-propionate group of heme a_3 through water molecules by the Grotthuss mechanism [155]. However, in a later work, they suggested that the construction water networks connecting proton donors and acceptors is not a necessary condition for transferring a proton, as an excess hydrated proton induces wetting of its surroundings and thus makes its own connectivities [158].

Cui *et al.* studied proton pumping using semi-empirical SCC-DFTB simulations focusing on the role of Glu-242 and hydration of the hydrophobic cavity above Glu-242 [159–162]. The D-propionate of heme a_3 was modeled as the PLS. In their simulations, upon protonation of the D-propionate group, the deprotonated Glu-242 does not flip "down", but rather stays "up", pointing towards the hydrophobic cavity and maintains a water-mediated connectivity to the protonated D-propionate group of heme a_3. The authors thus claim that the valve mechanism is unlikely to serve as a gating element in the pumping mechanism of C*c*O [159, 160]. Moreover, the authors claim that the water-gating of proton transfer [146] might not be a viable gating model due to simplistic models employed in the study (however, see Chapter 5). They also calculated the pK_a dependence of Glu-242 and of the D-propionate group (modeled as PLS) on the hydration levels of the hydrophobic cavity above Glu-242 [161]. The authors find that protonation of the D-propionate group of heme a_3 expands the cavity. The pK_a calculations suggest that the protonation of D-propionate group of heme a_3 lowers the pK_a value of Glu-242, facilitating the proton transfer to the BNC. However, no basis of protonation of the PLS is provided. Recent semi-empirical free energy calculations [162] by the group suggest that a proton is loaded to the D-propionate group semi-concertedly, while Glu-242 is reprotonated rapidly by an excess proton in the

D-channel, as previously suggested by Voth *et al.* [157]. The protonation of D-propionate group expands the cavity providing better hydration, which leads to a down-shifted pK_a of Glu-242, enabling proton transfer to the BNC.

Based on quantum chemical DFT models of the catalytic intermediates, Blomberg and Siegbahn [163, 164] calculated the thermodynamics and energetics of proton transfer in C*c*O. They evaluated the minimum energy path (MEP) to study the proton transfer and suggested a concerted mechanism for loading of the PLS. Mechanistically, the proton is loaded via a positively charged transition state which is electrostatically stabilized by the electron on the reduced heme *a*, establishing the coupling between protonation of the PLS and reduction of heme *a*. Thus, the electron on heme *a* directs the loading of the PLS, which is followed by electron transfer to the BNC, which is quickly reprotonated to prevent the back-flow of the loaded proton. However, it should be noted that quantum chemical cluster calculations do not account for dynamics, which often play a critical role in proton transfer reactions inside enzymes [93, 165].

Combining time-resolved electrometric and spectroscopic data, based on one-electron reduction of the fully oxidized enzyme, Belevich *et al.* [96] suggested a pumping mechanism in which, starting from the reduced R state, upon reduction by cytochrome *c*, a 10 μs equilibration phase of the electron between $\mathrm{Cu_A}$ and heme *a* is followed by uptake of a proton through the D-channel. The loading of the proton from Glu-242 to the PLS occurs in *ca.* 150 μs. This proton translocation is coupled to transfer of the electron from heme *a* to the BNC. However, the precise nature of the concerted PCET event remains unresolved. Thereafter, in 0.8 ms, another proton is taken up from the D-channel and transferred to the BNC via Glu-242 for the reduction chemistry. The transfer of proton to the BNC leads to relaxation of the protonated PLS in 2.6 ms and ejection of the proton to the P-side of the membrane.

3 Theoretical background of computational biochemistry

This chapter presents the theoretical framework underlying atomistic simulations of biomolecules. The first section is a short description of quantum mechanical methods involved in modeling biomolecular systems. The second section reviews the classical methods used for biomolecular simulations and the last section covers the hybrid quantum mechanics/molecular mechanics (QM/MM) approach to molecular simulations employed in this work.

3.1 Quantum chemical methods

Quantum biochemistry entails the application of the principles of quantum mechanics on physical models of biomolecules to obtain chemical and physical insights into their structure and function. In 1929, P. A. M. Dirac wrote [166]:

> The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are [thus] completely known, [and] the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.

Even now, close to a century later, obtaining exact solutions for even slightly more complex molecules (hydrogen molecular ion, ${H_2}^+$, is the largest molecular system with an exact analytical solution) still remains a highly intractable problem. Dirac's philosophy of reductionism served as the guiding light for development of a myriad of quantum chemical methods in the years following. Nowadays, with constantly increasing computing power, it is not unusual to apply quantum chemistry methods to models of molecular systems with hundreds of atoms and obtain accurate structural and energetic predictions.

3.1.1 The Schrödinger equation

Quantum mechanics postulates that there exists a wave function that completely describes the state of the system of electrons and nuclei. In 1926, Erwin Schrödinger introduced the time-dependent Schrödinger equation (TDSE) [167] to describe such a system in terms of its wave function,

$$\hat{H}\Psi(r,t) = i\hbar\frac{\partial\Psi(r,t)}{\partial t}, \tag{3.1}$$

where $\hat{H}$ is the Hamiltonian operator for the system and Ψ the wave function. The Hamiltonian is a quantum mechanical operator associated with the system's kinetic and potential energies. For bound systems where the Hamiltonian operator is independent of time, the equation can be solved as an eigenvalue equation, relating the wave function to the energy of the system by the time-independent Schrödinger equation (TISE),

$$\hat{H}\Psi(r) = E\Psi(r), \tag{3.2}$$

where E the energy of the system.

The Hamiltonian is given by the sum of all energy terms, the kinetic ($\hat{T}$) and the potential ($\hat{V}$) energies of the system. For a general molecular system consisting N_n nuclei and N_e electrons, the Hamiltonian can be written as

$$\hat{H} = \underbrace{-\sum_{I=1}^{N_n} \frac{\nabla_I^2}{2M_I}}_{\hat{T}^n} + \underbrace{\sum_{I=1}^{N_n}\sum_{J>I}^{N_n} \frac{Z_I Z_J}{|R_I - R_J|}}_{\hat{V}^{nn}} \underbrace{- \sum_{i=1}^{N_e} \frac{\nabla_i^2}{2}}_{\hat{T}^e} + \underbrace{\sum_{i=1}^{N_e}\sum_{j>i}^{N_e} \frac{1}{|r_i - r_j|}}_{\hat{V}^{ee}} + \underbrace{\sum_{i=1}^{N_e}\sum_{J=1}^{N_n} \frac{-Z_J}{|r_i - R_J|}}_{\hat{V}^{ne}} \tag{3.3}$$

in atomic units with M representing nuclear mass, Z representing nuclear charge, R for nuclear coordinates and r for electronic coordinates. Using this Hamiltonian formulation, one can, in principle, solve the TISE for a given $\Psi(r)$ to obtain energy eigenvalues of the system. The analytical solution is, however, not possible for systems exceeding two particles.

3.1.2 The Born-Oppenheimer approximation

The Born-Oppenheimer approximation (BOA) [168, 169] states that the motions of electrons and nuclei can be decoupled because nuclei are much heavier as compared to electrons. For instance, for the hydrogen atom, its nuclear mass is *ca.* 1840 times heavier than the mass of the electron. This disparity in masses causes nuclei to react much slower to external perturbations as compared to electrons. From an electron's point of view, nuclei appear to be fixed or the electrons experience nuclei as a fixed external potential ($\hat{V}^{ne}$), whereas with respect to the nuclei, electrons can be considered fully equilibrated. This enables us to write the wave function as product of electronic and nuclear wave functions,

$$\Psi(R, r) = \Psi_{nuc}(R)\Psi_{elec}(r), \tag{3.4}$$

where R and r represent nuclear and electronic coordinate variables, respectively. Using the BOA, one can solve the electronic Schrödinger equation (SE) as,

$$(\hat{H}_{elec} + \hat{V}^{ne})\Psi_{elec}(r; R) = E_{elec}(R)\Psi_{elec}(r; R), \tag{3.5}$$

where the electronic wave function, Ψ_{elec}, depends parametrically on the nuclear coordinates defined by the structure of the molecular system. The eigenvalues, $E_{elec}(R)$, give the electronic state energies of the molecule for the molecular structure. The electronic energy, E_{elec}, along with the nuclear repulsion term, $\hat{V}^{nn}$, gives the BO potential energy surface. The BOA sets up the notion of potential energy surface (PES), which is central for describing structure and dynamics of molecular systems [170].

3.1.3 The Hartree-Fock approximation

To solve the SE for more complicated multi-electron quantum systems, Hartree approximation was introduced by Hartree in 1927 [171], in which, the total electronic wave function is written as the product of individual one-electron functions, known as Hartree Product,

$$\Phi(r_1, ..., r_N) = \prod_{i=1}^{N} \phi_i(r_i). \tag{3.6}$$

For the electronic Hamiltonian, all terms except the electron-electron repulsion term, $\hat{V}^{ee}$, have only one electronic summation index. By making an approximation where each electron feels only the average effect of the rest of the electrons, the Hamiltonian can be written as a sum of one-electron Hamiltonians

$$\hat{H} = \sum_{i=1}^{N} \hat{h}_i. \tag{3.7}$$

This formulation gives us the framework to solve the N one-electron SEs for an N-electron system and to break down a $3N$ variable SE into N three variable equations, reducing the computational complexity. However, the Hartree approximation violates the antisymmetry requirement, imposed by Pauli's exclusion principle [172], in which,

$$\Phi(r_1, r_2) = \phi_1(r_1)\phi_2(r_2) \neq \phi_1(r_2)\phi_2(r_1) = \Phi(r_2, r_1) \neq -\Phi(r_1, r_2). \tag{3.8}$$

In multi-electron systems, one must also account for electronic spin, which is given by the spin quantum number s. The electron can exist in two states corresponding to $s = 1/2$, which is given by the quantum number m_s, which can have values of $+1/2$ or $-1/2$, referred to as "up" and "down" spins, respectively. Spin is effectively included in the SE formulation by writing each one-electron wave functions as a product of a spatial function and a spin function. The product is called a spin orbital. The spatial function corresponds to distribution of electron density in space while the spin function accounts for "up" or "down" (α or β) spin states of the electron. In 1929, Slater [173] therefore proposed the determinant,

$$\Psi(x_1, x_2, ..., x_n) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \cdots & \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \cdots & \chi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_N) & \chi_2(x_N) & \cdots & \chi_N(x_N) \end{vmatrix}, \tag{3.9}$$

where, χ represents spin orbitals, which depends on both the spatial coordinates r and the spin coordinate (α or β). Exchanging any two rows of the Slater determinant (SD), corresponds to exchange of two electrons of the system, leads to a change in the sign of the determinant, thus satisfying the antisymmetry principle. The $1/\sqrt{N!}$ ensures a normalized wave function. In the case when two rows of the SD are same, the SD vanishes, complying with Pauli's principle [172].

Fock extended the Hartree method by incorporating the Slater determinant formalism to give the Hartree-Fock (HF) equations [174, 175]. The HF method starts by guessing the N one-electron spin orbitals and construction of a Slater determinant. The HF equation is solved iteratively in a self-consistent manner, until the electron density and the energy converges. This is known as the self-consistent field (SCF) approach. The HF equation can be written as,

$$\hat{F}\psi_i = \epsilon_i \psi_i, \tag{3.10}$$

where $\hat{F}$ is the Fock operator for an electron defined as,

$$\hat{F}(1) = \hat{h}(1) + \sum_{j=1}^{N} [\hat{J}_j(1) - \hat{K}_j(1)]. \tag{3.11}$$

For closed shell systems, the Fock operator takes the form,

$$\hat{F}(1) = \hat{h}(1) + \sum_{j=1}^{N/2}[2\hat{J}_j(1) - \hat{K}_j(1)], \tag{3.12}$$

and,

$$\hat{h}(1) = -\frac{1}{2}\nabla_1^2 - \sum_j \frac{Z_j}{r_{1j}} \tag{3.13}$$

$$\hat{J}_j(1)\chi_i(1) = \left[\int \chi_i^*(2)\frac{1}{r_{12}}\chi_i(2)dx_2\right]\chi_i(1) \tag{3.14}$$

$$\hat{K}_j(1)\chi_i(1) = \left[\int \chi_j^*(2)\frac{1}{r_{12}}\chi_i(2)dx_2\right]\chi_j(1) \tag{3.15}$$

where $\hat{h}$ is the one-electron Hamiltonian operator. $\hat{J}$ and $\hat{K}$ are two-electron operators called the Coulomb and exchange operators, respectively. The Coulomb term accounts for the electrostatic repulsion felt by an electron in the mean-field of all other electrons, whereas the exchange term results from the antisymmetric requirement of the wave function. Unlike the former, the exchange term has no classical explanation associated to it. The Coulomb and the exchange terms account for electron-electron interaction in the multi-electron system. In the HF treatment, the unphysical self-interaction of an electron with itself, arising from the Coulombic interaction is explicitly accounted for, since $\hat{J}_{ii} - \hat{K}_{ii}$ explicitly cancels.

In 1951, Roothan and Hall [176, 177], independently reformulated the HF equations into a computationally tractable linear algebraic matrix form. The equations are called the Roothan-Hall equations, where one-electron orbitals are replaced by a linear combination of basis functions (see section 3.1.4),

$$\phi_i = \sum a_{ij}\chi_j, \tag{3.16}$$

The χ_i and ϕ_i are the atomic and molecular orbitals, respectively, and the treatment in Eq. 3.16 is referred to as the linear combination of atomic orbitals to make molecular orbitals (LCAO-MO). Using the formulation in Eq. 3.16, HF equations can be written as

$$F_i \sum_{\alpha}^{M_{basis}} c_{\alpha i}\chi_\alpha = \epsilon_i \sum_{\alpha}^{M_{basis}} c_{\alpha i}\chi_\alpha \tag{3.17}$$

Multiplying from the left by the basis functions followed by integration gives the eigenvalue Roothan-Hall equations:

$$FC = SC\epsilon \tag{3.18}$$

where,

$$F_{\alpha\beta} = \langle\chi_\alpha|F|\chi_\beta\rangle \tag{3.19}$$

$$S_{\alpha\beta} = \langle\chi_\alpha|\chi_\beta\rangle. \tag{3.20}$$

F is the Fock matrix, S is the overlap matrix of the basis functions, C is a matrix of atomic orbital coefficients and ϵ is a diagonal matrix of orbital energies. The equations are solved in a self consistent way, until convergence is achieved.

Although, the HF treatment accounts for the antisymmetry and indistinguishability of electrons, it does not account for the electron correlation, due to the mean-field approximation. The energy difference between the exact and HF energies is known as correlation energy as defined by Löwdin [178]. Another limitation of the HF method is the usage of a single Slater determinants to approximate the wave function (See section 3.1.5). For example, dispersion interactions, which are often important in biomolecular systems [179], result from electron correlation effects [180–182].

3.1.4 Basis sets

In quantum chemistry, basis sets refer to a set of one-particle functions, whose linear combination yields molecular orbitals, i.e., the LCAO-MO approach (Eq. 3.16). The basis sets need not be formed of exact atomic orbitals obtained as a solution to the HF equations, but they can be any set of convenient functions whose linear combination results in reasonable and well-behaved molecular orbitals.

The Slater-type orbitals (STOs) [175] introduced by Slater in 1930, were one of the first set of basis functions used to solve HF equations in a numerical way. The STOs are based on the orbitals obtained from solving SE for one-electron quantum systems. Though STOs may appear to be the natural choice, the calculation of the two-electron integrals gets computationally exhaustive. Gaussian-type orbitals (GTOs) are thus, more widely used in most modern quantum chemistry packages, as the evaluation of the two-electron integral is analytically possible and simpler to compute than STOs [183]. GTOs have the following functional form in 3D space,

$$\phi(x,y,z;\alpha,i,j,k) = \left(\frac{2\alpha}{\pi}\right)^{3/4} \left[\frac{(8\alpha)^{i+j+k} i!j!k!}{(2i)!(2j)!(2k)!}\right]^{1/2} x^i y^j z^k e^{-\alpha(x^2+y^2+z^2)}, \tag{3.21}$$

where, α governs the width of the gaussian function, and i, j, and k are represent the quantum numbers of the electronic orbital ($i+j+k=0$ for s function, and $i+j+k=1$ for p function, and so on). Typically GTOs are used in contracted form (linear combinations of several gaussians) to account for cusp in the electron density close to nuclei and exponential asymptotic behavior or atomic orbitals.
For a minimal basis set, one basis function is used to represent each atomic orbital. In computational chemistry, the STO-nG minimal basis sets by John Pople *et al.* [184] employ a contraction of n gaussians are fitted to model each Slater type function. The minimal basis sets, however, fail to produce accurate energies, which is essential to model molecular systems, where neighboring atoms further affect the shape of the orbital. Therefore, more gaussians are added to increase the accuracy of the basis set. Doubling and tripling the number of functions for each orbital gives double- and triple-ζ basis-sets, respectively. From the principles of chemical bonding, we know that mainly the valence electrons take part in the bonding. To account for this, split-valence basis sets were introduced, in which core orbitals are given by a single (contracted) basis function, while valence orbitals are split into multiple functions. Popular split-valence basis sets by Pople *et al.* [185], known as the Pople basis sets are denoted by the notation *X-YZg*. The number before indicates the number of contracted gaussians used for core orbital, and the numbers after the hyphen indicate the number of gaussians used in the valence functions. Two, three and four digits after hyphen indicate double-, triple- or quadruple-ζ basis functions. Furthermore, polarization (denoted by asterisk *) and diffuse functions (denoted by +) are added to basis sets to account for distortion and exponential decay of valence orbitals, respectively. Alternate split-valence basis sets developed by Ahlrichs and coworkers [186] known as Karlsruhe basis sets have been employed for this work. In the Karlsruhe nota-

tion, SV stands for split-valence, P stands for polarized, TZV denotes valence triple-zeta and D refers to inclusion of diffuse functions. In this work def2-SVP and def2-TZVP basis sets have been used most often for geometry optimization, dynamics and single point energy calculation and are computationally comparable to Pople basis sets, 6-31G** and 6-311G**, respectively.

3.1.5 Electron correlation methods

Electron correlation can be described as correlated motion or instantaneous interactions of each electron with every other electron in the quantum system. This effect is neglected in HF theory due to the mean field approximation. The HF method overestimates electron–electron repulsion and hence, overestimates the electronic energies. The HF method typically accounts for almost 99% of the total energy of the system [187]. However, the remaining 1% of the energy, termed the correlation energy [178], is often crucial for the chemical description, such as, bond breakage, bond formation and dispersive effects in the quantum system. There are two kinds of electron correlation effects: (1) Dynamic correlation, a consequence of electrons, especially of opposite spin, avoiding each other and, (2) Static correlation, due to the inability of a single Slater determinant to account for the state due to near degeneracy different electronic configurations [187].

A perturbation theory based approach to account for the correlation energy is given by Møller-Plesset (MP) correction [188]. The second order energy correction, MP2, can be written as,

$$E_{\mathrm{MP2}} = \frac{1}{4}\sum_{ijab} \frac{|\langle ij|ab\rangle - \langle ab|ij\rangle|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \tag{3.22}$$

where i, j correspond to the occupied orbitals, and a, b are the virtual orbitals. The numerator accounts for the Coulomb and the exchange energies between electron pairs, and ϵ denotes the respective orbital energies. A typical MP2 calculation can usually recover about 80-90% of the correlation energy. MP2 scales as N^5, where N is the number of basis functions, which is an order higher than HF calculations, which scale as N^4. It should be noted that MP methods are based on the HF orbitals subjected to the mean-field approximation. Moreover, MP treatment is perturbative and not variational, and hence there is no assured convergence. Therefore, for MPn treatments with increasing order of correction, n, the energies obtained are usually oscillatory in nature. MP2, however, accounts for dispersive interactions which are important in biomolecular systems. The computational cost for MP2 calculations can be reduced by using resolution of identity (RI) approach to the MP2 correlation energy (RI-MP2) [189]. MP2 calculations were used to benchmark protein affinities and dispersion interaction in a side project during this work (see Publications).

Configuration Interaction (CI) is based on variational principle and is similar in treatment to the HF method. A trial wave function is constructed as a linear combination of ground state HF solution and excited state. A truncation at single and doubles excitations is called CISD, and Slater determinants are constructed as in HF calculations. A full CI calculation (FCI), which accounts for all n electrons in K orbitals, the number of permutations in a CI calculations is $(2K!)/[N!(2K-N)!]$, which becomes exponentially large high values of K ($K^N/N!$). Hence, while FCI calculations in the basis set limit are exact, the calculations are practically not possible. Truncated CI approaches are good at retrieving the correlation energies, and are computationally tractable (CISD scales as N^6), but they suffer from size-extensivity and need of large basis sets for convergence [187, 190].

Coupled cluster (CC) theory also builds on HF method, but unlike CI treatment, which is based on linear combination of excited determinants, the trial wave function is given by a non-linear exponential formulation. CC methods do not suffer with size-extensivity and computationally scale similarly to CI methods at the level of excitations taken into consideration. The CCSD(T) method, which involved coupled cluster treatment with single and double excitations with a perturbative correction for triples contribution is popular in computational theory and is often called as "gold standard" in single reference calculations for organic molecules [191, 192]. CCSD and CCSD(T) calculations have been employed in this work for benchmarking purposes.

3.1.6 Density Functional Theory

Until this point, in all descriptions pertaining to the electronic structure of atoms and molecules, the key descriptor has been the wave function Ψ. It is through complex operations on the wave function that we obtain meaningful chemical insights into the structure and energetics of molecules. The wave function itself is a complicated entity depending $3N$ spatial coordinates for an N-electron system, with each increasing electron, making this problem even more difficult to solve. Density Functional Theory (DFT) is a reformalism of the electronic structure problem based on the electron density, $n(r)$, reducing the $3N$ dimensions to 3 dimensions.

Shortly after the introduction of the wave equation by Erwin Schrödinger in 1927, Thomas and Fermi proposed in 1927, the first version of DFT, wherein they described the total energy of the system as a functional of the electron density $n(r)$. Though the Thomas-Fermi model and later improved Thomas-Fermi-Dirac model provided some estimates for atomic properties, the models failed to describe the molecular chemistry in any considerable detail due to approximation of the kinetic energy term, errors in the exchange energy, and no accounting for the correlation energy [193–196].

The Hohenberg-Kohn Theorems

The basis of modern DFT is based on two fundamental theorems by Hohenberg and Kohn (HK). In DFT, the electrons ($n(r)$) interact with the an external potential, V_{ext}. The quantity of interest $n(r)$ in terms of the N-electron wave function Ψ is given by

$$n(r) = N \int \dots \int \Psi^*(r_1, r_2, \dots, r_N)\Psi(r_1, r_2, \dots, r_N) dr_2 \dots dr_N, \tag{3.23}$$

where,

$$N = \int n(r) dr. \tag{3.24}$$

The first HK theorem states that the external potential and hence the total energy, is a unique functional of the electron density [197].

For a non-degenerate ground state, let us assume two external potentials v_a and v_b consistent with ground state density $n_0(r)$. The corresponding Hamiltonians H_a and H_b would also have ground state wave functions Ψ_a and Ψ_b also consistent with the density $n_0(r)$.

Using the variational principle,

$$\begin{aligned}\langle\Psi_a|H_a|\Psi_a\rangle &< \langle\Psi_b|H_a|\Psi_b\rangle\\ E_a &< \langle\Psi_b|H_b|\Psi_b\rangle + \langle\Psi_b|H_a - H_b|\Psi_b\rangle\\ E_a &< E_b + \int n_0(r)[v_a(r) - v_b(r))]dr\end{aligned} \tag{3.25}$$

and similarly exchanging indices a and b, we have,

$$E_b < E_a + \int n_0(r)[v_b(r) - v_a(r))]dr. \tag{3.26}$$

Adding the two inequalities, the density integral vanishes to give the impossible condition,

$$E_a + E_b < E_b + E_a. \tag{3.27}$$

Hence, the assumption leads to a contradiction and proves that the electron density has a unique one-to-one mapping with the external potential and also the wave function as given by the first HK theorem. Thus, if we have the ground state electron density $n_0(r)$, we have all the information about the ground state system in question.

The second HK theorem states that there exists an energy functional, $E[n(r)]$, depending on the electron density, $n(r)$, for any external potential. The ground state energy of this system is the global minimum value of this functional, and the electron density, $n(r)$, that minimizes the functional is the exact ground state electron density, $n_0(r)$ [197]. This energy functional is given by,

$$E[n] = F[n] + \int v(r)n(r)dr, \tag{3.28}$$

and it can be variationally optimized for a set of trial densities to reach closer to the ground state electron density. The HK theorems establish the correspondence between the electron density, wave function and the external potential while also providing a variational principle for the energy functional. However, the HK theorems still lack to provide a practical recipe for electronic structure calculations.

The Kohn-Sham Formulation

The Kohn-Sham (KS) theory [198] is based on an approximation that the ground state density of an N-particle system equals that of some other system, in which, the particles do not interact. This reduces the complexity of treating a system of N interacting electrons, to a system with N individual noninteracting electrons, an approach similar in philosophy to Hartree treatment for wave functions. The energy functional can be written as,

$$E[n] = \hat{T}_0[n] + \int [\hat{V}_{ext}(r) + \hat{U}_{el}(r)]n(r)dr + E_{xc}[n], \tag{3.29}$$

where $\hat{T}_0$ corresponds to the kinetic energy of a system with density n and lacking electron–electron interactions. The next term due to electrostatic electron-electron interaction is given by,

$$\hat{U}_{el}(r) = \int \frac{n(r)}{|r - r'|}dr'. \tag{3.30}$$

This is a purely classical term which, and unlike the HF treatment, where self-interaction is exactly canceled, $\hat{U}_{el}$ term includes the self-energy of the density with itself that needs to be corrected for. The term as a result of an external potential is given by,

$$\hat{V}_{ext}(r) = \sum_{a} \frac{-Z_a}{|R_a - r|}. \tag{3.31}$$

The last term E_{xc} is known as exchange-correlation energy, which accounts for all the missing descriptions: electron exchange, electron correlation effects, and a correction to the kinetic energy due non-interacting assumption and self-interaction.

From HK theorems, we can apply the variational principle to the energy functional to get,

$$\begin{aligned} \mu &= \frac{\delta E[n(r)]}{\delta n(r)} \\ &= \frac{\delta T_0[n(r)]}{\delta n(r)} + \hat{V}_{eff}(r), \end{aligned} \tag{3.32}$$

where, $\hat{V}_{eff}(r)$ is the effective potential from combining all remaining terms and is given by,

$$\hat{V}_{eff}(r) = \hat{V}_{el}(r) + \hat{V}_{ext}(r) + \hat{V}_{xc}(r), \tag{3.33}$$

where,

$$\hat{V}_{xc}(r) = \frac{\delta E_{xc}[n(r)]}{\delta n(r)}. \tag{3.34}$$

Eq. 3.32 can be solved for as SE for non-interacting particles,

$$\left[\frac{1}{2}\nabla^2 + \hat{V}_{eff}(r)\right] \phi_i(r) = \epsilon_i \phi_i(r), \tag{3.35}$$

where, ϕ_i are the Kohn-Sham (KS) orbitals. It should be noted the KS orbitals only depend on the coordinate r. The orbitals obtained from solving the KS equation can be used to calculate the electron density,

$$n(r) = \sum_{i=1}^{N} |\phi_i|^2. \tag{3.36}$$

The kinetic energy term is evaluated directly from the KS orbitals. The obtained density can be used to obtain a better description of $\hat{V}_{eff}$ providing an SCF procedure. The procedure proceeds in a similar way as in the HF methods except that instead of the exchange term of HF, in the KS treatment there is an exchange-correlation term. This makes the KS treatment formally exact, though the exact form of exchange-correlation functional is still unknown. Approximations to estimate the exchange-correlation functionals in DFT formalism are discussed in the next subsection.

The Exchange Correlation Energy

Ever since the introduction of DFT, a large number of approximate exchange-correlation functionals, both non-empirical (based on theoretical considerations) and semi-empirical (based on fitting to experimental data) have been suggested. The choice of approximation strategy, nonetheless, depends on physical quantities we are interested in obtaining and computing power at hand. A brief discussion of different functionals is provided below.

The local density approximation (LDA) [198, 199] is given as,

$$E_{xc}^{LDA} = \int n(r)\epsilon_{xc}[n(r)]dr, \tag{3.37}$$

where, ϵ is the exchange and correlation energy functional of a homogeneous electron gas with density n. For spin polarized systems, the Local Spin Density Approximation (LSDA) was also given [200]. The exchange-correlation is split into individual exchange and correlation terms. The analytic form for the exchange term for the homogeneous electron gas is known, while analytic forms for the correlation term have been parametrized on the basis of accurate quantum Monte Carlo calculations [201, 202]. The LDA approach is known to fail at the surface of the atoms and hence rendering it unsuitable for application to chemical systems. The LDA approach, however, works well for solids, which have a slowly varying electron density [199].

The Generalized Gradient Approximation (GGA) attempts to overcome the shortcomings of the LDA by including the gradient of the density in the description of the exchange-correlation functional. In molecules, the electron density can vary rapidly over a small region of space leading to the breakdown of LDA. The exchange correlation term, E_{xc}, is expanded in the Taylor series versus density n and truncated at a linear term to also include the gradient. The general form of GGA functionals can be written as,

$$E_{xc}^{GGA} = \int n(r)\epsilon_{xc}[n(r), \nabla n(r)]dr. \tag{3.38}$$

The first popular GGA functional was given by Becke in 1988 [203], often known as Becke88 as a correction to the LDA exchange energy

$$\begin{aligned} E_x^{B88} &= E_x^{LDA} + \Delta E_x^{B88} \\ \Delta E_x^{B88} &= -\beta n^{1/3} \frac{x^2}{1 + 6\beta x \sinh^{-1} x} \\ x &= \frac{|n|}{n^{4/3}} \end{aligned} \tag{3.39}$$

where x is a dimensionless ratio, and β is a parameter used to fit the exchange functional to known exchange energies for noble atoms.

Another popular GGA functional for correlation energy was given by Lee, Yang and Parr (LYP) [204], which comprises four parameters obtained by fitting data for helium atom. Other functionals in chemistry given by Perdew *et al.* [205–207] include PW86 (Perdew-Wang), PW91 and PBE (Perdew-Burke-Ernzerhof), which depend on non-empirical parameters. The LYP correlation function is often used with the B88 exchange functional for exchange-correlation energy. This treatment is referred to in literature as the BLYP functional and is widely used in computational chemistry. GGA functionals present a large improvement over LDA results with energetics comparable to correlated methods like MP2 [208]. BP86 functional, comprising B88 exchange functional [203] and the PW86 correlation functional [205], has been used in this work for DFT calculations.

In an effort to improve on GGAs, meta-GGAs were introduced which, along with the gradient of density, also used the Laplacian as a variable for evaluating the exchange-correlation function. The general form of meta-GGAs can be written as,

$$E_{xc}^{m-GGA} = \int F(n(r), \nabla n(r), \nabla^2 n(r))dr. \tag{3.40}$$

meta-GGAs perform better than GGAs and meta-GGAs like TPSS [209] have been widely used in computational chemistry and biological systems [210]. The TPSS functional was used for benchmarking purposes in this work.

Hybrid functionals are based on the premise that HF calculations provide an exact form of exchange energy. Initially, attempts were made to write the exchange-correlation energy as a sum of HF exchange energy and correlation energy from LDA. This approach, however, failed to accurately account for exchange-correlation energy. Becke [211] suggested an alternative approach for exchange-correlation energy,

$$E_{xc} = \int_0^1 U_{xc}^{\lambda} d\lambda, \tag{3.41}$$

where, λ is an inter-electronic coupling parameter going from 0 to 1 and U_{xc}^{λ} is the potential energy of exchange correlation for a coupling value of λ. A value of zero corresponds to no coupling, this is, no Coulombic repulsion between electrons or a non-interacting system. Whereas, a value of one corresponds to the real system with full Coulomb repulsion. In practice, the integral is difficult to evaluate analytically and is evaluated by numerical approximations, for example, the half and half linear interpolation

$$E_{xc} = \frac{1}{2}(U_{xc}^0 + U_{xc}^1) \tag{3.42}$$

where U_{xc}^0 accounts for exact exchange due to non-interacting system of electrons. The U_{xc}^1 corresponds to the exchange-correlation energy from LDA treatment. Naturally, more accuracy can be achieved by introducing a GGA corrections. Becke introduced three parameters semi-empirically fitted hybrid functional commonly known as the B3 functional. The B3 functional can be written as,

$$E_{xc} = a * E_x^{Slater} + (1-a)E_x^{HF} + b * \Delta E_x^{B88} + E_c^{VWN} + c\Delta E_c^{non-local}. \tag{3.43}$$

Becke [212] also introduced the B3PW91 functional in 1993. The functional can be written as,

$$E_{xc}^{B3PW91} = (1-a)E_x^{LDA} + aE_x^{HF} + b\Delta E_x^{B} + E_c^{LDA} + c\Delta E_c^{PW91}, \tag{3.44}$$

with parameter values, a=0.2, b=0.72 and c=0.81. Among most modern functionals in computational chemistry, the B3LYP (Becke, 3-parameter, with correlation from LYP functional) [213] is the most popular functional of choice. It is similar in form to B3PW91 functional, except PW91 is replaced by LYP. Unlike PW91, LYP attempts to account for full correlation energy and can be written as,

$$E_{xc}^{B3LYP} = (1-a)E_x^{LDA} + aE_x^{HF} + b\Delta E_x^{B} + (1-c)E_c^{LDA} + cE_c^{LYP}, \tag{3.45}$$

with same parameter values as in B3PW91 functional. B3LYP functional was also prominently used in this work for DFT calculations.

Dispersion corrected DFT

DFT calculations are very accurate and produce $< 1\%$ energy error for chemical processes [214]. However, DFT fails to account for dispersion interactions, due to electron correlation effects, which are often important in biomolecular systems. It is known that LDAs and GGAs cannot describe the long-range behavior of the vdW interactions (r^{-6} term) [182]. Grimme *et al.* [214] introduced dispersion corrected DFT to include dispersion interactions in the DFT formalism. The correction to the energy can be given as,

$$E_{\text{DFT-D3}} = E_{\text{DFT}} - E_{\text{disp}}, \tag{3.46}$$

where, E_{DFT} is the energy from DFT, and E_{disp} is the energy correction term due to dispersive interactions. E_{disp} is an addition of the dispersion energy due to two- and three-body terms. The correction is semi-classical in nature. DFT-D3 [215] was employed in this work for all DFT calculations.

3.2 Classical biomolecular simulations

Quantum mechanical methods described in the previous sections can be very accurate at the atomic level and are essential for understanding the chemistry of bond breakage and bond formation. Biomolecules are generally orders of magnitude larger in size than the typical molecular systems for which full quantum mechanical treatment is possible. Moreover, for large enough systems and for processes pertaining to longer timescales, independent quantum events are averaged over as a consequence the Law of Large Numbers, making the system deterministic and classical in behavior. In order to model such large biomolecular systems, approximate methods to calculate potential and kinetic energies are needed. The theoretical framework for classical atomistic simulations employed in this work is discussed in this section.

3.2.1 Molecular dynamics

Classical molecular dynamics (MD) simulation techniques are based on classical physics where the time evolution of a set of interacting atoms is followed by integrating Newton's equations of motion based on a parametrized force field (see subsection 3.2.2). Useful equilibrium properties can then be obtained from the simulation by subjecting it to statistical mechanics treatment. MD was originally developed in 1957 by Alder and Wainwright [216]. In 1971 and 1974, Rahman and Stillinger [217, 218] used MD to simulate a realistic system of liquid water, following which, the first MD simulation of the protein, Bovine pancreatic trypsin inhibitor (BPTI) was carried out by McCammon *et al.* in 1977 [219]. With increasing computing power, and an increasing number of experimentally determined structures available, MD has become a central tool, serving as a "computational microscope" in biochemistry and molecular biology. [220]

Classical mechanics

For a set of N interacting particles, the classical equations of motion are given by Newton's second law of motion [221],

$$F_i = m_i \frac{d^2}{dt^2} r_i(t). \tag{3.47}$$

Force can also be written as the gradient of the potential energy,

$$F_i = -\nabla_i U, \tag{3.48}$$

where the potential energy U is a function of all $3N$ coordinates. Eq. 3.47 and Eq. 3.48 give the differential equation,

$$F_i = -\nabla_i U = m_i \frac{d^2}{dt^2} r_i(t) \tag{3.49}$$

In MD, this differential equation is numerically integrated to obtain the time evolution or the trajectory for the simulated system.

Statistical mechanics

On an algorithmic level, MD simulations deal with microscopic quantities. MD algorithms take positions and velocities as inputs and give positions and velocities as outputs. Macroscopic properties, like pressure, diffusion constants, relaxation properties, free energies, etc., can be computed by relationships from statistical mechanics.

The phase space is defined as a $6N$ dimensional space consisting positions ($\boldsymbol{x}^N$) and momenta ($\boldsymbol{p}^N$) for all N particles and a microstate corresponds to a point in phase space. The time ordered series of microstates obtained from the simulation is referred to as the trajectory of the simulation. Macroscopically, the thermodynamic state of a system is given by observable parameters like pressure, volume and temperature, such that all other thermodynamic information can be derived [222]. An ensemble can thus be defined as a collection of microstates of the system subjected to macroscopic constraints.

Average macroscopic quantities as observed in experiments are ensemble averages due to Avogadro number of copies (microstates) under investigation. Hence, in order to make connection with the experiment, ensemble average must be evaluated as,

$$\langle A \rangle = \frac{\int A(\xi) w(\xi, t) d\xi}{\int w(\xi, t) d\xi}, \tag{3.50}$$

where, w represents the probability density, $A(\xi)$ is an observable property as a function of ξ and $\langle A \rangle$ is the ensemble average. We know after a sufficiently long time, system will reach an equilibrium, i.e., the probability density, w, and thus the ensemble average, $\langle A \rangle$ will become time-independent,

$$\lim_{t \to \inf} \langle A \rangle (t) = \langle A \rangle_{eq}. \tag{3.51}$$

The ergodic hypothesis in statistical mechanics allows to establish equivalence of ensemble averages and time ensembles in the limit of large number of microstates and long time-evolution, respectively. This allows us to obtain macroscopic observables by time averaging MD trajectories, given the system has reached a state of equilibrium.

In MD simulations, when total number of atoms N, the volume V, and the temperature, T are kept constant, the simulation models a canonical (NVT) ensemble. NVT ensemble is the most commonly used ensemble for biomolecular systems. In NVT ensemble, the probability of finding the system in a microstate i with energy E_i is given by the Boltzmann distribution,

$$w_i = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}}, \tag{3.52}$$

where, β is the inverse thermodynamic temperature ($(k_B T)^{-1}$). Other simulation ensembles used in MD simulations are NVE, NPT, μVT ensembles where E is energy, P is pressure, and μ is chemical potential [223, 224].

3.2.2 Force fields

The force field is a functional parametrized form that allows rapid evaluation of the force on each atom due to the interactions with all other particles (atoms) in the system. [225–227]

In the force field approach, electrons are not explicitly considered, while the nuclei move on a effective parametrized PES. The parameters are obtained from experimental values or quantum chemical calculations [225–227]. The potential energy function representing interaction of all particles in the system can be written as,

$$U(r) = U_{bonded} + U_{non-bonded}. \tag{3.53}$$

The bonded term corresponding to covalently bonded atoms along with their neighboring atoms comprises bond and angle terms, modeled as harmonic terms, a dihedral term taking into account the periodicity of dihedral conformations, and improper torsions also modeled as harmonic terms to account for planar connectivities.

$$U_{bonded} = \sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\phi(1+cos(n\phi-\delta)) + \sum_{impropers} k_\varphi(\varphi-\varphi_0)^2, \tag{3.54}$$

where, b and b_0 are bonds and equilibrium bond lengths, θ and θ_0 are angles and equilibrium angles, ϕ, n, δ are dihedrals, periodicity and equilibrium shifts in dihedral, and φ and φ_0 and impropers and equilibrium impropers. The non-bonded interactions consist the electrostatics term given by a Coulombic potential and van der Waals (vdW) term modeled as a Lennard-Jones (LJ) potential. These terms are evaluated pairwise for all atoms in the system,

$$U_{non-bonded} = \sum_{ij} \frac{q_i q_j}{4\pi\epsilon r_{ij}} + \epsilon_{ij}\left[\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{6}\right], \tag{3.55}$$

where q is the charge of atoms, ϵ is the dielectric constant of the medium, ϵ_{ij} is the well depth of the LJ potential for atoms i and j, σ_{ij} is the inter-atomic distance between i and j corresponding to a minimum in the LJ potential.

Due to the pairwise nature of non-bonded interactions, their evaluation is a computational bottleneck to the MD algorithm ($\binom{N}{2} \approx N^2$ evaluations for N particles). Cutoff schemes can be used where, the weak vdW term is smoothly truncated at a cutoff distance, whereas, for the electrostatics term, which has a persistent r^{-2} dependence, the interactions beyond a cutoff distance are evaluated only every few steps periodically. However, this approach may lead to errors and artifacts in the simulation [228]. This can be avoided by evaluating the long range electrostatic term using Ewald summation techniques [229].

3.2.3 Periodic boundary conditions and PME

Biomolecular MD simulations are usually carried out with periodic boundary conditions (PBC) to avoid unphysical boundary effects and finite size of the MD setup. In PBCs, the MD setup is replicated in all directions making it a pseudo-infinite setup. When evaluating the potential energy function, each atom interacts with the nearest atom or image in the periodic lattice according to the minimum image convention. The error due to cutoff schemes for electrostatics can be avoided when using PBC and evaluating all electrostatic interactions using Ewald summation schemes for periodic systems. Ewald summation [230] of long-range electrostatics can be written as,

$$U_{\text{Ewald}} = U_{\text{direct}} + U_{\text{reciprocal}} + U_{\text{self}} + U_{\text{surface}}, \tag{3.56}$$

where, the terms on the right refer to a direct sum, a reciprocal sum, self-energy and surface energy. The self-energy term is a constant while the surface energy term vanishes due to dielectric screening by water. The direct sum is calculated in real space while the reciprocal sum is calculated in the reciprocal space. Ewald summation scales as $N^{3/2}$. The Particle Mesh Ewald (PME) [229] is a faster algorithm for calculating long-range electrostatics, where, the reciprocal sum in Ewald summation is instead, evaluated using fast Fourier transform (FFT). For a N particle system, PME scales as $N\log(N)$, offering considerable speed-up for large systems.

3.2.4 Time integration

At the heart of the MD simulation technique is the numerical evaluation of Newton's equations of motion for a N particle system, based on potential function, depending all $3N$ coordinates and produces a $6N$-dimensional coordinate-momentum space (phase space) spanning trajectory describing the time evolution of the system. The differential equation, Eq. 3.49, can be integrated numerically by the Taylor expansion based velocity-verlet algorithm as,

$$\begin{aligned}\vec{x}(t+\Delta t) &= \vec{x}(t)+\vec{v}(t)\Delta t+\frac{1}{2}\vec{a}(t)\Delta t^2\\ \vec{v}(t+\frac{1}{2}\Delta t) &= \vec{v}(t)+\frac{1}{2}\vec{a}(t)\Delta t\\ \vec{v}(t+\Delta t) &= \vec{v}(t+\frac{1}{2}\Delta t)+\frac{1}{2}\vec{a}(t+\Delta t)\Delta t,\end{aligned} \tag{3.57}$$

where $\vec{x}$, $\vec{v}$ and $\vec{a}$ are position, velocity, and acceleration vectors, respectively and Δt is the integration time step. In velocity-verlet algorithm, first, positions at $t+\Delta t$ are evaluated, followed velocity evaluation at $t+\frac{1}{2}\Delta t$. Thereafter, forces are computed for $t+\Delta t$ followed by velocity evaluation at $t+\Delta t$, giving the kinetic energies. The choice of an appropriate time step involves a trade-off between accuracy and speed of the algorithm. The time step must be large enough to sample as much of phase space as possible, but it should also be small enough to generate trajectories that produce time-reversible energy conserving ensembles. The time step of choice must be smaller than the fastest possible vibrations in the system, typically 3000 $\text{cm}^{-1} \approx 10$ fs. Hence, a time step of 1-2 fs is employed for biomolecular simulations. To propagate the system, starting from a static structure, the initial velocities in an MD simulation are assigned from a Maxwell-Boltzmann distribution, for a simulation temperature (often 310 K for biomolecular systems). [223],

Langevin thermostat and barostat

The Langevin equation is a stochastic differential equation, an extension to Newton's equation of motion, where, the system is assumed to be immersed in a reservoir of fictitious particles that are microscopic compared to the macroscopic system of interest. These interactions lead to a modified Newton's equation with additional terms, a random force and a friction term. The friction term is due to the damping induced by the smaller particles and it lowers the velocities compensating for the energy influx due to the random force. The equation can be written as,

$$\vec{\nabla}_i U - \gamma_i \vec{v}_i + \vec{R} = m_i \vec{a}_i, \tag{3.58}$$

where, R_i is the random force and γ_i is the friction coefficient for particle with index i. The random force is usually drawn from a gaussian distribution with mean value of zero and is completely uncorrelated in time. The frictional force is related to the random force by the fluctuation-dissipation theorem as,

$$\langle \vec{R}_i(t)\cdot\vec{R}_i(t')\rangle = 6k_BT\gamma_i. \tag{3.59}$$

This relation can be used to maintain temperatures in constant temperature simulations.

For constant pressure treatment, an alternate form of Langevin equation where an analogy of a fictional piston is used to regulate the pressure by shrinking or expanding the volume of the simulation cell to maintain constant pressure. The equation can be written as,

$$\frac{d^2}{dt^2}V = \frac{1}{W}[P(t) - P_{ext}] - \gamma\frac{d}{dt}V + R(t), \tag{3.60}$$

where, V is the volume, $P(t)$ and P_{ext} the instantaneous pressure and imposed pressure, respectively. W is an additional degree of freedom serving as the pressure piston [231].

3.2.5 Free energy calculations

Computation of free energy is fundamental to understanding biomolecular structure, function and behavior. In thermodynamics, the Helmholtz free energy is defined in the NVT ensemble as,

$$A = -\frac{1}{\beta}\ln(Z) \tag{3.61}$$

where β is the inverse thermodynamic temperature ($(k_BT)^{-1}$) and Z is the partition function in NVT ensemble given by,

$$Z = \sum_i e^{-\beta E_i}, \tag{3.62}$$

where, E is the potential energy of the system. In the NPT ensemble, the Gibbs free energy (G) is obtained. In the condensed phase, however, the systems are usually not very compressible so both free energies are essentially the same.

In biomolecular simulations, we are generally interested in computing the free energy along a given reaction coordinate, ξ, a suitable geometric or chemical degree of freedom along which, the PES is projected. Such reaction coordinates can model structural or chemical changes during a biomolecular process, for example, opening/closing of an ion channel as a function of inter-residue distances in the protein. MD simulations typically spend most of time in minimum energy basins of the PES whereas regions of higher energy are sampled rarely. The knowledge of free energy along the reaction coordinate at such rarely sampled regions is often essential to obtain barriers using biasing methods like umbrella sampling for the biomolecular process of interest (see next subsection).

Umbrella sampling

The umbrella sampling (US) formalism was introduced by Torrie and Valleau in 1977 [232]. In US, a biasing potential, e.g., a harmonic potential is imposed on the system along the reaction coordinate to ensure sampling of unfavorable zones on the free energy surface. The potential modified by addition of windows (usually harmonic heaps) is given by,

$$V^b(r) = V^{ub}(r) + w_i(\xi), \tag{3.63}$$

where, b and ub superscripts indicate biased and unbiased potentials, respectively, and w_i is the added biasing potential for the harmonic window, i. The unbiased free energy as a function of reaction coordinate can be recovered as,

$$A_i(\xi) = -\frac{1}{\beta}\ln P_i^b(\xi) - w_i(\xi) + F_i \tag{3.64}$$

where P_i^b is the Boltzmann average with biased potentials, and F_i is the free energy shift associated with introduction of the bias for the window i and is given by

$$F_i = -\frac{1}{\beta}\ln\langle\exp[-\beta w_i(\xi)]\rangle \tag{3.65}$$

The F_is are initially unknown but can be evaluated by numerical procedures like Weighted Histogram Analysis Method (WHAM) [233]. The WHAM equations can be written as,

$$p(\xi) = \frac{\sum_{j=1}^{N} n_j p_j^{biased}(\xi)}{\sum_{j=1}^{N} n_j \exp[-\beta(V_j(\xi) - F_j)]}, \tag{3.66}$$

and,

$$\exp[-\beta F_i] = \sum \exp[-\beta V_i(\xi)]p(\xi)d\xi. \tag{3.67}$$

The WHAM equations can be solved self-consistently to obtain the unbiased probability distribution.

3.2.6 Continuum electrostatics

The function of many proteins depend on the protonation states of their amino acids. For example, a change in the protonation state of a residue in an ion channel may trigger a myriad of conformational changes leading to open or closing of the channel or some other biochemical function. Additionally a protonation change involves a change in long-range electrostatic interactions between all redox groups, ions, and other titrable residue within the protein. Many experimental and computational studies over the years have reported highly up- or down-shifted pK_a values of amino acids [112, 234–237]. Thus, calculation of pK_as of amino acids in proteins is central for understanding catalytic processes [238].
For an acid HA,

$$\mathrm{HA} \overset{\mathrm{K_a}}{\rightleftharpoons} \mathrm{H^+} + \mathrm{A^-},$$

with the dissociation constant, K_a,

$$K_a = \frac{[\mathrm{H^+}][\mathrm{A^-}]}{[\mathrm{HA}]},$$

the pK_a for HA can be written as,

$$\mathrm{p}K_a = -\log_{10} K_a = \frac{\Delta_a G}{2.303RT} \tag{3.68}$$

where $\Delta_a G$ is the free energy of deprotonation.

In the continuum electrostatics method, water is treated implicitly, with a dielectric constant of 80, corresponding to bulk water, and the interior of the protein is usually modeled with a dielectric constant of 4.

Th pK_a of a titrable amino acid can be calculated using the approach of thermodynamic cycle (Fig 3.1). The pK_a shift is evaluated as the free energy of transferring a titrable amino acid from solvent (water) to its position in the protein. The pK_a value of an amino acid in water is called its model pK_a value, which is generally obtained from experimental data. The change in pK_a value of the amino acid in the protein is calculated by adding the Born desolvation energy, ΔG_{Born}, the interaction energy due to background charges, ΔG_{back}, with all the other amino acids in the protein in their neutral (reference) states, and a term, ΔG_{ij}, corresponding to the electrostatic interaction between titrable amino acid pairs as protonation state of a residue depends on the protonation state of all other

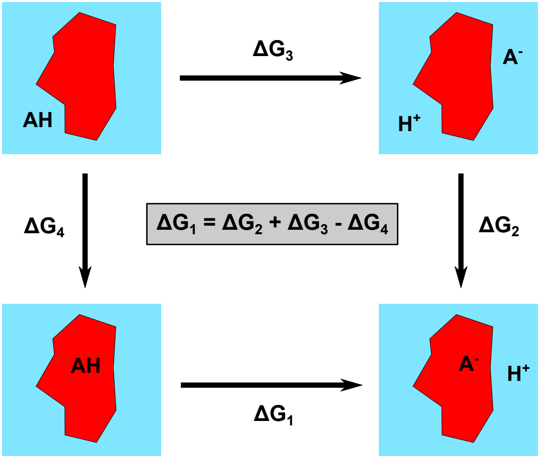

Figure 3.1: Thermodynamic cycle used for calculating pK_a of residues in protein environment. Red indicates the protein medium ($\epsilon = 4$) and cyan indicates aqueous phase ($\epsilon = 80$). $\Delta \mathrm{G}_1$ gives the pK_a of the amino acid, HA, in the protein medium. $\Delta \mathrm{G}_1$ can be evaluated in terms of $\Delta \mathrm{G}_3$, which is known from experiments, and, $\Delta \mathrm{G}_2$ and $\Delta \mathrm{G}_4$, which are evaluated computationally.

protonable residues in the protein. The total free energy is thus,

$$\Delta G = \Delta G_{Born} + \Delta G_{back} + \Delta G_{ij}. \tag{3.69}$$

The ΔG_{ij} term involves a Boltzmann sum over all pairs of protonable residues. Even for a moderately sized protein, the calculation of interaction term is computationally very demanding as it involves at least 2^N contributions for a protein with N titrable residues. Usually Monte-Carlo procedures are used to effectively estimate the ΔG_{ij} term.

The electrostatic terms are evaluated by solving for linearized Poisson-Boltzmann equation (LPBE) on a discrete grid,

$$\nabla[\epsilon(\vec{r})\nabla\phi(\vec{r})] - \kappa^2(\vec{r})\phi(\vec{r}) = 4\pi\rho_0(\vec{r}), \tag{3.70}$$

where, $\phi(\vec{r})$ is the electrostatic potential, $\epsilon(\vec{r})$ is the dielectric constant, $\rho_0(\vec{r})$ is the charge density and $\kappa(\vec{r})$ is the inverse Debye length. The electrostatic potential of the system is calculated iteratively based on LPBE until convergence and the electrostatic energy can then be evaluated using the charges of each of the titrable residue in the protein. [235, 239, 240]

Continuum electrostatics pK_a calculation can be used for predicting pK_a shifts of buried residues in a protein. However, several approximations, like implicit treatment of solvent, assigning a dielectric constant of 4 to protein and 80 to solvent often lead to large errors and thus the results should be interpreted in a qualitative manner [241, 242]. Benchmarking studies, however, suggest that accuracies of *ca.* 1 pK_a can be achieved [243].

3.3 QM/MM calculations

Quantum mechanical methods are necessary to study chemical reactions as they produce accurate description of the structure and energetics of the system. Moreover, they are almost parameter free in principle, i.e. they do not need input parameters from force fields. Enzymes are biological catalysts, and enzymatic reactions often involve breaking and forming of bonds and thus require a QM treatment. The protein environment plays a key role in facilitating the reactions at the active site, often due to long-range electrostatics. In practice, QM methods can not be used for systems more than *ca.* 200-300 atoms in size. This makes accurately studying enzymatic reactions with QM methods difficult. MM methods, on the other hand, can be used to simulate the whole enzyme in its native-like environment but are based on parametrized force fields. This renders them inflexible for enzymatic reactions, which involve bond-forming and bond-breaking along with subtle energetic changes which are largely inaccessible to MM methodology.

Quantum Mechanics / Molecular Mechanics (QM/MM) calculations strive to overcome these limitations by treating the region of chemical interest, for example the active site of the enzyme at QM level, while the rest of the system is described at molecular mechanics level. The QM/MM methodology was first introduced by Warshel and Levitt in 1976 [244] for which they were awarded the Nobel prize in 2013. Over, the past few years QM/MM methodology has become increasingly popular for studying biochemical reactions and has yielded many insights into the inner workings of enzymatic machineries [245–247]. This section is devoted the QM/MM methodology applied to biomolecular systems.

QM/MM energy schemes

The energy term in QM/MM comprises three types of interactions terms: (1) QM-QM energy term, (2) MM-MM energies and (3) QM-MM energy describing the interactions between QM and MM regions. While QM-QM and MM-MM terms are fairly easy to describe using QM and MM algorithms, respectively, the description of QM-MM coupling term is rather challenging. Two schemes, namely the subtractive and additive schemes, are used most frequently in QM/MM calculations to tackle this problem.

In the subtractive scheme, the energy of the full system is evaluated at MM level. Thereafter, the energy of the QM system is calculated at the QM level and is added to the MM term. To account for double-counting, the energy of QM region is evaluated at the MM level and is subtracted from the energy term,

$$V_{QM/MM} = V_{MM}(MM + QM) + V_{QM}(QM) - V_{MM}(QM), \tag{3.71}$$

where, the terms in parenthesis indicate the region of evaluation and subscripts indicate the method of evaluation. The subtractive QM/MM scheme is easy to implement and is computationally efficient [248]. However, the simplistic approximation leads to several shortcomings. The QM-MM interaction is effectively treated at the less accurate and force field based MM level. Moreover, the QM region needs to be parametrized to obtain energy at the MM level.

The additive scheme involves the sum of MM region energies at MM level, QM region energies at QM level and the third coupling term describing the interaction of MM and QM (QM-MM) regions,

$$V_{QM/MM} = V_{MM}(MM) + V_{QM}(QM) + V(QM - MM). \tag{3.72}$$

The $V(QM - MM)$ term accounts for vdW, electrostatics and bonded interactions at the boundary between QM and MM region.

$$V(QM-MM) = V^t extvdW(QM-MM) + V^{elec}(QM-MM) + V^{bond}(QM-MM) \tag{3.73}$$

The vdW interaction term $V^{vdW}(QM-MM)$ is given by the LJ potential,

$$V^{vdW}(QM-MM)=\sum_{QM}\sum_{MM}\epsilon_{ij}\left[\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{12}-2\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{6}\right], \tag{3.74}$$

where, i and j are indices of atoms in QM and MM regions, respectively, and LJ parameters are as described previously [245].

The electrostatic interaction term $V^{elec}(QM-MM)$ can be described by three different approaches: mechanical embedding, electrostatic embedding or polarized embedding schemes. In mechanical embedding, the MM region interacts with a set of fixed partial charges for the atoms in the QM system and is the same as subtractive scheme and is usually avoided.

Electrostatic embedding involves including the MM fixed point charges in the QM calculations, which are used to evaluate electrostatic QM-MM interaction as based on pairwise point charges (similar to LJ treatment above). The electrostatic embedding scheme is the most popular treatment in QM/MM simulation of enzymes due to right balance of speed and accuracy.

In polarized embedding, both the MM region and the QM region polarize each other. Theoretically, this is the most accurate description of the QM-MM interaction term, but polarized embedding schemes are rarely used for QM/MM calculation as polarizable force fields for MM regions are still in their developmental phase and hence often impractical to use [245–247, 249].

QM/MM boundary

Partitioning the simulation system into QM and MM regions is generally not trivial. It depends largely on chemical intuition and the level of QM theory needed to describe the biochemical reaction studied in the QM region.

Proteins are long chains of covalently bonded amino acids. During selection of the QM region, the strategy is to have interesting fragments of protein in the QM region, while the remaining protein fragments are placed in the MM region. An unavoidable consequence is that covalent bonds are broken at the boundary and need to be saturated in respective regions to prevent unphysical behavior. Several methods are implemented in QM/MM methodology for boundary treatment.

The link-atom scheme [250] introduces a fictional link-atom, typically a hydrogen atom, to saturate the QM atoms where covalent bonds are broken. The link atom is placed along the broken bond at a distance similar to covalently bonded hydrogen. The forces on the link-atom in the simulation are redistributed to the MM and QM atoms, whose bond was cut at the boundary. Link-atom scheme is a frequently used boundary scheme in QM/MM implementations.

In the boundary-atom scheme, the first atom of the MM region next to the boundary is included in the QM calculation to emulate a saturated bond in the QM calculation, whereas in the MM calculation, it acts as a MM atom. The boundary-atom needs to be parametrized in order to compensate for broken covalent bond. The need of parameters makes this approach less attractive [245].

In the frozen-orbital scheme [251], hybrid orbitals are placed at the boundary where a covalent bond is cut between MM and QM. The orbitals emulate a saturated bond and are frozen during the calculations. Though theoretically elegant, this scheme is less popular as it is more computationally intensive, more difficult to implement, and practically gives similar results as the link-atom scheme [245, 247, 249].

QM/MM-MD simulations

QM/MM calculations at the DFT/MM level of QM theory are computationally very intensive, but with the increasing computing power and more efficient simulation algorithms it is possible to perform QM/MM-MD simulations for enzymes. The QM/MM-MD treatment is based on BOA, where, in each time step, the energy and the gradients are computed for the QM region with positions of nuclei fixed. Once the electronic part has been evaluated, and the gradient and energy terms from the QM region and added to the MM region, the system is propogated by Newton's laws of motion (see Eq. 3.49). The time-evolution of the electronic structure, which governs the chemistry, depend on the nuclear point-charge coordinates of the MM region, which are updated at every step of QM/MM-MD algorithm. A typical time step of 1 fs is usually employed in QM/MM-MD simulations for algorithmic stability and to capture all vibrational frequencies of the QM system. Evaluation of the gradient term for the QM region with all MM atoms as background charges, is often the computational bottleneck in the QM/MM calculations. One way to circumvent this problem is to discard atoms beyond a certain threshold radius from QM region. In such cases it should be noted that new system should be neutralized, which can be difficult to achieve in practice. In this work, complete set of atoms from the MD setups have been used for QM/MM simulations. Other QM/MM-MD approaches include Car-Parrinello (CP) MD (CPMD) simulations which are based on DFT pseudo potential plane wave framework. [252, 253] Although, the approach is computationally rapid, it requires shorter integration time steps in order to account for electronic motion that must average out during the vibrations.

4 Aims of the study

The proton pumping mechanism of the redox driven proton pump, C*c*O, has been subjected to decades of rigorous experimental and theoretical studies [22, 23, 37, 47, 59, 82, 95, 98, 99, 116, 145–147, 254, 255]. There is a consensus on many key mechanistic features, e.g., the existence of a redox-sensitive transient proton-loading site (PLS) [96, 98], the crucial role of internal water molecules in proton transfer processes [22, 82, 256], and the roles of the D- and the K-channels [22, 94, 116]. However, several aspects, such as, the exact location of the PLS, the order of proton and electron transfer events and the mechanism of redox-linked proton pumping, nonetheless, remain unclear. This work aims to provide molecular insight into some of the aforementioned unclear aspects of the proton pumping machinery of C*c*O using multi-scale molecular simulations.

The key questions this work aims to answer are the following:

1. How does C*c*O regulate the sorting and delivery of protons to different sites, *i.e.*, to the BNC and to the PLS, depending on the redox state of the enzyme? What prevents the backflow of these transferred proton towards their thermodynamically favorable directions?

2. What is the character and location of the PLS? What is the basis of loading and release of a proton at the PLS? What is the mechanism of the proton transfer? What prevents the loaded protons from leaking back?

3. Why does the K-channel only supply protons in the latter half of the catalytic cycle and what prevents the proton transfer from K-channel in the first half of catalysis? What is the mechanism and energetics of the proton transfer reactions through the K-channel?

5 Redox-state dependent water dynamics

[†]

5.1 Introduction

Water molecules influence the structure, dynamics, and function of biomolecules in a multitude of ways [257–259]. Water molecules in the solvation shell of proteins, the so-called "biological water" [260], modulate the secondary structure and dynamics of the protein [261], whereas, intra-protein water molecules interact with the protein at specific hydrophilic residues through hydrogen bonds, and are central to proton transfer processes [256, 262, 263]. In proton pumps, such functional water molecules provide protonic connectivity to transfer protons along hydrophilic channels by Grotthuss mechanism [89]. In a Grotthuss-like proton transfer reaction, an “excess” proton moves through a hydrogen bonded network of proton carrier molecules by concomitantly forming and breaking bonds of adjacent hydrogen bonded molecules. The reaction proceeds by forming a covalent bond with an incoming proton while transferring a covalently bonded proton to the next carrier and so on until final acceptor species is protonated. This leads to diffusion of the charge, rather than the movement of the proton itself [92, 93]. For proton transfer in proteins, these proton carriers constitute of intra-protein water molecules and protonable residues. Study of water dynamics is thus important for gaining insights into proton transfer processes in proteins.

5.1.1 Proton transfer in C*c*O

In the oxidative half of the catalytic cycle of C*c*O, the D-channel starting at Asp-91 and terminating at Glu-242, supplies the protons needed for oxygen reduction chemistry, the chemical protons; and the protons that are pumped across the membrane, the pumped protons [97, 101, 102]. The protonable Glu-242 residue acts as a fork, relaying protons to both sites depending on the redox-state of the enzyme [21, 99, 106, 146]. Reduction of the BNC couples to transfer of a chemical proton to the BNC and pumped proton across the biological membrane [22, 82, 96–98, 100–105]. Interestingly, several experiments suggest that the pumped proton is transiently stored a proton-loading site (PLS), before it is ejected across the membrane following reduction and protonation of the BNC [98, 103].

5.1.2 Catalytic water

The X-ray structures of C*c*O [50] lack catalytically active water molecules in the hydrophobic cavity above Glu-242 leading the BNC and to the putative proton-loading site (PLS), which is assumed to reside close to the D-propionate of heme a_3 [138, 142]. Intra-protein water molecules have been shown to play a vital role in the proton-pumping machinery of light-driven proton pump, bacteriorhodopsin [256], and in other bioenergetic proteins [257, 264]. Also, several computational studies have suggested water molecules to be present in

---

[†] Part of this chapter will be published in the manuscript (in preparation) titled "Activation of the K-channel in Cytochrome *c* Oxidase" (see Publications)

the hydrophobic cavity above Glu-242 [146, 265, 266]. Since, Glu-242 is located *ca.* 10 Å away from the PLS and BNC, it is likely that water molecules are present in the hydrophobic cavity in order to catalyze the proton transfer reactions in the active site. The hydrophobic nature of the cavity makes water molecules align in quasi one-dimensional hydrogen bonded arrays [267–269], which might provide the wiring requisite to transfer protons to different catalytic sites of the enzyme. Although it is known the Glu-242 supplies both pumped and chemical protons in the oxidative phase, the exact mechanism of such bifurcation of the proton transfer paths remains unresolved.

5.1.3 Water-gated proton transfer

The notion of redox-state dependent water gating was first suggested by Wikström *et al.* in 2003 [146]. The authors argued that water molecules could, by their orientations direct proton transfer to different catalytic sites in the enzyme. These water-wires would thus distinguish protons originating from the singular source, the D-channel, into pumped and chemical protons. However, the simulations performed in this work were performed on ps timescales and were based on a minimal model of the subunit-I of C*c*O (rhodobacter sphaeroides; PDB ID: 1M56) and without the membrane-water surroundings. This led to a series of reevaluations of the proposed water-gated mechanism of proton delivery in C*c*O. Several groups have suggested that such water arrangements may not contribute in gating of protons in C*c*O and are likely an artifact of the limited simulation setups employed in the original study [155, 160, 161]. In a recent MD study of C*c*O in several redox states, the probability of redox state dependent formation of water-wires based on hydrogen bonding criteria was evaluated [270]. They reported the validity of the water-gated mechanism of proton delivery dependent on the redox states of the enzyme, but suggested that such a water-gating might not be sufficient to prevent the back-flow of protons, leading to a short-circuit of the proton pumping machinery. Moreover, a similar water-gated mechanism was proposed for aquaporin [271, 272], but was later proven to be inadequate [273].

Biological proton transfer processes are governed by two factors: (1) the thermodynamic control, which can be given by proton affinities of the proton donor and the proton acceptor species, (2) the kinetic control, by modulating barriers of proton transfer. In case of C*c*O, the catalytic water molecules in the hydrophobic cavity above Glu-242 may play an important role in the sorting mechanism by reorientating to alter connectivity from Glu-242, selecting for the proton-transfer, either to the BNC or to the PLS.

In the presence of a proton donor, a proton acceptor and a strong electric field in the direction of the proton acceptor may result in tightly hydrogen bonded water-wires ideal for proton transfer (Fig. 5.1A). However, if the fields are weaker, or the proton-affinity of the acceptor group is relatively low, configurations with water molecules not best oriented for proton transfer can occur (Fig 5.1B). Such metastable water configurations can however reorient into water-wire-like configurations. In presence of a field in the opposite direction, water molecules would orient in the opposite direction (Fig 5.1C) blocking the proton transfer by Grotthuss mechanism characteristic to intra-protein proton transfer processes. Such "forbidden" pathways could, if stable enough, act as gating elements preventing proton transfer along a direction depending on electrostatics in the active site of the enzyme [146, 274].

C*c*O has multiple redox cofactors, which are bound to exert a strong electrostatic influence on their environment. Due to inverse squared dependence of electric field on the distance, the field effects are likely to be felt even at appreciable distances. Water

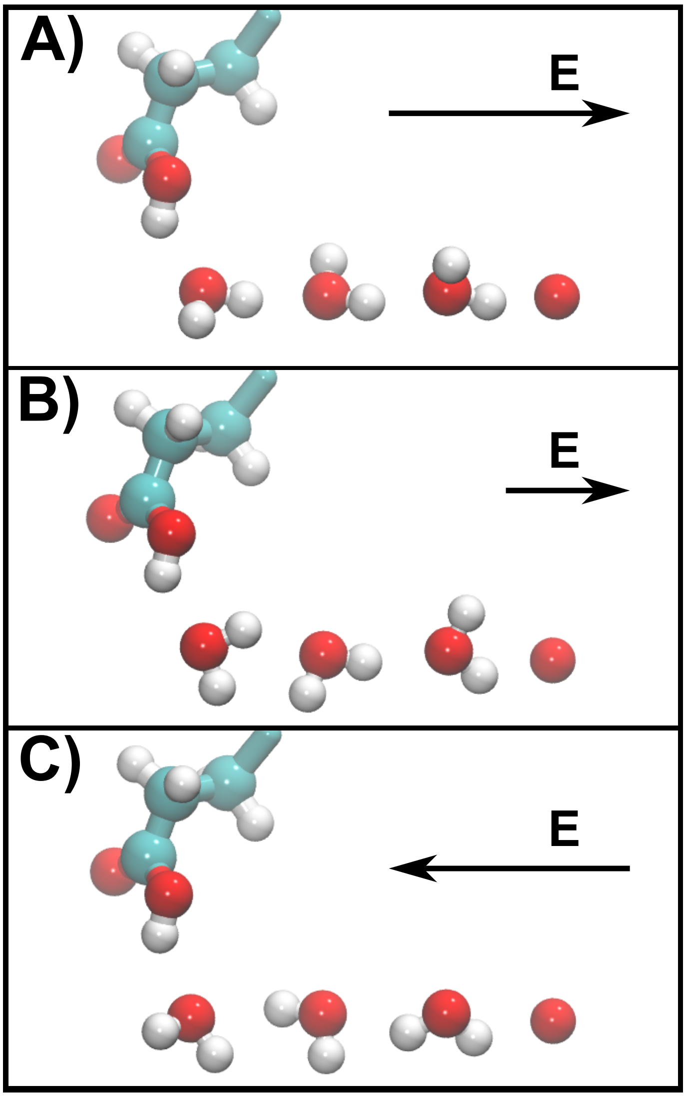

Figure 5.1: Putative structures of water wires to transfer a proton from a protonated glutamic acid residue to a proton accepting site (red sphere), depending on the electric field felt by the water molecules.

molecules, though neutral, are polar, and hence effectively act as dipoles in response to electric fields. In presence of an external electric field, a dipole feels a torque aligning the dipole in a direction anti-parallel to the field. Hydrogen bonds formed between these water molecules stabilize the configuration further by *ca.* 0.25 eV for every hydrogen bond [275]. Moreover, the electric field would further polarize the oriented hydrogen bonded water-wire configuration adding to its stability. The alteration of electric field in the hydrophobic cavity as the electron moves from heme *a* to BNC might be responsible for selecting pumping or chemical pathways, based on the location of the electron, for the protons to be transferred from Glu-242. The orientation of the water molecules in the hydrophobic cavity, at least in principle, may also prevent the leak of protons from the protonated sites back to Glu-242 [146].

5.2 Materials and methods

5.2.1 MD simulations

The X-ray structure of bovine C*c*O was obtained from the Brookhaven Protein Databank (PDB ID: 1V54 [50]). Subunits I and II of C*c*O was embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane, TIP3P water molecules, and a 100 mM NaCl concentration (29 Na^+, 10 Cl^-), comprising in total 88 837 atoms (Fig. 5.2). Residues Glu-242 and Asp-364 of subunit I were modeled in their protonated states, with Glu-242 flipped "up" towards the non-polar active site, and 4-6 water molecules added the non-polar cavity above Glu-242, according to result from previous studies [99, 146, 265, 266, 270]. The pumping process was studied with heme *a* modeled in the oxidized or reduced states, and the BNC in the P_M ($a_3{}^{IV}$=O^{2-}/Cu^{II}-OH^- TyrO$^\bullet$) and P_R (a_{3IV}=O^{2-}/Cu^{II}-OH^- TyrO$^-$) states, based on DFT parameterization of the redox-active groups [276]. The MD setups were subjected to energy minimization following gradual heating of the systems to 310 K. The systems were equilibrated in isobaric *NPT* ensemble for 10 ns. Finally, 500 ns MD simulations were performed in *NVT* ensemble on all studied catalytic states using the CHARMM36 force field [227], a 2 fs integration step, and long-range electrostatics treated with the Particle Mesh Ewald approach [229]. The MD simulations were performed with NAMD 2.9 [277] and the MD trajectories were analyzed using VMD [278].

5.2.2 Longest connectivity along shortest path (ζ)

The evaluation of longest hydrogen bonded connectivity along the shortest path (ζ) was based on single-source single-sink Dijkstra's algorithm by E. W. Dijkstra in 1959 [279]. The algorithm has its roots in graph theory and aims to find the shortest path (sum of edge lengths) between nodes of a graph, given a source and a sink node (Fig. 5.3). From snapshots at every 2 ps obtained from MD simulations, oxygen atoms of water molecules in the hydrophobic cavity above Glu-242, which might be involved proton transfer were considered as intermediary nodes. Glu-242 was modeled as the source node and either, the oxygen of Cu_B ligand or, the carboxylate oxygens of the D-propionate of heme a_3 were considered as sink nodes. Once the shortest connecting graph was evaluated by the algorithm, the longest connectivity (edge) was evaluated by reverting back to the full atomic description (including hydrogen atoms) to obtain the longest hydrogen like connectivity.

The graph to be subjected to Dijkstra's algorithm is initialized by selecting water molecules in the ellipsoid volume (spindle-like) between the source and sink atoms. The graph is constructed in a dictionary like data structure indexed by nodes connected by an edge and the edge length as a value for the indexed dictionary entry. Entries with edge lengths longer than distance between source and sink nodes are discarded and not added to the graph.

For connected nodes *s* (source node) and *u* (intermediary node) in the graph, the algorithm at each step, finds the node, *u*, with the minimum path length from *s* and updates the shortest path lengths of the nodes connected to *u* in the graph, employing a tree-like search pattern iteratively. The procedure follows in a way, such that, given the shortest path length from *s* to *u*, and *u* is connected to another node *v*, one can then update the shortest path length from *s* to *v*. If the path length obtained from the connected path through *u* is smaller than the current record of the shortest length to *v*. The algorithm proceeds breadth-wise, traversing the graph, while only considering the immediate neighbors

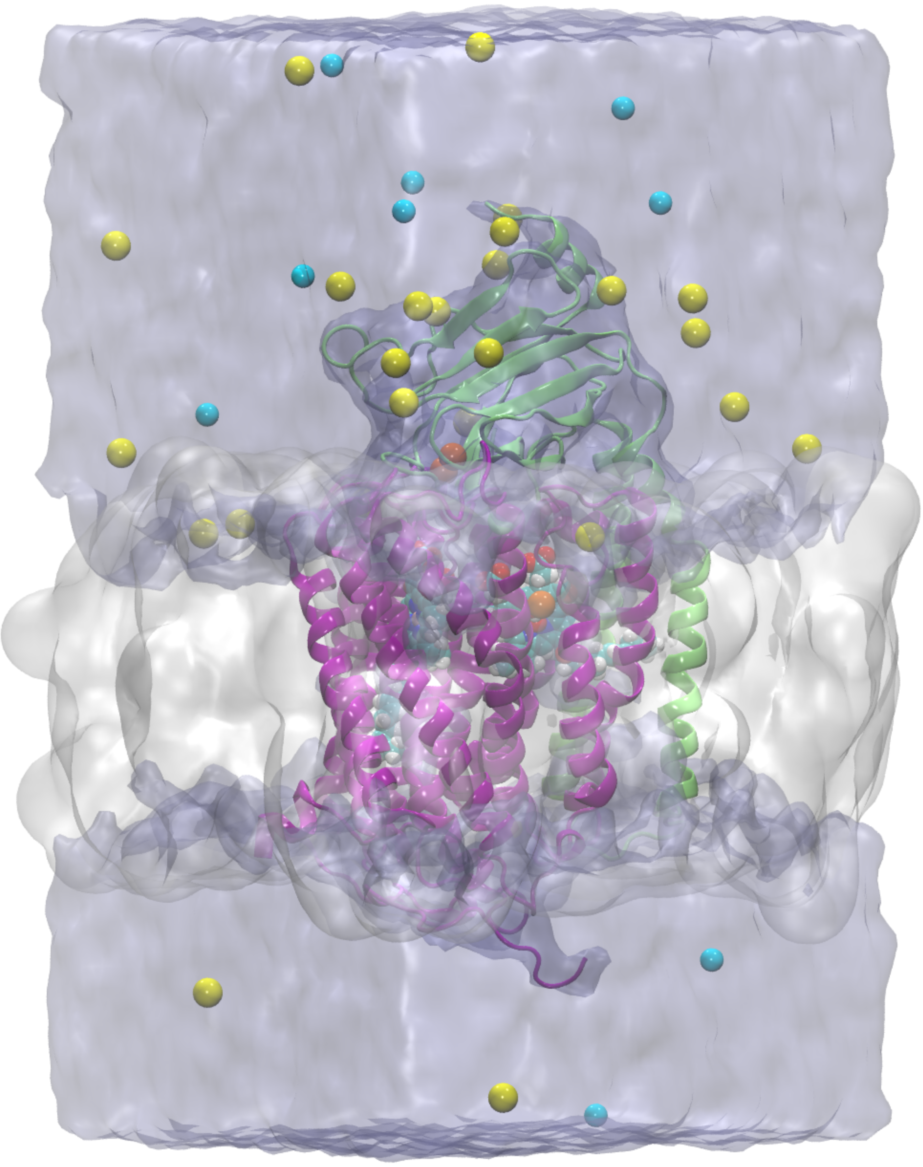

Figure 5.2: Molecular dynamics setup of cytochrome *c* oxidase (PDB ID: 1V54) embedded in lipid membrane (white medium) and solvated by water molecules (blue medium), neutralized with sodium (yellow) and chlorine (cyan) ions at 100 mM concentration. The total system consisted of 88 837 atoms.

at each step without making a conscious effort to reach the source directly. A bookkeeping array, A, which is indexed by nodes, stores the current estimate of the shortest path length from node u to every neighbor node v. The array is a reduced form of the graph and stores the list of neighbor nodes for each node. Once all the neighbors of a node are visited, the node is added to a list of visited nodes, for which A need not be evaluated anymore. A list of parent nodes for each node is also maintained enabling walking back once the sink node is visited to evaluate the shortest path.

At the start of the algorithm, all path lengths are initialized to infinity to signify that these nodes have not been visited by the algorithm. The path length from the source node s to itself is set to 0. The shortest path is evaluated in a recursive fashion which calls the algorithm with new source as the shortest path length node from the previous call.

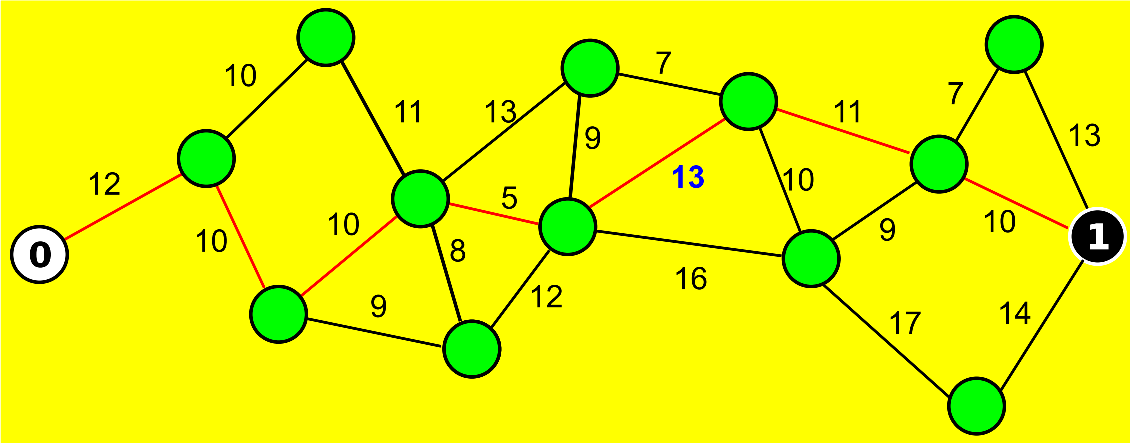

Figure 5.3: Longest connectivity along the shortest path. Example demonstrating shortest path determined by the Dijkstra's algorithm for finding the shortest path (in red) from source (0) to destination (1) traversing the intermediary nodes (green) on a graph. The longest edge (blue label, 13) along the shortest path (red) is the quantity of interest.

The recursion ends when the new source and the sink nodes are the same nodes and the shortest path is evaluated by walking back along the array, A.

Direct implementation of Dijkstra's algorithm to structures from MD can lead to several complications like, (1) graphs not leading to sink nodes causing the algorithm to fail; (2) cyclical referencing of predecessors due to spatial arrangement of three or more nodes leading to infinite recursion; (3) lack of water molecules owing to the selection criterion used for making the graph. Care must be taken before interpreting results directly as there is a possibility that the evaluated shortest path, and thus the longest edge might be unphysical. In the current implementation, checks were placed to identify of the aforementioned inconsistencies during the runtime. Upon encountering the exceptions, the graph was remade with relaxed selection criterion for water molecules expanding the graph that is fed to the algorithm. This procedure was repeated until consistent solutions were obtained.

5.2.3 Snippet averaging analysis

Mechanistic assertions based on MD simulations directly depend on the quality of sampling of the equilibrium ensemble that simulations have been subjected to. Absolute global statistical convergence of the whole ensemble is generally not possible to achieve. However, as long as the simulation leads to canonical sampling, the statistical quality of the observables should increase with longer simulation times. Moreover, protein systems exhibit a multitude of timescales depending on the nature of the process; for example, small backbone movements in a protein backbone can be classified as a fast processes (ps-μs), whereas, processes like amyloid-β peptide aggregation occur at notoriously slow timescales (ms-days). Hence, clearly, some statistical observables (fast) are bound to converge much faster than others, even in the case when the two observables are obtained from the same biomolecular simulation [280].

Historically, time autocorrelation analysis have been the mainstay for error estimation and convergence analysis of statistical observables. The relaxation lifetime estimate from time correlation has an intuitive explanation, that of decorrelation time, implying the loss of memory of its value beyond the relaxation time in history, and hence reaching statistical independence or convergence beyond the relaxation lifetime [217, 223]. Time autocorrelation analysis is a power tool for analyzing MD simulations, but lacks a description for assessing statistical error associated with the observable quantity. Moreover, the involvement of processes with slower timescales in the tail region make the relaxation times harder to estimate [258].

An alternative to time autocorrelation analysis is snippet averaging analysis [281, 282]. In snippet averaging analysis, the MD trajectory is chopped into N contiguous snippets. The mean of the statistical quantity of interest is calculated for each snippet yielding N values. The analysis is started with small snippet widths, w (for e.g., the timestep employed in MD simulations). The snippet width w is incrementally increased and the mean for each snippet is calculated and stored for each such w. For each w, standard deviation of all the means with the snippet width w is evaluated. As the width increases, the snippets get more statistically independent (decorrelated) from each other, until the error estimate converges, yielding the true measure of the standard error of the observable. The Snippet Standard Error (SSE) is given by,

$$SSE(\eta, \omega) = \sigma_{\omega}/\sqrt{N}, \tag{5.1}$$

where, η is the observable of interest and σ_{ω} is the standard deviation among all snippet means for snippet width ω and N is the number of snippets. The SSE monotonically increases until plateauing to the true error estimate of the observable. The snippet width (time) at the error convergence limit is an upper bound to correlation time as estimated from time autocorrelation analysis. Snippet averaging analysis is computationally more efficient than time autocorrelation analysis and the SSE plateauing time gives us a cheaper means to achieve the same result *viz.* estimate convergence of the dynamical observable of interest.

5.2.4 Electrostatic field calculations

Continuum electrostatics calculations were performed to estimate the electrostatic profiles for transferring a probe charge with a radius of 1.6 Å, modeled as the effective radius of a proton, in the hydrophobic cavity between above Glu-242 either towards the D-propionate of heme a_3 or the ligand of $\mathrm{Cu_B}$. The electric field profiles were obtained by solving the linearized Poisson-Boltzmann (PB) equation in MEAD [235]. For the PB calculations, 19 structures for each redox state were selected from the 500 ns MD trajectories. The protein domain was modeled using explicit atoms and a polarizable medium with ϵ=4. For all MD structures from each redox state, the water molecules in the hydrophobic cavity were removed and the positions of the water molecules forming water-wires were used to generate an interpolated array of spatial points along the pump or the chemistry paths from Glu-242 to the D-propionate group of heme a_3 or to the $\mathrm{OH^-}$ ligand of $\mathrm{Cu_B}$, respectively. The electric field was obtained by evaluating PB electrostatics of translating the probe charge along the generated spatial arrays.

5.3 Results

5.3.1 Water dynamics

In order to examine the role of water molecules in the proton transfer processes from Glu-242 residue, 0.5 μs classical atomistic molecular dynamics (MD) simulations of C*c*O in the redox-states, heme $a_{\mathrm{red}}/\mathrm{P_M}$ and heme $a_{\mathrm{ox}}/\mathrm{P_R}$, were performed to study the water dynamics in the hydrophobic cavity, which might be linked to loading of the PLS and protonation of the reduced BNC. Protonated Glu-242 was modeled flipped-up towards the active site, which in line with several mechanistic and computational studies [99, 148, 149, 283]. The MD simulations suggest that water molecules can provide hydrogen bonded connectivity between Glu-242 and the D-propionate group of heme a_3, which is in close proximity to the PLS; as well as between Glu-242 and the oxygen of hydroxo ligand of the $\mathrm{Cu_B}$, the BNC (Fig. 5.4; top panel). Longest connectivity along the shortest path (ζ) was also calculated from MD simulations for each state to quantify water-mediated connectivity, which is likely to play a key role in proton-transfer processes in the active site of C*c*O. ζ is a measure of the propensity of forming a hydrogen bonded path connecting the proton donor to the proton acceptor, with the path constituting water molecules in the hydrophobic cavity above Glu-242. The shortest connected path is the kinetically favored path, while the longest connectivity within this pathway would represent the rate limiting step in the proton transfer process. Thus, a shorter value of ζ implies the presence of a favorable pathway for proton-transfer processes (as seen in Fig. 5.4 and Fig. 5.5). The reduction of the BNC ($\mathrm{TyrO^{\bullet}} \rightarrow \mathrm{TyrO^{-}}$) strongly increases the probability of forming a hydrogen bonded water wire between Glu-242 and the $\mathrm{Cu_B}$ center, while reducing the probability of a hydrogen bonded connectivity to the D-propionate of heme a_3. (Fig. 5.4; bottom panel).

From the ζ profiles, it is clear that the behavior of the ζ is sensitive to the redox state of the enzyme. With the electron on heme a, a strongly connected, hydrogen bonded water-mediated connectivity from Glu-242 to the PLS is observed, favoring protonation of the PLS. Whereas, when the electron is transferred from heme a to the BNC, the connectivity flips towards the BNC (Fig. 5.4). This behavior is further reflected in the cumulative density of ζ for the simulated redox states (Fig 5.5).

In the MD simulations, we also find that the absence of water molecules in the hydrophobic cavity above Glu-242 leads to flipping of Glu-242 towards the D-channel, away from the catalytic site, which could prevent the proton transfer. The evaluation of water occupancy along the vectorial paths, emanating from Glu-242 and leading to the BNC and the PLS, show similar hydration pattern for pump and chemical paths in heme $a_{\mathrm{red}}/\mathrm{P_M}$ state with the electron on heme a, whereas, there is an increased hydration towards the chemical pathway in the heme $a_{\mathrm{ox}}/\mathrm{P_R}$ state, with the electron at BNC along with reduced water occupancy in the path towards the pump site (Fig. 5.6).

A water mediated hydrogen bonded path linking Glu-242 to PLS or BNC has been characterized by $\zeta < 2.5$ Å, which is within the hydrogen bonding limit. For the distribution of ζ, $\zeta = 2.5$ Å serves a role similar to a node or transition state separating hydrogen bonded paths ($\zeta < 2.5$ Å; below the lower dotted line in the bottom panel of Fig 5.4) from the next class of water arrangements ($2.5 < \zeta < 4.0$ Å; between dotted lines in the bottom panel of Fig 5.4) constituting configurations which, often have a water molecule being in an orientation disrupting the potential hydrogen bonded network along the shortest path leading to larger ζ values. Occasionally, in this ζ range, the hydrophobic cavity has insufficient number of water molecules to form a fully connected path to the proton donor to the acceptor (second peak). Water molecules in hydrophobic regions have characteristically reduced translational entropy and heightened orientational entropy [265]. These altered characteristics can account for the flippant behavior of ζ, with frequent inter-conversions between the two classes (Fig. 5.4), often on fs timescales. The regime beyond $\zeta > 4$ Å (above the higher dotted line in the bottom panel of Fig 5.4) is predominantly, a result of the hydrophobic cavity being devoid of water molecules.
To quantify the water dynamics, the correspondence of ζ regimes depending on water oc-

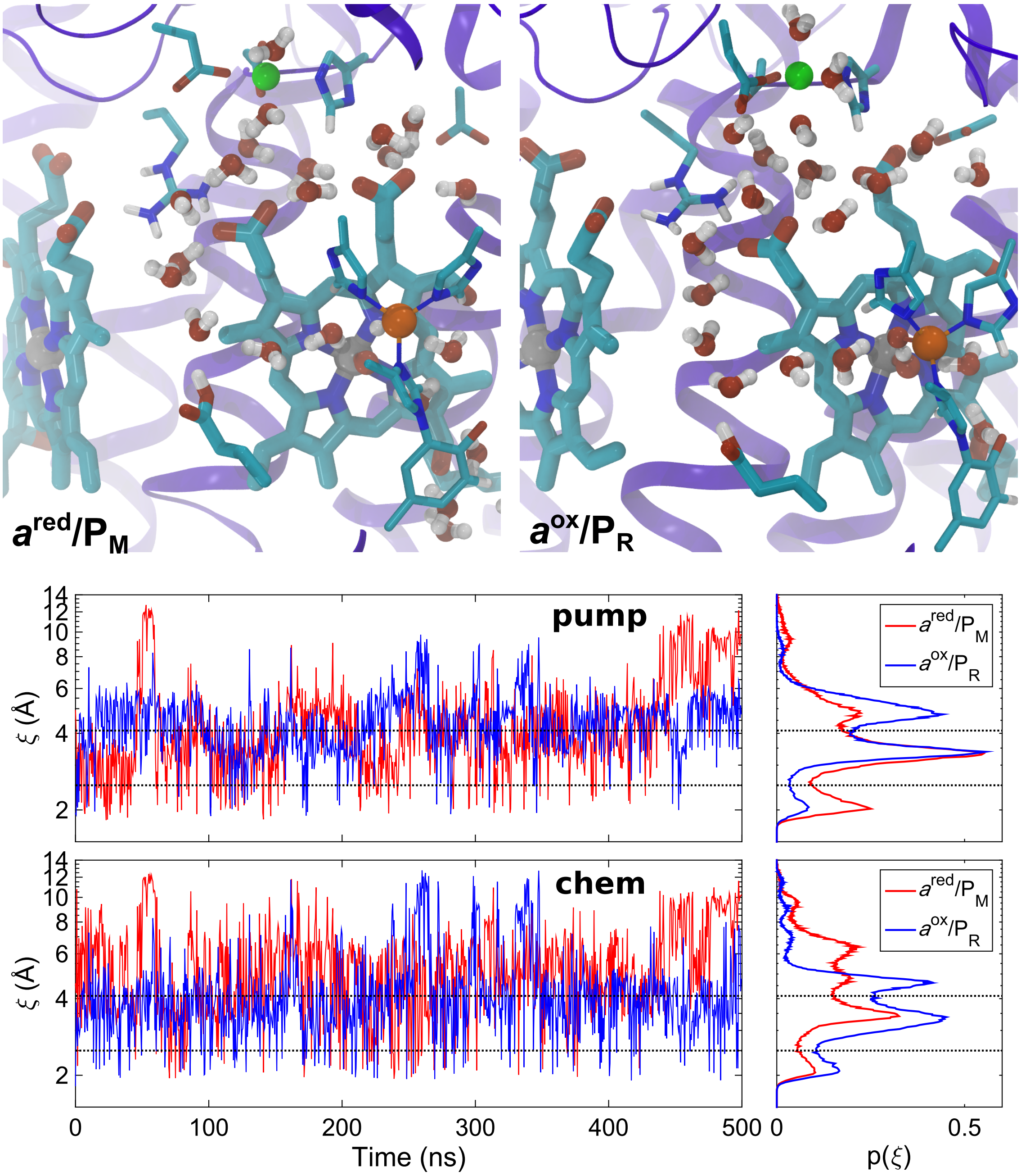

Figure 5.4: Redox-state dependent water dynamics in the active site of C*c*O. Top: Snapshots of the water dynamics obtained from MD simulations of heme $a_{\mathrm{red}}/\mathrm{P_M}$ state on the left, and heme $a_{\mathrm{ox}}/\mathrm{P_R}$ state on the right, resembling transient states involved in proton transfer events. Bottom: Time-evolution and probability distribution of the longest hydrogen bonding distance along the shortest path, ζ, from Glu-242 to the D-propionate of heme a_3 (top sub-panel, ”pump”) and to the hydroxide of $\mathrm{Cu_B}$ (bottom sub-panel, ”chemistry”). The probability distributions were calculated with Dijkstra's algorithm from 500 ns MD simulations of the heme $a_{\mathrm{red}}/\mathrm{P_M}$ and $a_{\mathrm{ox}}/\mathrm{P_R}$ states. The bottom right panel describes the histograms of ζor “pump and “chemical” pathways.

cupancy in respective proton transfer directions was evaluated as conditional probabilities (Fig. 5.7). The conditional probability distributions indicate that regimes with $\zeta < 4.0$ Å show similar correspondence to water occupancy with higher water content leading to higher population of connected paths in both redox-states, heme $a_{\mathrm{red}}/\mathrm{P_M}$ and heme

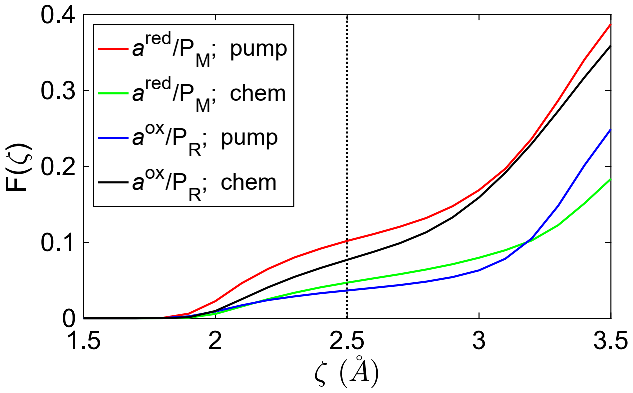

Figure 5.5: Cumulative distribution function of ζ over chemical and pump paths for redox states heme $a_{\mathrm{red}}/\mathrm{P_M}$ and $a_{\mathrm{ox}}/\mathrm{P_R}$.

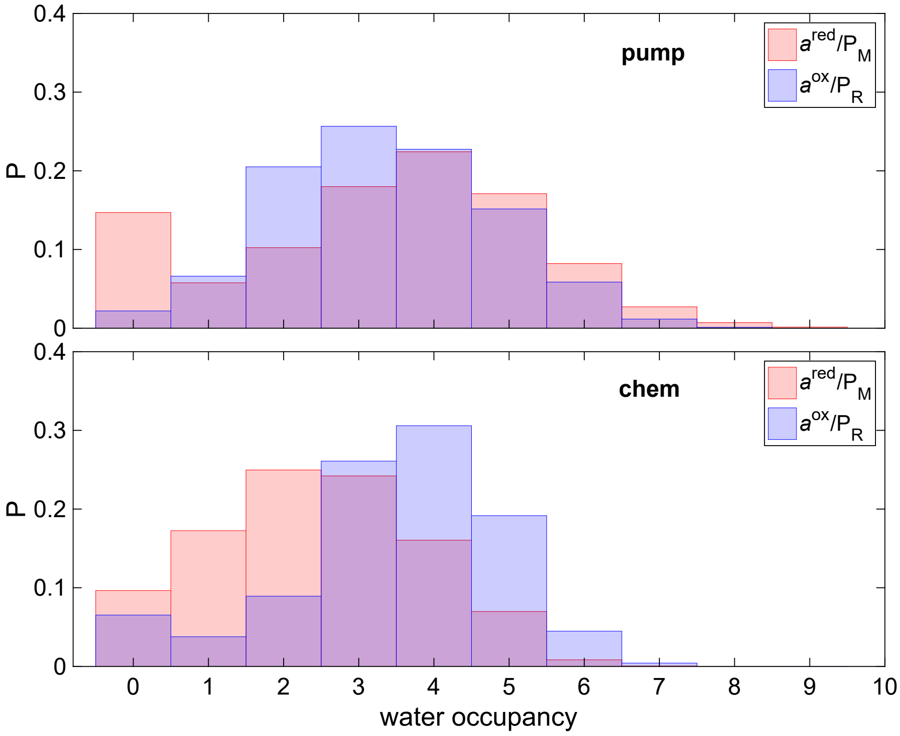

Figure 5.6: Water occupancy along the putative proton transfer paths leading to pump or chemical sites for catalytic states heme $a_{\mathrm{red}}/\mathrm{P_M}$ and heme $a_{\mathrm{ox}}/\mathrm{P_R}$.

$a_{\mathrm{ox}}/\mathrm{P_R}$. Moreover, the similar correspondence between $2.5 < \zeta < 4.0$ Å and $\zeta < 2.5$ Å (hydrogen bonded connectivity) explains the flipping behavior of ζ (as seen in temporal plot, Fig. 5.4) is due to frequent inter-conversions among the two regimes. The correspondence for $\zeta < 4.0$ Å also suggest that these water configurations effectively belong to the same regime, given that inter-conversions occur at ps timescales, which is much faster than the timescales for proton transfers these water molecules catalyze. Thus, a strictly evaluating only hydrogen bonded paths may lead to erroneous inferences. The conditional probability plots are in line with the observations from MD simulations and further suggest that the formation of water-mediated protonic connectivities to be sensitive to the water content at the active site and to the redox states of the enzyme.

In order to evaluate the statistical error associated with ζ, for each redox state along each path, snippet averaging error analysis was performed. The statistical error associated

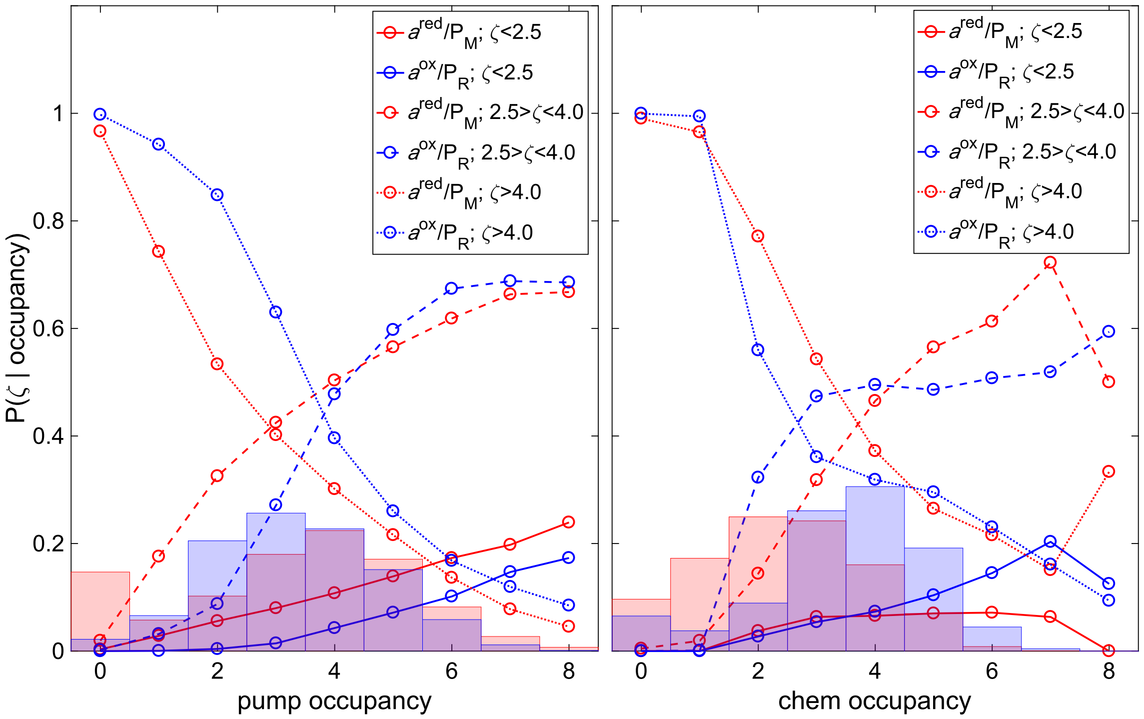

Figure 5.7: Conditional probability of connectivity regimes (ζ ranges) based on water occupancy along the chemical and pump paths for redox-states heme $a_{\mathrm{red}}/\mathrm{P_M}$ and heme $a_{\mathrm{ox}}/\mathrm{P_R}$. The histograms indicate probability of water occupancy along the respective paths and redox states. The solid, dashed and dotted lines represent the conditional probability of ζ regimes ($\zeta < 2.5$ Å; $2.5 < \zeta < 4.0$ Å; $\zeta > 4.0$ Å), given the water occupancy along the respective paths and redox states.

with ζ lies in the range of *ca.* 0.01-0.05 Å (Fig. 5.8). Importantly, the analysis shows error in ζ reached convergence after 75-100 ns. Thus, the 500 ns of MD simulations performed for each state, ensure adequate sampling for the observable ζ. It should also be noted that previous studies examining water-gating of protons in C*c*O have performed much shorter MD simulations ranging from 10-70 ns [270], which are unlikely to capture the complete equilibrium behavior of the water dynamics in the hydrophobic cavity above Glu-242.

5.3.2 Electric field effects

In order to probe the redox-state dependent water dynamics in C*c*O, the energy for transferring a cation with a proton radius (1.4 Å) between Glu-242 and the D-propionate or $\mathrm{Cu_B}$ was estimated using Poisson-Boltzmann (PB) continuum electrostatic calculations. These calculations do not consider the energetics of actual bond breaking, but can, nevertheless, estimate electrostatic contributions of the proton transfer process. The PB calculations indicate that reduction of the BNC increases the exergonicity of transferring a proton along the chemistry pathway, consistent with the increased probability of forming a hydrogen bonded water pathway between the Glu-242 and $\mathrm{Cu_B}$ (Fig. 5.9). We also find that reduction of heme a ($a_{\mathrm{red}}/\mathrm{P_M}$) favors the proton transfer along the chemistry pathway in comparison to the oxidized state ($a_{\mathrm{oc}}/\mathrm{P_M}$), while the transfer along the "pump" pathway is favored in the one-electron reduced states ($a_{\mathrm{red}}/\mathrm{P_M}$,$a_{\mathrm{ox}}/\mathrm{P_R}$) relative to the $a_{\mathrm{ox}}/\mathrm{P_M}$ state. The PB calculations thus suggest that the increased hydrogen bonded connectivity between Glu-242 and $\mathrm{Cu_B}$ may originate from formation of an electric field between the two sites upon reduction of the latter, while reduction of heme a/BNC also affects the elec-

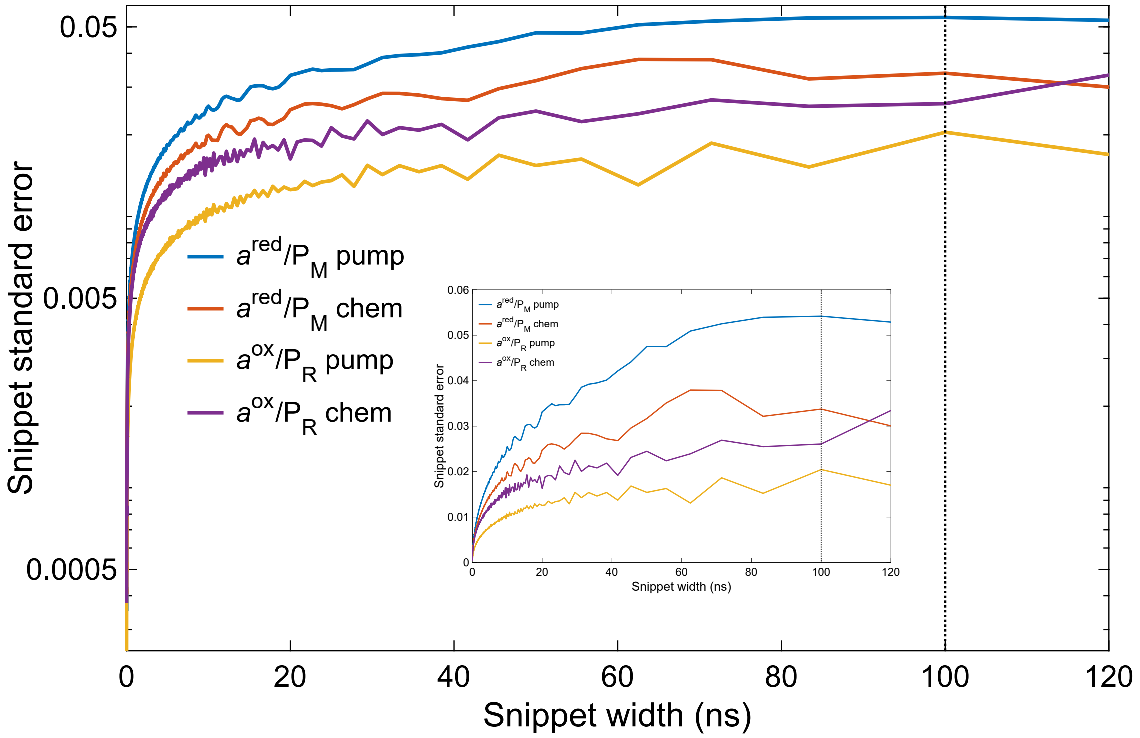

Figure 5.8: Error analysis for ζ observable. Snippet averaging analysis of ζ statistic on chemical and pump paths for redox states heme $a_{\mathrm{red}}/\mathrm{P_M}$ and $a_{\mathrm{ox}}/\mathrm{P_R}$ with standard error on log scale. Inset: non log-scale plot.

tric field along the pump pathway. These electric field variations in the non-polar cavity are consistent with the observed water dynamics (Figs. 5.4-5.7), and may be responsible for orienting the water dipoles to the electric field, and thus modulating the barrier for proton transfer, selecting between the two delivery sites.

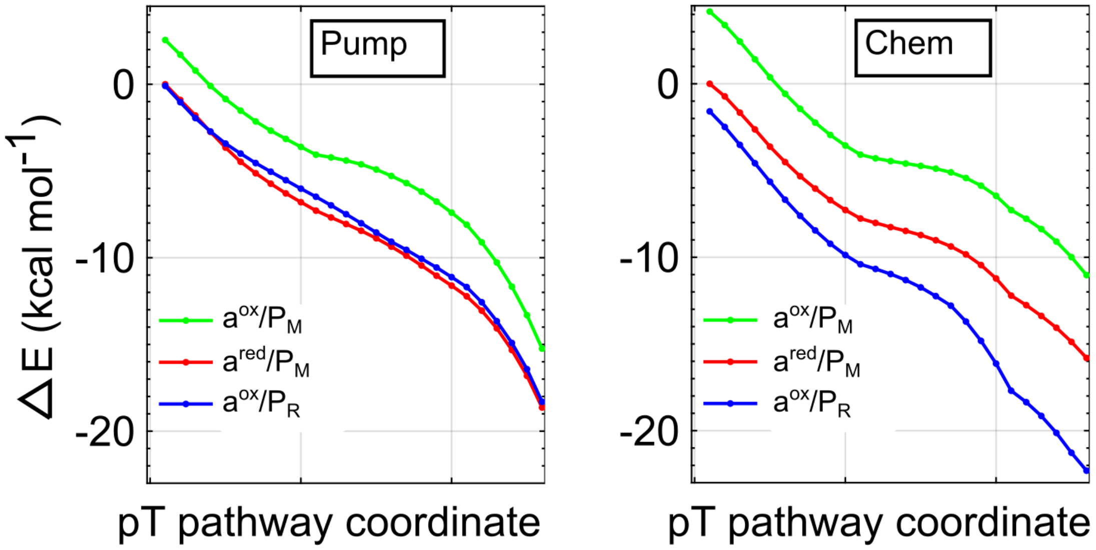

Figure 5.9: Electric field along the proton transfer paths going from Glu-242 to either BNC (hydroxo ligand of $\mathrm{Cu_B}$) or PLS(D-propionate group of heme a_3). The ox-ox, red-ox and ox-red labels indicate states with both heme a and BNC oxidized, heme a reduced and BNC oxidized, and heme a oxidized and BNC reduced, respectively.

5.4 Discussion

The functional water molecules in the active site of C*c*O may play a crucial role, switching the proton delivery routes based on the redox state of the enzyme. This rerouting is achieved by reorienting water molecules along the proton-transfer paths. The orientational behavior can be explained in terms of electric field fluctuations as a result of changes in the redox states of the enzyme. The role of water molecules has been subject to controversy due to lack of water molecules in the crystal structures of C*c*O. However, C*c*O produces two water molecules in each turnover, and thus the active site can easily undergo hydration and dehydration processes.

Warshel *et al.* [152] suggest that the primary effect governing the proton transfers from Glu-242 to the BNC and to the PLS, is the energetics for placing water molecules along the proton transfer pathways and the energetics for protonating these placed water molecules. The authors further suggested that orientations of water molecules have a minimal influence on the proton transfer energetics and that free energies profiles are the only way to characterize the proton transfer reaction and the gating mechanism. Recent theoretical studies by Voth *et al.* [158] suggest that prior formation of a water-wire, is not necessarily required for proton transfer reactions. Instead, protons can create their own paths via water wires that form as a result of the electric field generated by the hydronium (H_3O^+) species. However, it should be noted that without water molecules occupying the hydrophobic cavity, proton transfer over distances of 10 Å is not possible. Moreover, the presence of water molecules and their respective orientations favoring proton transfer are likely to minimize the kinetic cost of transferring a proton. It can be thus suggested that catalytic molecules in the hydrophobic cavity in the active site of C*c*O act as elements of a kinetic gate regulated by the redox-state of the enzyme by switching proton delivery paths and possibly preventing the leak of protons back to Glu-242.

5.5 Summary

In this study, the mechanism of proton gating in C*c*O was explored using MD simulations and PB calculations. Water-gated mechanism of proton delivery was probed by analyzing water dynamics from 0.5 μs MD simulations by implementation of Dijkstra's algorithm to evaluate water-mediated connectivities from Glu-242 to the BNC and the PLS. ζ, the longest connectivity along shortest path, obtained from Dijkstra's algorithm was used to quantify water-mediated connectivities. It has been shown from the ζ distributions that reduction of heme *a* drastically enhances the connectivity to the PLS while the connectivity to the BNC is lower, with a smaller probability. This suggests that the propensity of proton transfer to the PLS is higher when heme *a* is reduced, but nonetheless, though in smaller probability, the observation of hydrogen bonded pathways to BNC in heme *a* reduced state suggests that water orientation alone is unlikely to prevent the short-circuit reaction. Upon reduction of the BNC, connectivity to the PLS reduces and connectivity to the BNC increases, favoring proton transfer to the BNC. Statistical analysis of ζ suggested that the quantity requires 75-100 ns to converge, which is longer than the MD simulation timescales reported in literature used to study the water-gate mechanism in C*c*O. Electric field changes in the hydrophobic active site were also evaluated using PB calculations along the proton transfer directions from Glu-242 to the PLS and the BNC. Electric fields were sensitive to the redox states of the enzyme, and might provide the physical basis for orientating water molecules to facilitate proton transfer to different sites, depending on the redox state of the enzyme.

6 Proton-loading and proton transfer dynamics in cytochrome *c* oxidase

†

6.1 Introduction

The electron transfer (eT) in C*c*O takes place via the Cu_A center and an electron-queuing heme *a*, from which the electrons are transferred further to the BNC (Fig. 6.1). To satisfy electroneutrality [144], reduction of the BNC is coupled to its protonation, and the free energy released in this process is employed for proton pumping to the positively charged (P-side) of the membrane [22, 98] (Fig. 6.1). Experiments suggest that the pumped proton is transiently stored at an unknown proton-loading site (PLS), before it is ejected across the membrane [21, 22, 96, 98, 103]. The PLS thus plays a central role in transducing the free energy released from the reduction chemistry at the BNC into a proton motive force (pmf) across the membrane. Moreover, proton transfer (pT) to the PLS must be tightly gated in order to prevent the proton from leaking backwards in the pumping step towards its thermodynamically favorable direction.

Belevich et al. suggested based on time-resolved spectroscopic and electrochemical studies that the PLS is protonated in *ca.* 150 μs by a conserved glutamic acid residue, Glu-242, located at the end of the proton-conducting D-channel (Fig. 6.1), while reduction and protonation of the BNC led to relaxation of the PLS, and release of the proton to the P-side of the membrane in *ca.* 2.6 ms. Experimentally, it is, however, not possible to rule out that reduction of the BNC could take place concertedly with loading of the PLS [82]. By combining electrometric and structural information, it was suggested that the PLS could reside above heme a_3, e.g. at the D-propionate or A-propionate sites of heme a_3 [96]. Brzezinski and co-workers observed structural changes around the D-propionate of heme a_3 (D-prp, Fig. 6.1) when Glu-242 was mutated into a glutamine, suggesting that the observed structural transition might mimic a constrained conformation of the PLS [54]. Moreover, Ferguson-Miller and co-workers [284] found that the Mg^{2+} center (Fig. 6.1), located *ca.* 6 Å from D-propionate of heme a_3, transiently interacted with a proton in the pumping process. Transient Fourier-transformed infrared (FTIR) spectroscopy can provide valuable insight into the protonation dynamics of enzymes, but no direct FTIR signals have so far been linked to the protonation of the PLS. Nevertheless, Iwaki and Rich found that reduction of the electron-queuing heme *a* leads to spectral changes that might reflect vibrational shifts in guanidinium groups of arginines. This vibrational shift was assigned to Arg-438 (Fig. 6.1), which forms an ion-pair with the D-propionate of heme a_3, and could therefore also indirectly reflect the protonation of the PLS [133, 136, 137, 256]. Continuum electrostatic calculations [132, 138, 142, 285] further suggest that reduction of heme *a* might lead to an increase in the pK_a of several sites above heme a_3,

---

[†] This chapter has been published as "A Protonated Water Cluster as a Transient Proton-Loading Site in Cytochrome *c* Oxidase", S. Supekar, A. P. Gamiz-Hernandez, V. R. I. Kaila, Angew. Chem. Int. Ed. 2016, 55, 11940. (see Publications) and has been reproduced with permission from John Wiley & Sons, Inc.

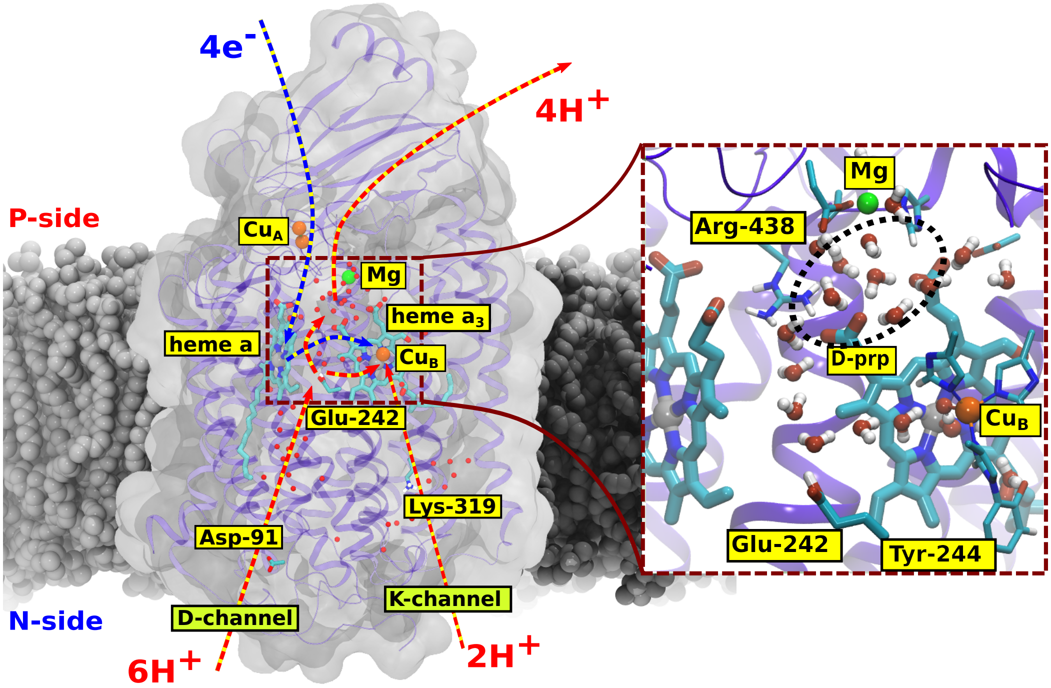

Figure 6.1: The structure and function of cytochrome c oxidase (C*c*O). Electron transfer from cytochrome c (not shown) at the positively charged (P) side of the membrane via $\mathrm{Cu_A}$, and heme *a* to the binuclear site (BNC), heme a_3/$\mathrm{Cu_B}$ center (blue arrows) leads to uptake of protons from the negatively charged (N) side of the membrane via the D-channel (Asp-91) and K-channels (Lys-319) (red arrows). Glu-242 at the end of the D-channel shuttles protons both the BNC and to the proton-loading site (PLS) above heme a_3, from which the protons are released to the P-side of the membrane. Inset: The BNC, and the hydrophobic active site cavity above Glu-242, and its surroundings.

but the exact location of the PLS, nevertheless, still remains unclear. Identifying the exact location and function of the PLS is central for understanding the molecular mechanism of proton-coupled electron transfer in C*c*O. This study aimed at identifying the location of the PLS, to elucidate the mechanism of proton transfer from Glu-242, and the gating of the loaded pumped protons.

6.2 Models and methods

6.2.1 QM cluster models

Quantum chemical cluster models with *ca.* 340 atoms were optimized based on the relaxed MD structure of the $a_{\mathrm{ox}}/\mathrm{P_R}$ state. The models included heme a_3, $\mathrm{Cu_B}$, $\mathrm{Mg^{2+}}$, and residues Glu-198, Glu-242, Val-243, Tyr-244, Trp-126, Trp-236, His-290, His-291, His-240, His-368, His-376, Asp-364, Asp-369, Arg-438 as well as 17 water molecules forming a hydrogen bonded network (Fig. 6.2). The amino acids were cut at Cβ positions, which were fixed during the structure optimization to account for protein strain [286]. The structures were optimized at the BP86-D3 level [203, 205, 214] using the multipole accelerated resolution of identity approximation (RI-MARIJ) [287], the m4 integration grid, and def2-SVP/def2-TZVP (Fe, Cu) basis sets [186]. The protein environment was modeled using a polarizable dielectric medium with a dielectric constant set to 4 using the COSMO model [288]. Single point energy calculations of the optimized structures were performed at the B3LYP-D3 level [212, 214], with def2-TZVP (Fe, Cu) and def2-SVP basis sets (C, H, O, N, Mg) and def2-TZVPPD basis sets for the oxygen atoms of the protonated water molecules (Fig. 6.2) above the D-propionate site of heme a_3. The vibrational spectrum for the PLS was calculated at RI-BP86/def2-SVP/def2-TZVP (Fe, Cu)/$\epsilon = 4$ level by estimating the molecular Hessian numerically. The DFT calculations were performed with TURBOMOLE v 6.5-6.6 [289].

IR spectra for DFT cluster models

One of the most popular methods for evaluating IR spectra computationally involves normal-mode analysis. This involves calculation of the Hessian matrix, the second derivative of energy with respect to atomic displacements, whose eigenvalues and eigenvectors are evaluated by diagonalization of the Hessian matrix. The obtained eigenvalues are the square of the vibration frequencies and the eigenvectors are the normal modes. The eigenvalues and eigenvectors can then be used to calculate IR spectra and thermodynamic quantities,such as, zero-point energy, entropy and enthalpy. The IR spectra for the quantum cluster models were obtained by numerically evaluating the Hessian for using DFT with Numforce module of Turbomole [289] for obtaining Hessian numerically. The protein environment was modeled using a polarizable dielectric medium (COSMO) [288] with a dielectric constant (ϵ) of 4.

6.2.2 Hybrid QM/MM simulations

Hybrid quantum mechanics/molecular mechanics (QM/MM) calculations were performed based on relaxed structures from MD simulations for each redox state. The QM region, comprising *ca.* 125 atoms, included the propionic groups of heme a_3, Mg^{2+}, Asp-364, Asp-369, His-368, Arg-438, Glu-198 and 18 surrounding water molecules (Fig. 6.5).The QM region was described at the RI-MARIJ/BP86-D3 [203, 205, 214, 287] level and def2-SVP basis sets [186], and link atoms were introduced between the Cα-Cβ atoms, with the remaining system was treated classically with classical CHARMM27 force field [226]. The MM region consisted of the remaining *ca.* 87,000 atoms including the protein, lipids, water and ions. The QM/MM systems were minimized for 100 steps, followed by 5 ps of QM/MM molecular dynamics simulations at T=310 K, and using an integration step of 1 fs. The QM/MM models were studied in the state $\mathrm{a_{ox}/P_M}$, $\mathrm{a_{red}/P_M}$, and $\mathrm{a_{ox}/P_R}$ with the water cluster above the D-propionate modeled in both its protonated (Fig. 6.6) and deprotonated forms (Fig. 6.7), and also with the A-propionate in a protonated state (Fig. 6.8, Fig. 6.9). QM/MM MD simulations were also initiated after 10 ns classical relaxation of the Zundel species above heme a_3 (Fig. 6.12). Classical force parameters for the Zundel species were obtained from B3LYP-D3/def2-TZVP [186, 212, 214] calculations of $H_5O_2{}^+$. The Particle Mesh Ewald ($\mathrm{P_M}$E) method [229] has not yet been implemented in the employed QM/MM scheme, but the QM system was explicitly polarized by *ca.* 87,000 MM atoms without employing cutoffs. The QM/MM MD calculations were performed using the CHARMM/Q-Chem interface [290].

QM/MM reaction pathway optimizations were performed with CHARMM/TURBOMOLE interface [289, 291] at the B3LYP-D3/def2-SVP level [186, 212, 214] with Glu-242, the propionic groups of heme a_3, Mg^{2+}, Asp-364, Asp-369, His-368, Arg-438, Glu-198 and 13 surrounding water molecules (N=123 atoms) to study the electrostatic dependence of redox states on proton transfer energetics from Glu-242 to the PLS in the $a_{\mathrm{ox}}/\mathrm{P_M}$, $a_{\mathrm{red}}/\mathrm{P_M}$, and $a_{\mathrm{ox}}/\mathrm{P_R}$ states. The pT pathways were optimized by using a linear combination of all water O-H bond distances between Glu-242 and the PLS, harmonically restrained with a force constant of 500 kcal mol^{-1} Å^{-2} between Q=[-5 Å, 5 Å], where, $Q = \sum_i (r^i_{\mathrm{OH}} - r^{i+1}_{\mathrm{H...O}})$. Harmonic constraints of 5 kcal mol^{-1} Å^{-2} were employed used to restrain the position of the oxygen atoms of the water molecules involved in the pT reaction. The QM/MM reaction pathway optimizations were initiated from relaxed MD structures of the $a_{\mathrm{ox}}/\mathrm{P_M}$ state, and the MM surroundings were kept fixed during the optimizations to avoid converging into a different minimum. QM/MM single point energies at B3LYP-D3/def2-TZVP level [186, 212, 214] were calculated for the optimized reaction profiles (Fig. 6.15).

6.2.3 Continuum electrostatics calculations

Continuum electrostatics calculations were performed to estimate the effect of heme *a* reduction on the protonated water cluster by solving the linearized Poisson-Boltzmann (PB) equation with the SOLINPROT module of MEAD [235]. To this end, structures obtained from 5 ps QM/MM MD simulations of the $a_{\mathrm{ox}}/\mathrm{P_M}$ state were employed. Solvation free energies of the solute were computed from the electrostatic energy of the solute in a dielectric media (reaction field) and from the interaction of the solute with the background protein charges. The solute consisted of the Zundel cation ($H_5O_2{}^+$) ($\epsilon_{solute} = 4$) and the protein together with internal water molecules were modeled as a set of atomic partial charges in a dielectric continuum with $\epsilon_{protein} = 4$, and $\epsilon_{external} = 4$. Charges and solvation radii were generated from our CHARMM force field parameter set [227]. The solvation free energy for the Zundel ($H_5O_2{}^+$) species in the $\mathrm{a_{ox}}/\mathrm{P_M}$ and $\mathrm{a_{red}}/\mathrm{P_M}$ states was calculated, and its dependence to the dielectric constant (Fig. 6.11) was examined. The atomic partial charges of the solute were computed at the B3LYP/def2-TZVP level [186, 212].

6.2.4 IR spectra from QM/MM trajectories

In order to evaluate IR spectra at ambient temperature (T=310K) with full description of the protein environment in QM/MM MD simulations, we calculated time-autocorrelation function for dipole operator. Fourier transform of the dipole autocorrelation function provides the IR spectra,

$$I(\omega) \propto \int_0^\infty \langle \vec{M}(0) \cdot \vec{M}(t) \rangle \cos(\omega t) dt, \tag{6.1}$$

where, $I(\omega)$ is the spectral density, $\vec{M}(t)$ is the total dipole moment of the system at time t, and ω is the vibrational frequency. The angular brackets represent an average taken over all time origins. The total dipole moment, $\vec{M}(t)$, is the vector sum of dipole moments of all individual molecules in the system

$$\vec{M}(t) = \sum_{i=1}^{m} \vec{\mu}_i(t) \tag{6.2}$$

where $\vec{\mu}_i(t)$ is the dipole moment of i^{th} molecule at time t, and m is the number of molecules in the system. The dipole moment $\vec{\mu}_i(t)$ is given by,

$$\vec{\mu}_i(t) = \sum_{j=1}^{n} q_j \vec{r}_j(t) \tag{6.3}$$

where n is the number of atoms in the i^{th} molecule, q_j is the partial charge of the j^{th} atom and $\vec{r}_j(t)$ is the position vector of the j_{th} atom at time t.

The dipole operator was evaluated for all residues and water molecules within 5 Å sphere of the protonated water cluster. ESP atomic charges from the last frame of QM/MM

MD simulations were used as partial charges for evaluating the dipole operator. The vibrational IR spectra was simulated using the IR spectral density plugin in VMD [278].

6.3 Results

6.3.1 Quantum cluster models (DFT)

Classically relaxed MD structures of C*c*O in P_M, P_R and F states were used to make quantum-cluster models to address the energetics and structure of likely protonation sites above heme a_3 with large-scale quantum chemical density functional theory (DFT) models. The quantum cluster models contain in addition to the active site, also its nearby protein residues and water molecules and range in size from 150-340 atoms. To study the protonation of PLS, starting from an optimized QM cluster model of protonated Glu-242 in the P_M state, the proton was transferred from Glu-242 via the hydrogen bonded water network to the D-propionate group of heme a_3, obtained from classical MD simulations (see chapter 5) followed by re-optimization of the structure with the proton on PLS (Fig. 6.2). This new state that mimics the protonation of the PLS, and is only *ca.* 8 kcal mol^{-1} higher in energy relative to the initial state with Glu-242 protonated. The optimized structures of the quantum chemical active site models are shown in Fig. 6.2. The endergonic energetics of protonation of PLS are in agreement with previous studies of the thermodynamics of proton pumping [106, 163, 292, 293]. Remarkably, we find that the proton, which was initially placed on the D-propionate group of heme a_3 moves in the structure optimization to the water cluster above this site, forming a protonated Zundel-like ($H_5O_2^+$) water cluster. The protonated water cluster was also found to be stable in $P_R$and F states, and resembled either Zundel ($H_5O_2^+$) or Eigen ($H_9O_4^+$) type structures [294], possibly reflecting the dynamical nature of the proton (Fig. 6.3). Protonating the A-propionate of heme a_3 resulted in structures that were much higher in energy, or ended in relaxation of the proton to the water cluster near the Mg^{2+} site. Optimized structures obtained after transferring proton from Glu-242 to PLS in $P_R$and F states were *ca.* 14 and 25 kcal mol^{-1} higher in energy, respectively. The energetics of proton transfer to PLS imply the PLS could be loaded in the P_M state.

The quantum chemical DFT models allow us to estimate relative pK_a changes of the water cluster upon changes in the protonation and redox state of the surroundings. In the cluster models of the $P_R$state, obtained by one-electron reduction of BNC, the proton affinity of the PLS increases by *ca.* 7 kcal mol^{-1} upon BNC reduction, whereas, protonation and reduction of the BNC, yielding the F state, results in the decrease of the proton affinity of the water cluster by *ca.* 9 kcal mol^{-1}. We also find that protonation of the PLS increases the electron affinity of the BNC by *ca.* 0.3 eV. The energetics from quantum cluster models (see Table 6.1) suggest that the protonation of the PLS leads to favorable reduction of the BNC, with the electron at BNC stabilizing the proton loaded at PLS; whereas, protonation and reduction of the BNC leads to destabilization of the protonated PLS leading to ejection of the proton to the P-side of the membrane. These results are consistent with experimental and theoretical studies of the pumping mechanism of C*c*O [22, 96, 98, 164, 292].

6.3.2 IR Spectra from DFT cluster models

To characterize the nature of the protonated PLS and its surroundings, vibrational IR spectrum was simulated by evaluating the molecular Hessian for the quantum cluster models. By comparing the IR spectra of protonated and deprotonated PLS in the P_M state, it was observed that protonation of the water cluster results in a large vibrational

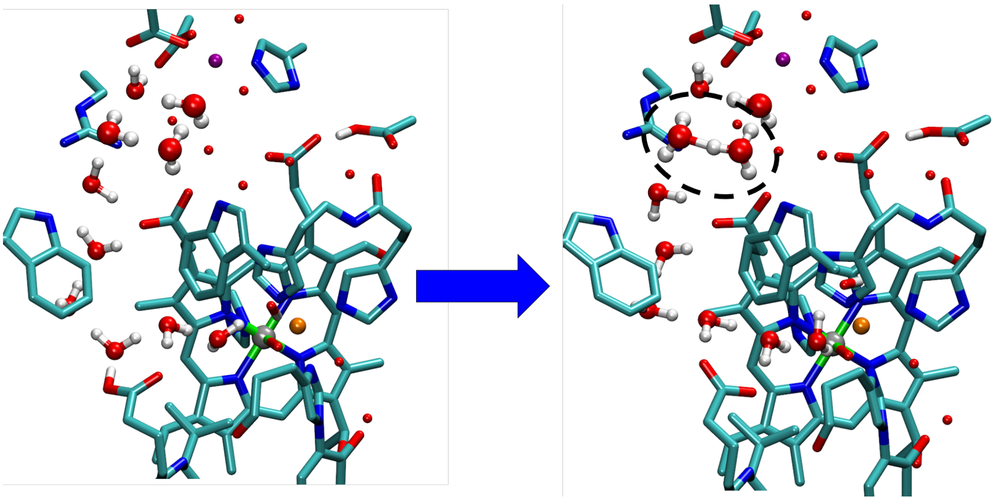

Figure 6.2: Optimized structures of transfer of a proton from Glu-242 to the D-propionate group of heme a_3. The starting structure for protonated Glu-242 was obtained from MD simulations. The structure was optimized using DFT at D3-B3LYP/def2-SVP level. From the optimized structure with protonated Glu-242, the proton from Glu-242 was moved the D-propionate group of heme a_3 by reorganizing the water molecules providing protonic connectivity. Optimization of this structure led to movement of the proton at D-propionate to the water cluster above heme a_3 forming a Zundel-like species. Other water molecules as shown as red spheres.

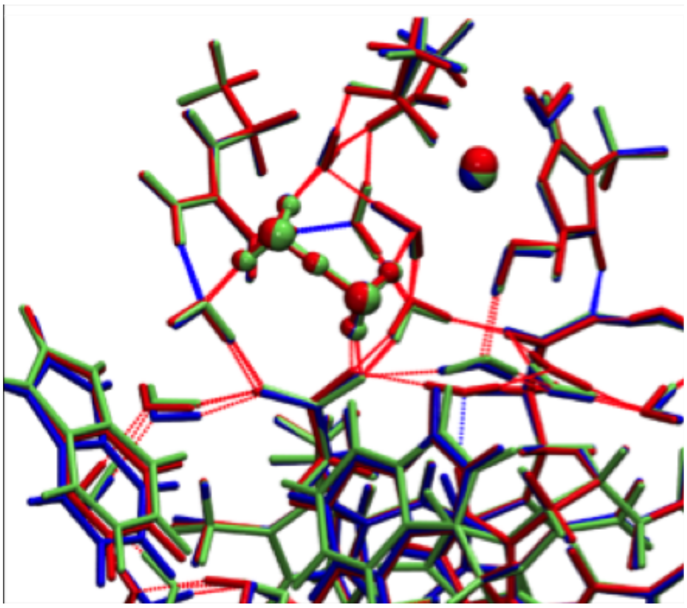

Figure 6.3: Structure of the protonated water cluster. DFT optimized structure of the protonated water cluster (PLSH) above heme a_3 in different proton and redox states of the surroundings with GluH/PLSH/P$_{\mathrm{M}}$ (in blue), GluH/PLSH/P$_{\mathrm{R}}$ (in red), and Glu^{-}/PLSH/F (in green).

shift in the water O-H bond stretches. Additionally, vibrational shifts linked to the guanidinium group of Arg-438 were also observed. The vibrational shifts coupled to protonated water cluster might be weak due to the large water content of C*c*O. Experimental IR spectroscopic studies [136] on the loading of PLS have reported vibrational shifts around 1592 and 1675 cm^{-1}, which were assigned to guanidinium groups of Arg-438. We observe similar

Reaction	**Method**	**State**		
Proton transfer		**P_M**	**P_R**	**F**
GluH/PLS → Glu^-/PLSH	QM	7.6	13.7	25.0
Relative proton affinity		**P_M**	**P_R**	**F**
PLS → PLSH	QM	0.0	-6.7	8.5
	QM/MM	9.8		
	Electrostatics	4.0		

Table 6.1: Energetics of the water cluster protonation in different redox states of C*c*O. The energetics are calculated at B3LYP-D3/def2-SVP/def2-TZVP(Cu,Fe)/$\epsilon = 4$ level for the DFT cluster models of the active site with N=340 QM atoms (labeled QM), or as an ensemble average from the 5 ps QM/MM trajectories (labeled QM/MM) with heme *a* reduced (a^{red}) or oxidized (a^{ox}), with BNC in P_M state. Glu and GluH refer to deprotonated and protonated forms of Glu-242, respectively. Labels PLS and PLSH refer to an unloaded and a loaded proton loading site, respectively. The standard deviation of the electrostatic stabilization energies is *ca.* 0.2 kcal mol^{-1} (see Fig. 6.10). Benchmarking calculations suggest that the B3LYP-D3 functional has an error of *ca.* 2 kcal mol^{-1} for protonation energies of simple model systems relative to CCSD(T) calculations (see Table 6.3, Table 6.4).

modes for guanidinium groups of Arg-438 at 1530 cm^{-1} and 1676 cm^{-1}, which shift to 1553 cm^{-1} and 1677 cm^{-1} upon protonation of PLS (Fig. 6.4, Table 6.2). Maréchal and Rich [137] reported structural reorganization in hydrogen bonded networks of water molecules upon reduction of metal centers in C*c*O, with some resemblance to protonated water clusters in bacteriorhodopsin [256]. This redox-state dependent signal of such dangling water molecules at 3600-3800 cm^{-1} could thus arise from the putative protonated water cluster which serves as PLS in DFT calculations. It should be noted that vibrational frequencies for hydronium (H_3O^+) like species are reported to be strongly anharmonic in nature [295].

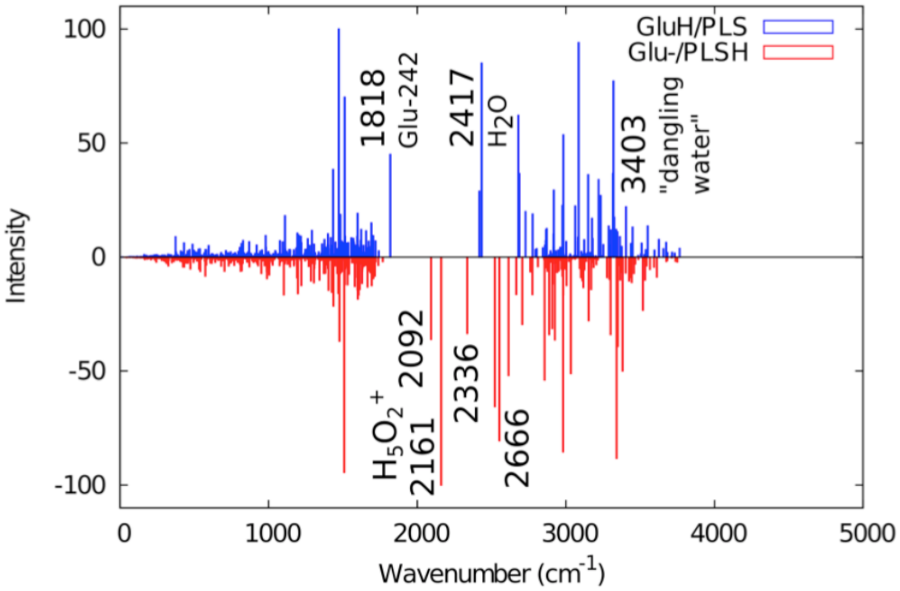

Figure 6.4: The vibrational spectrum obtained from the Hessian calculation at RI-BP86-D3/def2-SVP/def2-TZVP(Cu,Fe)/ϵ=4 level of theory. GluH/PLS and Glu-/PLSH labels indicate the location of proton; being either at Glu-242 or the proton loading site (PLS), respectively.

Group	**Protonation state**	
	GluH/PLS	**Glu$^-$/PLSH**
Glu-242	1818	
Arg-438	1546	1530
	1676	1676
	2727	2766
H3O+/$H_5O_2^+$		1768
		2092
		2666

Table 6.2: Vibrational normal modes (in cm^{-1}) linked to protonation of the water cluster above the D-propionate group of heme a_3 from QM models. The vibrational transitions were calculated at RI-BP86-D3/def2-SVP/def2-TZVP(Cu,Fe)/$\epsilon = 4$ level of theory. GluH/PLS and Glu$^-$/PLSH refer to the respective states with a deprotonated/protonated proton loading sites and Glu-242 protonated/deprotonated, with the BNC in the P_M state for both QM models.

6.3.3 QM/MM MD simulations

In order to examine whether the protonated water cluster remains dynamically stable at T=310 K, hybrid QM/MM MD simulations were performed on the protonated and deprotonated forms of the PLS. The starting structures for QM/MM MD simulations were obtained from relaxed MD structures for P_M (a_3^{IV}=O^{2-}/Cu^{II}-OH$^-$ TyrO$^\bullet$) state with heme a in oxidized or reduced forms (a_{ox}/P_M and a_{red}/P_M) and F (a_3^{IV}=O^{2-}/Cu^{II}-OH$_2$ TyrO$^-$) state formed after reduction and protonation of BNC with heme a in oxidized form (a^{ox}/F) (see Fig 2.6 for catalytic cycle). Time averaged snapshots of the QM/MM MD trajectories for the a_{ox}/P_M state with a proton loaded at the PLS are shown in Fig. 6.5. In the QM/MM MD simulations, similar to the DFT cluster models, the proton moves from the D-propionate group of heme a_3 to the water-cluster above heme a_3 and the protonated water cluster indeed remains stable on ps timescales. In the a_{red}/P_M state, the Zundel ion partially dissociates to Glu-198, while in the a_{ox}/P_M and /a^{ox}/F states, the protonated water cluster remains delocalized on the two water molecules between the D-propionate of heme a_3 and Asp-364/Glu-198 (subunit-II) residues which act as ligands to the Mg^{2+} cation (Fig. 6.5, Fig, 6.6). The partial dissociation may reflect that the PLS may redistribute among several sites, having a rather shallow energy surface, as suggested by recent continuum electrostatic calculations [142].

The water-cluster above heme a_3 bridging the D-propionate group of heme a_3, the aspartate and glutamate ligands of Mg^{2+}-center also remains stable for *ca.* 4 ps QM/MM MD simulations for all three redox sates without the proton loaded at PLS (Fig. 6.7). The positional robustness of these bridging water molecules in a highly polar and water filled region above heme a_3 is highly unusual for their environment. The water filled cavity encompassing the protonated water-cluster constantly exchanges water molecules, either through the Trp-126/D-propionate region or from the water filled H-channel.

To evaluate the possibility of the A-propionate group to serve as PLS in its full protein environment, QM/MM simulations were also performed by placing a proton on the A-propionate of heme a_3. After energy minimization, in the QM/MM MD dynamics, the proton dissociates from the A-propionate group and moves to the water cluster above heme a_3 bridging the D-propionate group of heme a_3 and the ligands of the Mg^{2+}-center (Fig. 6.8). In order to rule out the dependence of starting structure in the region above heme a_3 on the fate of the proton, another independent QM/MM simulation, starting with a different MD relaxed structure was performed with a proton at the A-propionate group. In this trajectory, similar to the previous calculations, the proton at A-propionate dissociates, moves to the water cluster and eventually diffusing towards the D-propionate forming a Zundel-like structure bridging the D-propionate and the carboxylate of the ligands of the Mg^{2+}-center (Fig. 6.9). These simulations further reinforce the notion that a protonated water-cluster could indeed serve as the PLS in C*c*O.

In order to investigate the role of redox state of heme *a* in the event of loading of a proton at the PLS in P_M state, we calculated solvation free energies of the protonated water-cluster in a_{ox}/P_M and a_{red}/P_M states using continuum classical electrostatics. The structures were obtained from the QM/MM MD simulation of the a_{ox}/P_M state. The effect of heme *a* reduction was studied by swapping between oxidized and reduced classical charge distributions of heme *a*. Another estimate of effect of the redox state of heme *a* was obtained as the difference of average energies from single point energy calculations on the QM/MM MD structures for a_{ox}/P_M state after swapping the redox state of heme *a* from oxidized to reduced. The reduction of heme *a* led to an increase in the proton affinity of the PLS by 4-10 kcal mol^{-1} (Table 6.1). Furthermore, classical electrostatic calculations on the QM/MM trajectories (Fig. 6.10) also suggest a similar increases in proton affinity upon heme *a* reduction.

Owing to the computationally intensive nature of QM(DFT)/MM MD simulations and size of the model employed here, it is not feasible to sample more than a few ps of dynamics. Hence, in order to ascertain the viability of the water cluster above heme a_3 to serve as the PLS on longer timescales, 10 ns of classical MD simulations were performed of the full system (*ca.* 88,000 atoms) with protonated water clusters above heme a_3 modeled as protonated PLS in a_{ox}/P_M and a_{red}/P_M states. The starting structures were obtained from QM/MM MD simulations. The protonated water cluster was parameterized for the classical force field as a Zundel cation ($H_5O_2{}^+$). The protonated water-cluster bridges the D-propionate of heme a_3, Asp-369 and Glu-198 (subunit-II) remains stable for 10 ns. The was followed by 0.5 ps of QM(DFT)/MM MD simulation based on 10 ns relaxed protonated PLS structures from MD simulations. The protonated water-cluster retained its Zundel character with the proton shared between two water molecules bridging the D-propionate of heme a_3 and the aspartate and glutamate ligands of the Mg^{2+}-cluster, thus further supporting the notion that a protonated water-cluster could serve as the proton-loading site in C*c*O (Fig. 6.12).

6.3.4 QM/MM spectra

By calculating the vibrational spectrum from the QM/MM MD trajectories, we find that the protonation of water cluster leads to disappearance of vibrational peaks at *ca.* 3600-3800 cm^{-1} (Fig. 6.13), which have been assigned to weakly hydrogen bonded dangling water molecules in FTIR studies. Interestingly, similar signals have also been associated with protonated water clusters in bacteriorhodopsin [256]. The QM cluster models further suggest that hydronium has vibrational modes at 1770-2670 cm^{-1} (Table 6.2, Fig. 6.4), and shifts linked to the guanidinium group of Arg-438 upon protonation of PLS from that may relate to findings by Iwaki and Rich [136].

6.3.5 QM/MM Dynamics of proton transfer to the PLS

The hydrogen bonded water-wires in the hydrophobic cavity above Glu-242 provide the conducting pathway to transfer protons to the PLS. Such a water-mediated proton transfer

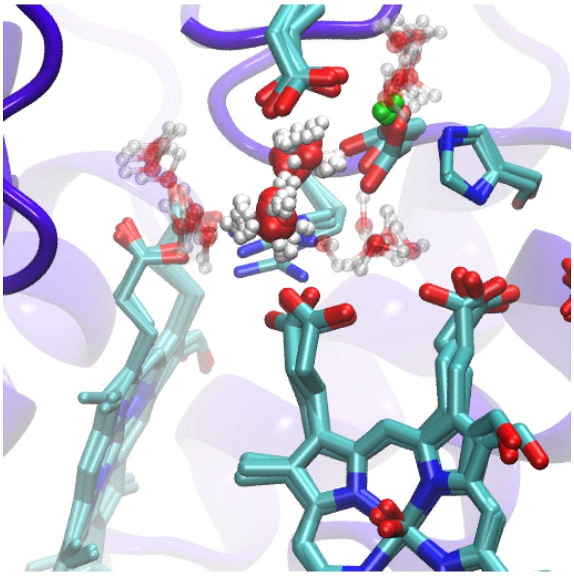

Figure 6.5: Snapshots from QM/MM dynamics of the protonated water cluster from a 5 ps trajectory with *ca.* 130 QM atoms and *ca.* 88700 MM atoms.

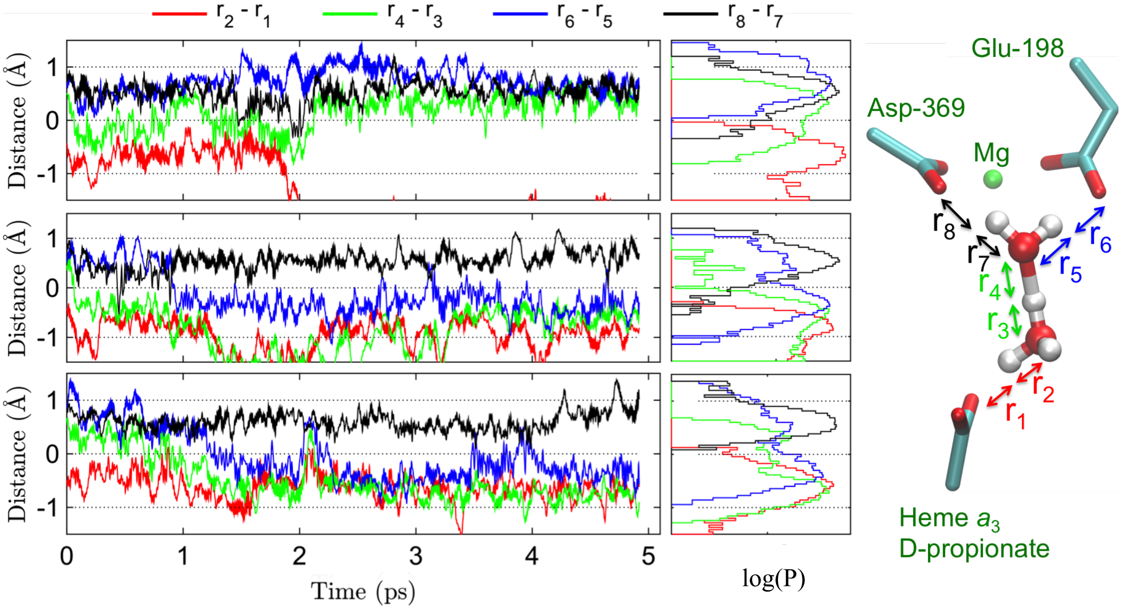

Figure 6.6: QM/MM dynamics and probability distributions of the central oxygen-hydrogen distance differences, shown in inset to the right. A value near 0 Å, indicates a proton shared between two oxygens, while values, $> +1.0$/ < -1.0 Å, indicate regular hydrogen bonds. The three panels show the dynamics in the $a^{\mathrm{ox}}/\mathrm{P_M}$ (top), $a^{\mathrm{red}}/\mathrm{P_M}$ (middle), $a^{\mathrm{ox}}/\mathrm{F}$ (bottom) states.

in the interior of proteins is generally perceived to occur by a Grotthuss mechanism. However, the dynamics and energetics of proton transfer from Glu-242 to the PLS situated at a distance of *ca.* 10 Å from it remains unclear and have not been explored at the DFT level. Loading a proton from Glu-242 to the PLS is an endergonic process with a high barrier and hence it is unlikely to observe proton transfer in unbiased QM/MM MD simulations. In order to facilitate the proton transfer process, a proton was introduced at the end of D-channel below the already protonated Glu-242. The introduction of an additional proton in order to trigger the proton transfer can be justified as such a configuration might be

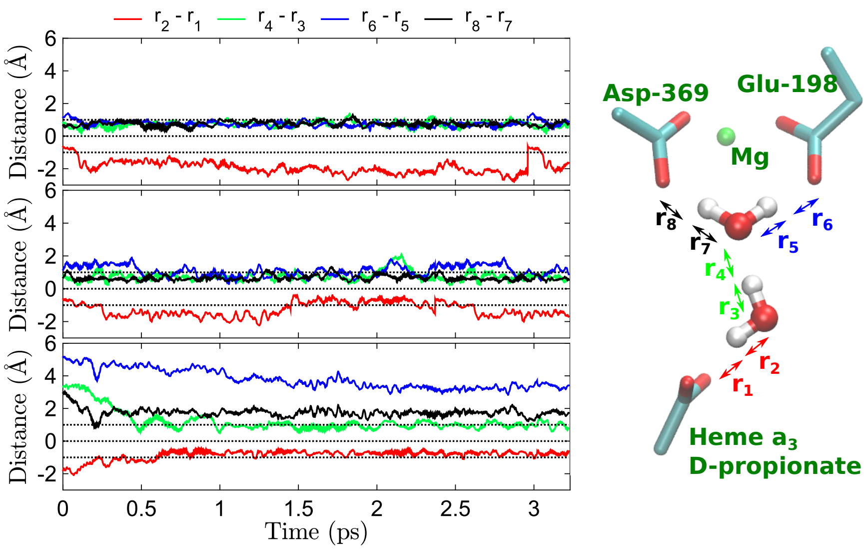

Figure 6.7: QM/MM MD dynamics of the redox states with a deprotonated PLS. A) QM/MM dynamics of the central oxygen-hydrogen distance differences, shown in inset to the right. A value near 0 Å, indicates a proton shared between two oxygen, while values $> +1.0/ < -1.0$ Å indicate regular hydrogen bonds. The three panels show the dynamics in the $a^{\mathrm{ox}}/\mathrm{P_M}$ (top), $a^{\mathrm{red}}/\mathrm{P_M}$ (middle), $a^{\mathrm{ox}}/\mathrm{F}$ (bottom) states.

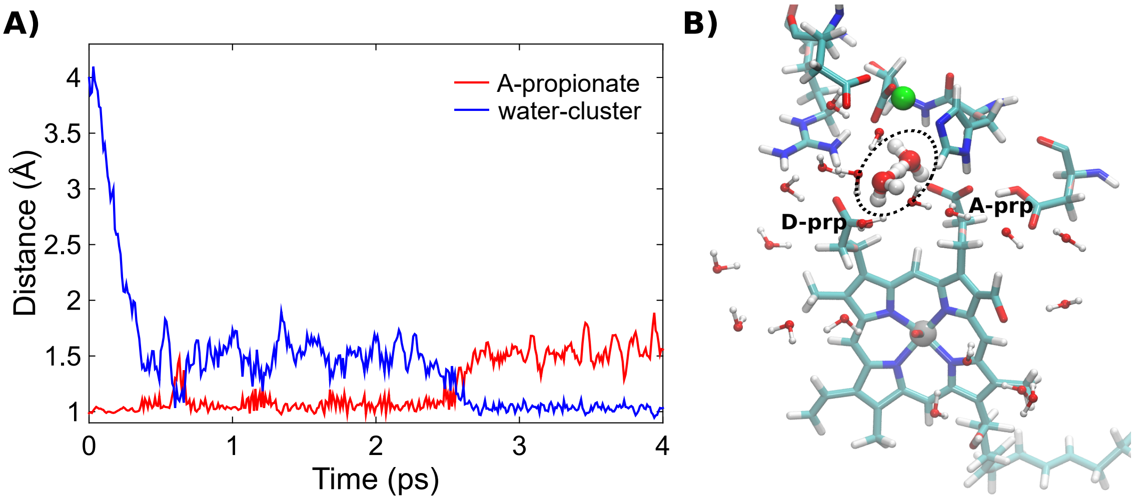

Figure 6.8: QM/MM MD simulations of state with a protonated A-propionate of heme a_3. A) The distance of the proton placed on A-propionate and the carboxyl oxygen of the A-propionate (in red), and between the proton and nearby water oxygen (in blue). B) Snapshot at 2.6 ps where the proton has moved into the water cluster above propionates (D-prp, A-prp), forming a Zundel ion (marked in a dashed black circle). See also Fig. 6.9 for an alternative trajectory with A-propionate modeled as the PLS.

transiently present in the event of a proton uptake occurring by a concerted proton transfer mechanism through the D-channel, as suggested by semi-empirical and reactive force field calculations [152, 296]. First-principles unbiased QM(DFT)/MM MD simulations were initiated with an excess proton below Glu-242, which results in proton transfer by a Grotthuss-type of hopping mechanism, shown in Fig. 6.14, typical for pT in short water wires, and it leads to the protonated water cluster above heme a_3. The protonation of the water cluster further supports the observations in QM and QM/MM MD simulations.

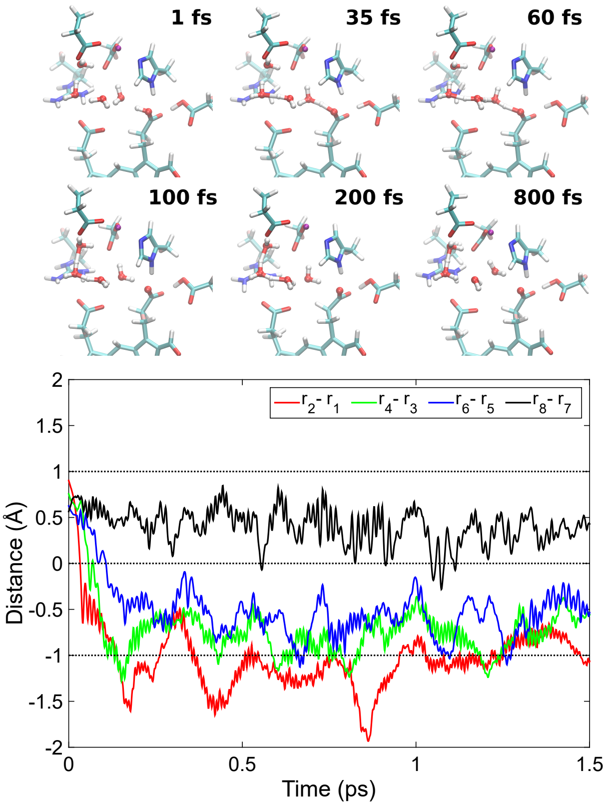

Figure 6.9: QM/MM dissociation trajectory of the proton from the A-propionate of heme a_3 to the water cluster above the D-propionate. r1-r8 are consecutive O-H bond distances from the A-propionate to the water molecule next to Asp-369.

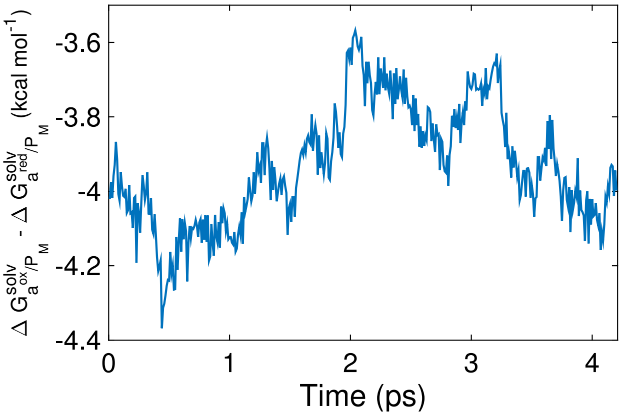

Figure 6.10: Classical electrostatics on the solvation energy of the protonated water cluster. $\Delta G/a^{\mathrm{ox}}/\mathrm{P_M}$ - $\Delta G/a^{\mathrm{red}}/\mathrm{P_M}$ is the difference in solvation free energy of the Zundel cation between with oxidized and reduced heme a states. The structures were obtained from QM/MM simulation of $/a^{\mathrm{ox}}/\mathrm{P_M}$.

The proton transfers to the water-cluster occurs in *ca.* 0.5 ps. The proton is transferred semi-concertedly along the water-wire via transient hydronium and Zundel-like intermediates to the putative water-cluster. Nevertheless, it should be noted that adding a proton below an already protonated Glu-242 introduces a large thermodynamics force inducing the proton transfer reaction. The dynamics observed, nonetheless, provide valuable insight into the molecular mechanism of loading of a proton at the PLS. The simulation shows

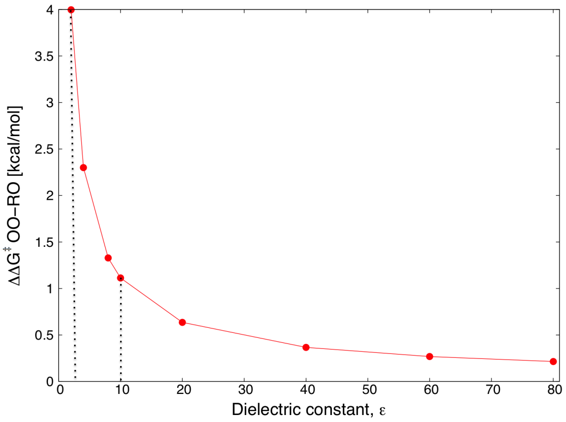

Figure 6.11: Solvation free energy difference of the proton transfer transition state (TS) structure between Glu-242 and PLS upon reduction of heme *a* ($\Delta\Delta G^{\ddagger}$ OO-RO). The solvation free energy is calculated as a function of the dielectric constant of non-polar cavity above Glu-242, which has been estimated to be around ϵ = 2-4, indicating that the heme *a* can stabilize the TS by up to 4 kcal mol^{-1}. The calculations were performed at Poisson-Boltzmann level with charge distributions of the TS structure obtained from B3LYP/def2-TZVP calculations.

that the proton is transferred by Grotthuss-like mechanism with formation of hydronium, Zundel-type cationic species along the path. Secondly, the simulation lends further support to the possibility of the water-cluster above the D-propionate of heme a_3 to serve as the PLS.

6.3.6 Energetics and reaction barriers of proton transfer

To further probe the redox state dependence of the energetics and barriers of proton transfer to the PLS, we performed QM(DFT)/MM reaction path optimizations of the proton transfer from a protonated Glu-242 to the PLS. The proton transfer path was optimized to mimic the structures formed in the Grotthuss type hopping mechanism seen in QM/MM MD proton transfer simulations (section). We obtain a proton transfer barrier of *ca.* 12-18 kcal mol^{-1} in the a_{red}/P_M, (Fig. 6.15), which is similar to the free energy profiles obtained from previous EVB calculations [152, 155]. Our calculations further suggest that in states with heme *a* oxidized (a_{ox}/P_M and a_{ox}/P_R) the barrier increases to *ca.* 20-22 kcal mol^{-1}, suggesting that the redox state of heme *a* electrostatically modulates the pT barrier, also in the case that proton transfer takes place from a neutral Glu-242, and may thus act as a gating element preventing proton back-flow Reaction scans can, however, be sensitive to local changes in the protein environment[297], for instance, the Trp-126/D-propionate hydrogen bond, which is broken in our simulations (Fig. 6.16), resembling configurations described by Goyal *et al.* [161]. Our solvation free-energy calculations, nevertheless, also support that the electron on heme *a* lowers the barrier by up to 4 kcal mol^{-1} (Fig. 6.10). Moreover, rotation of the anionic Glu-242 towards the D-channel, and its rapid re-protonation might further contribute in preventing the proton from leaking back to Glu-242 [99, 106, 149].

To evaluate the nature of Glu-242 with proton loaded at the PLS, MD simulations with a deprotonated Glu-242, BNC in F state and a proton loaded at the water cluster modeled as a Zundel cation were performed. The anionic Glu-242 side-chain during the simulation flips away from the hydrophobic cavity breaking the protonic connectivity to the PLS and hence preventing the back-flow of the proton loaded at PLS (Fig. 6.17). Glu-242 flips

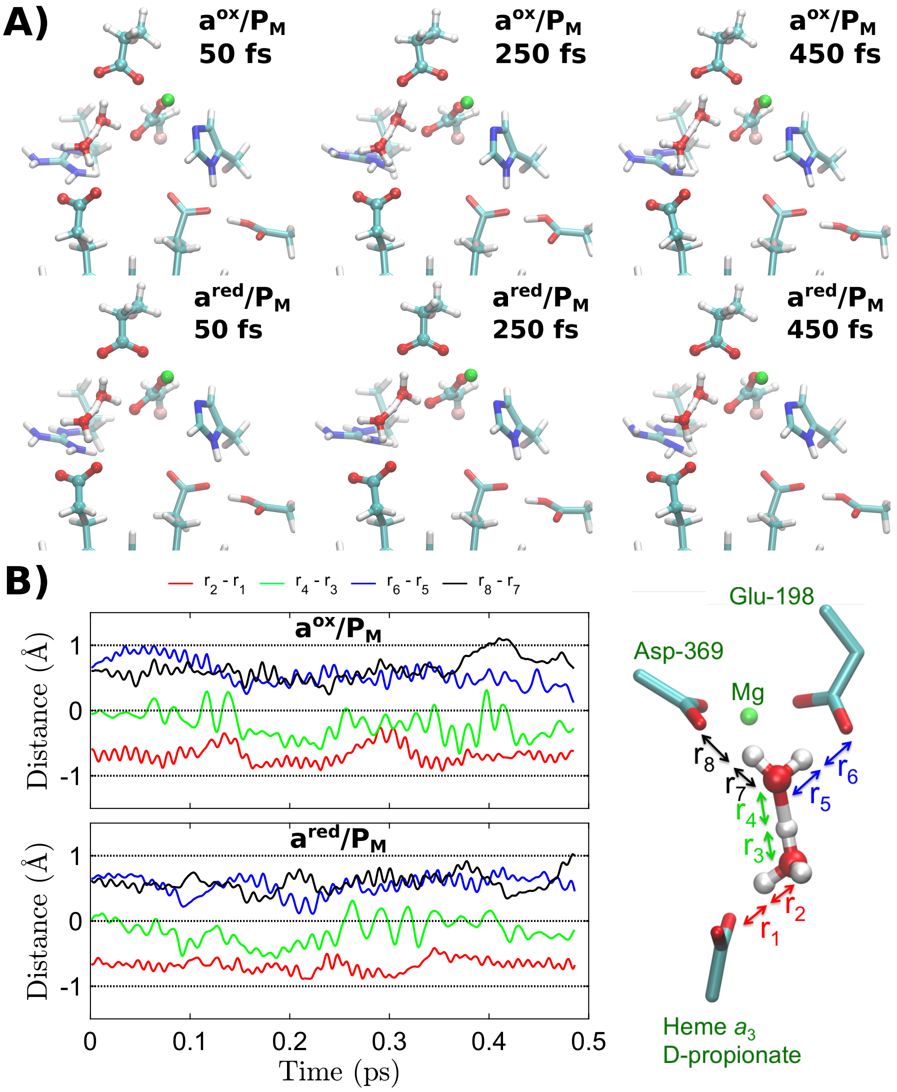

Figure 6.12: QM/MM MD of the protonated water cluster in the $a^{\mathrm{ox}}/\mathrm{P_M}$ and the $a^{\mathrm{red}}/\mathrm{P_M}$ states after 10 ns classical MD. A) Snapshots from QM/MM MD in the $a^{\mathrm{ox}}/\mathrm{P_M}$ (top) and $a^{\mathrm{red}}/\mathrm{P_M}$ (bottom) states initiated after 10 ns classical relaxation of the Zundel species. B) Probability distributions of the central oxygen-hydrogen distance differences in the $a^{\mathrm{ox}}/\mathrm{P_M}$ (top) and $a^{\mathrm{red}}/\mathrm{P_M}$ (bottom) states. A value near 0 Å, indicates a proton shared between two oxygens, while values $> +1.0/ < -1.0$ Å indicate regular hydrogen bonds. Inset (bottom right): Graphic definition of $\mathrm{r_1}$-$\mathrm{r_8}$.

down towards the D-channel to, ready to pick protons from the D-channel for further proton-transfer purposes in the catalytic cycle. The barrier obtained from the reaction path optimization for the PLS loading process in the $\mathrm{a_{red}/P_M}$ state is in good agreement with the reported experimental barrier of *ca.* 13 kcal mol^{-1}. To validate the DFT methodology used for performing reaction path optimizations, benchmarking calculations (Table 6.3, Table 6.4) at various levels of theory were performed. The hitherto employed DFT-methodology captures the proton affinity differences between carboxylates and water clusters with an accuracy of *ca.* 2 kcal mol^{-1} relative to correlated calculations. Moreover, entropic effects might further reduce the barriers by *ca.* 3 kcal mol^{-1} based on model calculations of proton transfer in water wires [298]. Tunneling effects based on measured H/D kinetic isotopic effects in C*c*O are expected to have a small contribution on the barriers [299, 300].

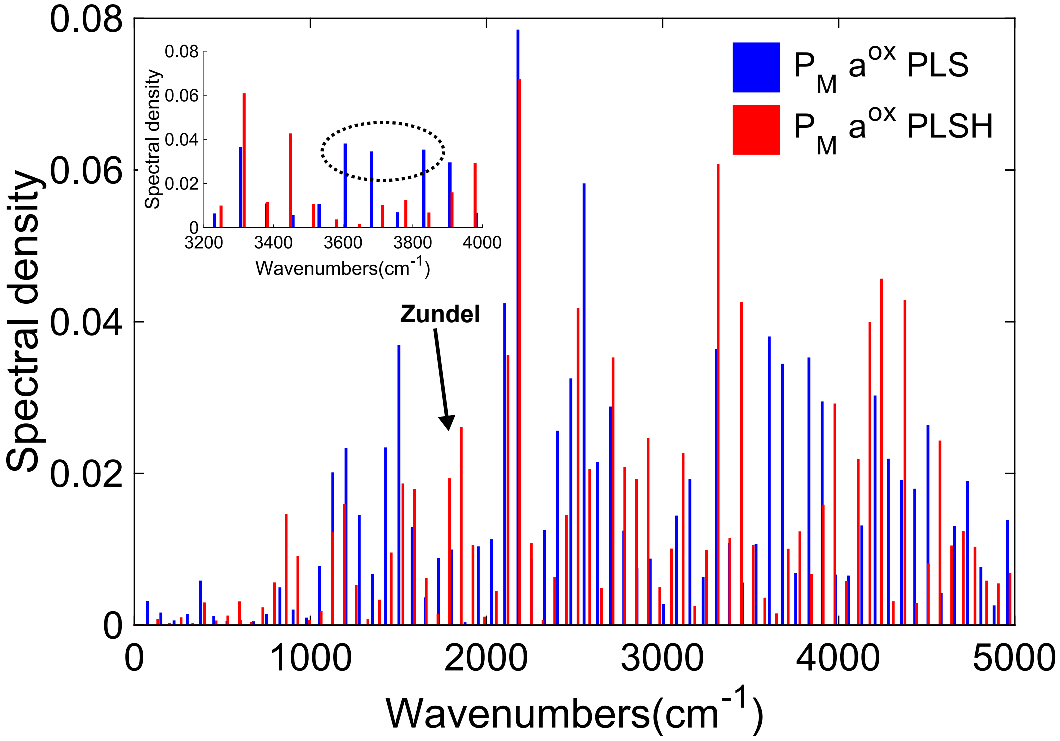

Figure 6.13: Vibrational spectrum of the protonated (PLSH) and deprotonated (PLS) water cluster above heme a_3 calculated from a 5 ps QM/MM trajectory. Vibrational shifts at *ca.* 1800 cm^{-1}, marked with a black arrow, arise from the Zundel species, while disappearance of vibrations at *ca.* 3600-3800 cm^{-1} (inset; dotted circle), indicate the absence of dangling bonds, a consequence of a tighter hydrogen bonding network.

DFT	E (kcal mol^{-1})	Ab initio	E (kcal mol^{-1})
B3LYP-D3 (c_x=0.20)	**19.9**	HF	32.5
B3LYP-D3 (c_x=0.10)	17.1	MP2	21.1
B3LYP-D3 (c_x=0.15)	18.5	SCS-MP2	23.6
B3LYP-D3 (c_x=0.25)	21.4	SOS-MP2	24.9
B3LYP-D3 (c_x=0.35)	24.1	CCSD	24.8
BHLYP	26.7	**CCSD(T)**	**22.6**
BP86	15.4		
PBE0	19.2		
TPSS	15.4		
TPSSh	19.2		

Table 6.3: Benchmarking the dissociation energy for GluH...$(H_2O)_4$ $\rightarrow$ Glu$^-$...H_2O $(H_3O^+...(H_2O)_2$ with def2-TZVPPD basis sets at different theory levels for structures obtained from QM/MM reaction scans (Fig. 6.15). The relative solvation energy from B3LYP-D3/def2-TZVPPD/$\epsilon = 80$ [186, 212, 214, 288] calculations has been added to each single point energy. DFT refers to DFT treatment with respective functionals. c_x refers to the amount of exact exchange employed in the Becke's three parameter functional [212]. Ab initio refers to wave function based methods. B3LYP-D3 underestimates the dissociation barrier by *ca.* 2 kcal mol^{-1} relative to CCSD(T).

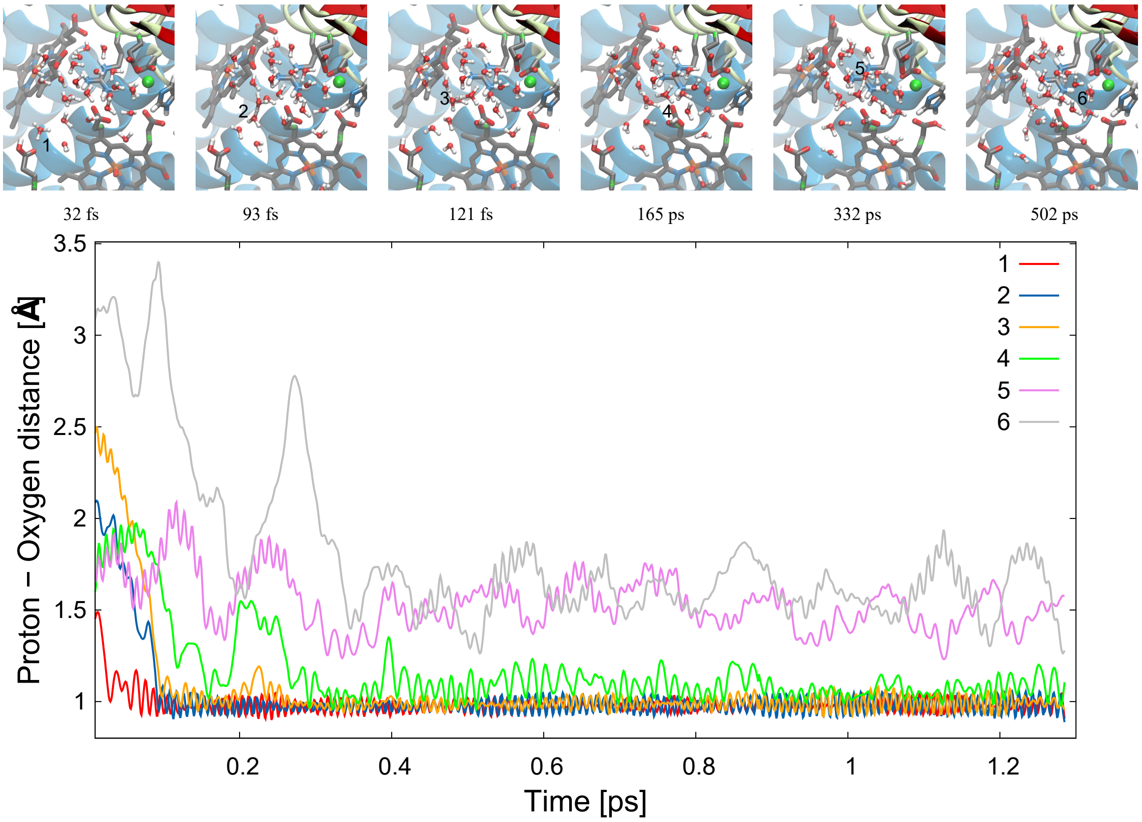

Figure 6.14: Proton transfer dynamics from Glu-242 to the PLS. Glu-242 protonates the water cluster above heme a_3 in *ca.* 0.5 ps along the water-chain that forms in classical MD simulations. The pT process takes place by a semi-concerted Grotthuss-type transfer mechanism. Water molecules 1-6 indicate the identity of the waters along the hydrogen bonded chain from Glu-242 (water 1) to the PLS (water molecules 5 and 6).

QM Method	**Proton affinity (kcal mol^{-1})**
B3LYP-D3	**-201.5**
HF	-201.3
MP2	-199.5
SCS-MP2	-199.9
SOS-MP2	-200.1
CCSD	-200.7
CCSD(T)	**-200.3**

Table 6.4: Benchmarking the proton affinity of a water dimer ($(H_2O)_2 \rightarrow H_5O_2^+$) using aug-cc-pVQZ basis sets [301]. The gas phase proton energy has not been subtracted from the reported values. The proton affinity of a water cluster is overestimated by *ca.* 1 kcal mol^{-1} at B3LYP-D3 level, relative to the CCSD(T) level of theory.

6.4 Discussion

The employed multi-scale molecular simulations of transient catalytic states in C*c*O suggest that a water cluster above the active site heme a_3 may function as a transient proton-loading site in the pumping process. This identified water cluster has an interesting resemblance to the proton-release group (PRG) of the light-driven proton pump, bacteriorhodopsin (bR) [256], in which the PRG is responsible for transiently storing the retinyl Schiff-base proton before it is released to the extracellular side of the membrane.

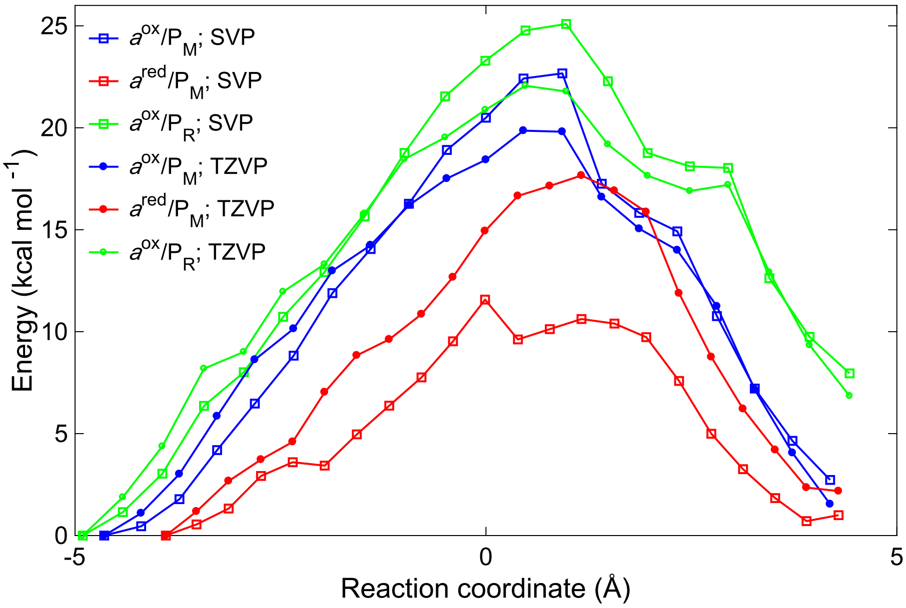

Figure 6.15: Basis set dependence of the proton transfer reaction profiles. Proton transfer reaction profiles for the loading of PLS from a neutral Glu-242 (without added hydronium) in the $a^{\mathrm{red}}/\mathrm{P_M}$ (red), $a^{\mathrm{ox}}/\mathrm{P_M}$ (blue), and $a^{\mathrm{ox}}/\mathrm{P_R}$ (green) states. Reaction profiles at B3LYP-D3/def2-TZVP level are shown as solid lines and profiles at B3LYP-D3/def2-SVP level are shown as dotted lines.

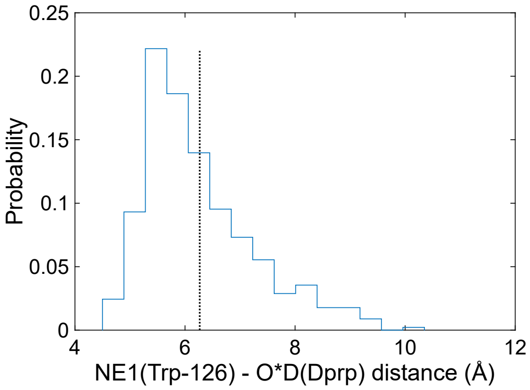

Figure 6.16: Trp-126 - D-propionate distance distribution from MD simulations. Distribution of minimum distance between the nitrogen atom (NE1) of Trp-126 and oxygen atoms (O1D or O2D) of the D-propionate of heme a_3 from 500 ns of MD in $\mathrm{P_R}$ state. Dotted black line indicates the mean of the distance distribution.

Site-directed mutagenesis experiments and FTIR data suggest that the PRG in bR may comprise a protonated water cluster bridging between Glu-194/Glu-204 and Arg-82, but recent semi-empirical tight-binding calculations suggest that the nearby protein residues prefer the proton to the water cluster [296]. Interestingly, experimental IR studies have reported the presence of similar dangling water signals for C*c*O too [137]. The protonated water cluster has also an interesting structural resemblance with the PRG of bR, with the cluster bridging between Arg-438, D-propionate of heme a_3, and Glu-198/Asp-368.

Functionally important elements in enzymes are usually structurally conserved. We find that in all available high-resolution X-ray structures of heme-copper oxidases, the water molecules involved in forming the protonated water cluster are also structurally

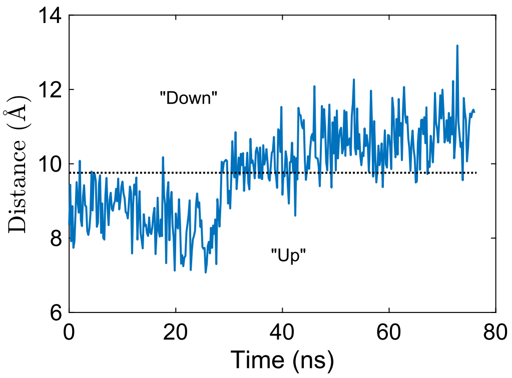

Figure 6.17: Glu-242 gating in F state. Distance between the anionic Glu-242 and the D-propionate of heme a_3 in the F state with a protonated water cluster above heme a_3. Glu-242 flips to a ”down” position, with connectivity towards the D-channel that may lead to its re-protonation and prevent the protonated water cluster from leaking backwards, in accordance with suggestions from several groups [99, 106, 147] but cf. also [160].

conserved (Fig. 6.18), although the MD simulations lead to an increased hydration state around the active site (Fig. 6.2, Fig. 6.3). This water cluster is also present in the X-ray structure of ba_3-oxidases, which may support the functional relevance of these water molecules in the pumping machinery. Interestingly, a perturbation in the structure of the water cluster above the D-propionate of heme a_3 was observed in the X-ray structure of some of the D-channel mutants, which do not pump protons but have regular or increased oxygen reduction activities (Fig. 6.18 B, C). If the water molecules involved in forming the putative protonated water cluster are functionally important, it also implies that a ”dry” PLS region might lead to a compromised proton-pumping activity, which can be observed under certain experimental conditions, such as in the resting oxidized (O) state [22, 82, 131]. Reduction and re-oxidation can be used for re-activation of the O state into a pumping OH state, which is also linked to synthesis of new water molecules at the BNC, and could ”re-wet” the water cluster at the PLS. Previous MD simulations indeed suggest that water molecules leave the non-polar cavity above Glu-242 via opening of the D-propionate/Arg-438 ion pair [283].

C*c*O contains several water molecules, which might explain why no direct spectroscopic signals have been assigned for the PLS protonation. Moreover, C*c*O synthesizes *ca.* 1000 water molecule s-1, which is likely to results in constant exchange of the catalytically active water cluster during turnover [22]. Our identified structure and dynamics of the protonated water cluster above the D-propionate region of heme a_3 can, nevertheless, be used to designing isotope-labelling and site-directed mutagenesis experiments that can further help to experimentally characterize the protonated water cluster at the PLS region.

6.5 Summary

It has been shown here that structurally conserved water molecules may function as a transient proton-loading site (PLS) and provide important functional coupling elements in the proton-pumping machinery of C*c*O. Multi-scale molecular simulations suggest that protonation of this water cluster is sensitive to the redox-state of the enzyme, and takes place by a concerted water-mediated Grotthuss-type transfer mechanism from Glu 242. The protonated water cluster remained stable in both large-scale quantum chemical calcula-

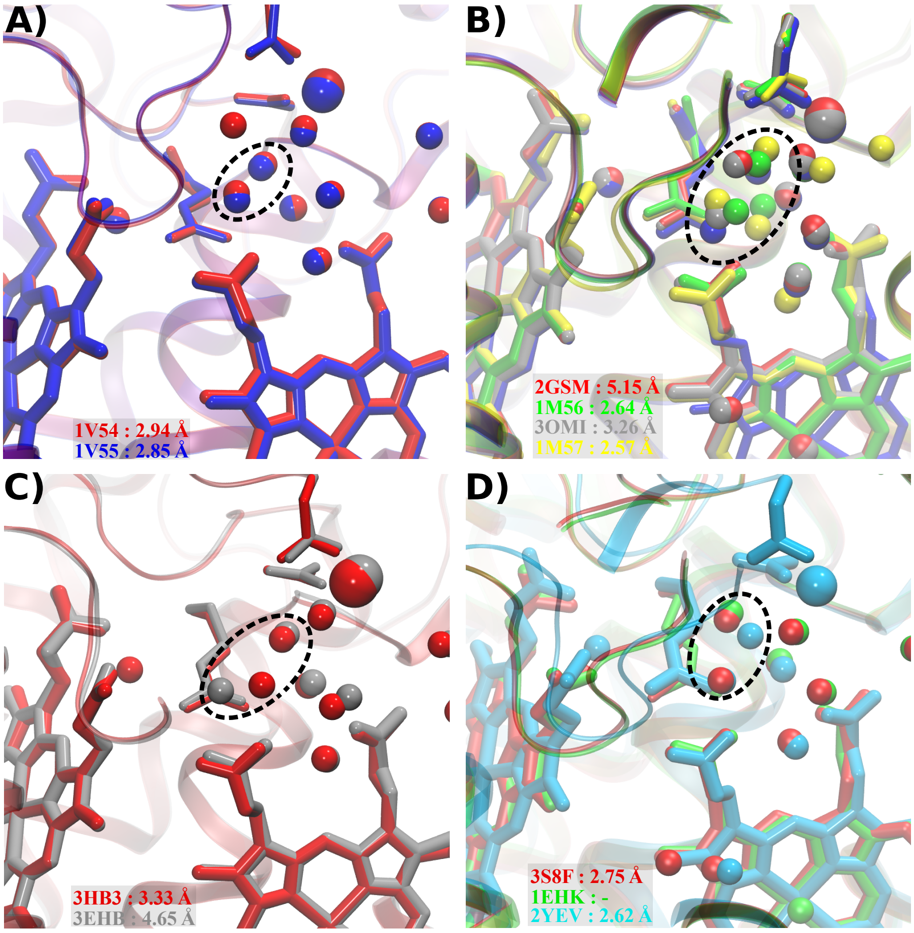

Figure 6.18: The central water molecules involved in forming a protonated water cluster in the quantum molecular simulations are marked with dashed ellipses. A) C*c*O from *Bos taurus*, with fully oxidized structures shown in red (PDB ID: 1V54), and fully reduced structure (PDB ID: 1V55) shown in blue, respectively. (B) C*c*O from *Rhodobacter sphaeroides* with fully oxidized wild type structures shown in red (PDB ID: 2GSM) and green (PDB ID: 1M56); the D132A (PDB ID: 3OMI) and E286Q (PDB ID: 1M57) mutant structures (rhodobacter numbering) shown in silver and yellow, respectively; and fully reduced structure (PDB ID: 3FYE) in blue (C) C*c*O from *Paracoccus denitrificans*: fully oxidized structure in red (PDB ID: 3HB3) and the non-pumping N131D mutant (Paracoccus numbering, PDB ID: 3EHB) shown in silver. (D) C*c*O from *Thermus thermophilus*: fully oxidized *ba*$_3$-type C*c*O in red (PDB ID: 3S8F) and green (PDB ID: 1EHK), and *caa*$_3$-type C*c*O in cyan (PDB ID: 2YEV).

tions and hybrid quantum mechanics/classical mechanics (QM/MM) molecular dynamics simulations on ps timescales. Consistent with experiments, the results also indicate that protonation of this putative water cluster increases the redox-potential of the active site, while reduction and protonation of the latter was coupled to the decrease of its pK_a that is likely to couple to the release of the proton to the P-side of the membrane. Computed vibrational shifts from protonation of this water cluster are also consistent with FTIR data. In summary, the multi-scale classical and quantum molecular simulations suggest that water molecules in C*c*O provide important coupling elements with distinct similarities to other energy converting proton-pumps such as bacteriorhodopsin and the respiratory complex I.

7 Activation and role of the K-channel in cytochrome *c* oxidase

[†]

7.1 Introduction

The reaction cycle of cytochrome *c* oxidase (C*c*O) [22] (Fig. 7.1) is initiated by binding and splitting of the dioxygen bond, which transfers the oxidative power of O_2 to the enzyme. The four electrons borrowed from the enzyme are stepwise regained in subsequent reaction steps by electron transfer (eT) from cytochrome *c* to the BNC, providing the thermodynamic driving force for the proton pumping process. The eT in C*c*O takes place via the Cu_A center and an electron-queuing heme *a* center, from which the electrons are transferred further to the binuclear center (BNC) [302]. Reduction of the BNC couples to loading of a physical proton to the proton-loading site (PLS)[21, 22, 82, 96]. This is followed by transfer of a chemical proton involved in oxygen reduction chemistry to the BNC leading to eventual pumping of protons across the membrane [22, 135, 303]. All physical protons are taken up from the D-channel, named after the conserved residue, Asp-91 [97, 101, 102]. The D-channel also supplies two chemical protons in the oxidative half of the reaction cycle, whereas the K-channel, named after the conserved residue, Lys-319, is activated in the reductive half of C*c*O reaction cycle (Fig. 7.1, Fig. 2.6), and is employed for the uptake of two chemical protons [97, 101, 102, 116, 304, 305]. It is also noteworthy that other oxidases, namely, B- and C-type C*c*Os entirely lack the D-channel, and are suggested to employ a K-channel analogue for conducting both, the physical and the chemical protons [41–43]. Although C*c*O has been extensively studied over the years, the role of K-channel and the molecular/enzymatic cause for a switch from the D- to K-channel going from oxidative to reduction half of the catalytic cycle for uptake of chemical protons still remains unclear [22, 94, 101, 116] .

In the oxidative phase (Fig. 7.1), following oxygen binding (A state), splitting of the dioxygen bond leads to the P_M state (a_3^{IV}=O^{2-}/Cu^{II}-OH^- TyrO$^\bullet$). Electron transfer to the BNC results in the experimentally observable P_R state (a_3^{IV}=O^{2-}/Cu^{II}-OH^- TyrO$^-$). This is coupled to loading of a proton at the PLS and is followed by transfer of a proton to BNC following pumping of the proton leading to F-state [22, 96]. Analogous to the P_M→P_R→F sequence, the F-state decays via a reduced state into the oxidized, O_H state. In the P_R→F→O sequence, Tyr-244 remains in the deprotonated form, based on the FTIR data by Gorbikova *et. al.* [68]. C*c*O can also relax to a resting oxidized state, O, which is a non-pumping state. The exact molecular differences between O and O_H are unknown and the two states exhibit same spectroscopic signatures but reduction and re-oxidation can be used for re-activation of the O state into a pumping O_H state [130, 131]. One electron reduction of the O_H state gives the $O_{H,R}$ state (a_3^{III}-OH^-/Cu^I-OH_2 TyrO$^-$). This transition is analogous to the P_M to P_R transition. This excess charge at BNC leads to loading

---

[†] This chapter will be published in the manuscript (in preparation) titled "Activation of the K-channel in Cytochrome *c* Oxidase" (see Publications)

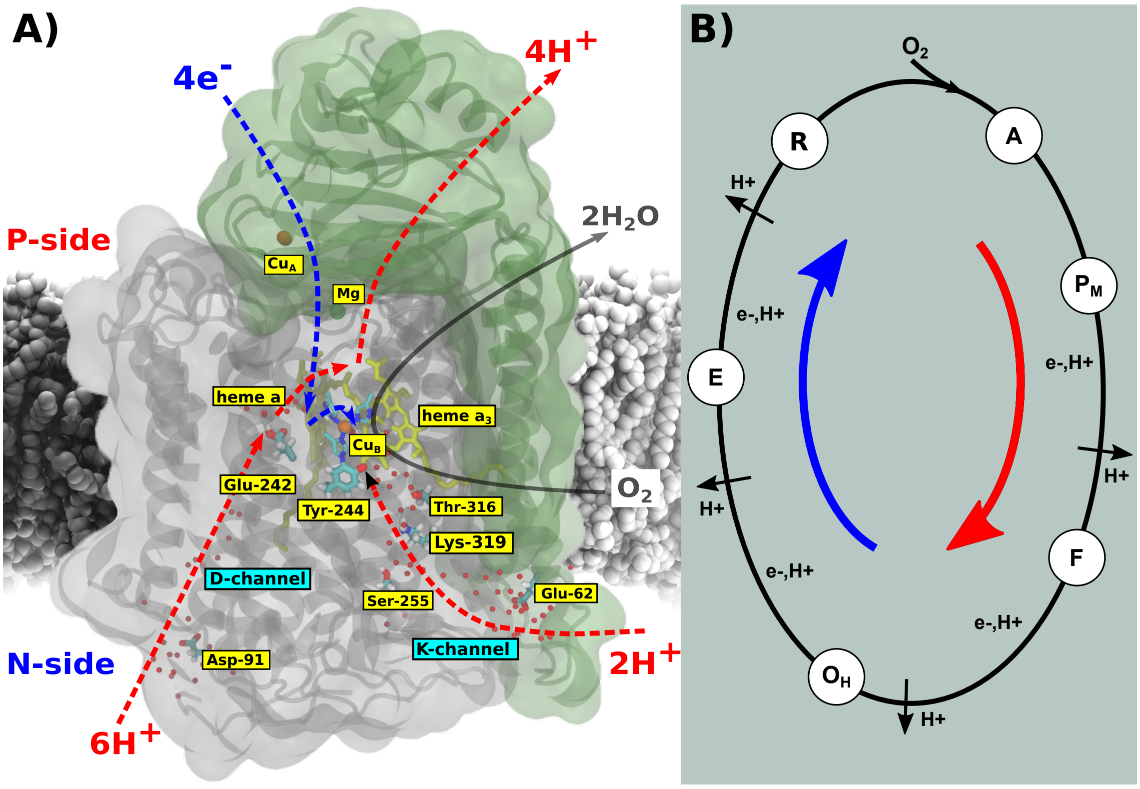

Figure 7.1: The structure, function, and reaction cycle of cytochrome c oxidase (CcO). A) Electron transfer from cytochrome c (not shown) at the positively charged (P) side of the membrane via Cu$_\mathrm{A}$, and heme a to the binuclear site (BNC), heme a_3/Cu$_\mathrm{B}$ center (blue arrows) leads to uptake of protons from the negatively charged (N) side of the membrane via the D-channel (Asp-91) and K-channel (Lys-319) (red arrows). The K-channel starts at Glu-62 at the N-side and proceeds via Ser-255, Lys-319, Thr-316 residues ending at Tyr-244 of the Cu$_\mathrm{B}$ center. The Lys-319 residue in the K-channel supplies two protons to the BNC for oxygen reduction chemistry [305]. B) The catalytic cycle of CcO with key catalytic states. Red and blue arrows indicate the oxidative and the reductive phases of catalytic cycle, respectively. Black arrows with H$^+$ labels indicate a pumping event.

of a physical proton to the PLS, following which TyrO$^-$ attains a chemical proton from the K-channel to yield the E state. Further protonation and reduction of the E state leads back to the reduced R state of the enzyme completing the catalytic cycle and producing two molecules of water and pumping four protons across the membrane [22, 68, 96, 98, 126].

The K-channel starts at residue Glu-62 (subunit II) and comprises Lys-319, Thr-316, the farnesyl group of heme a_3, and terminates at Tyr-244, providing a proton conduit from the negatively charged N-side to the BNC [94, 95]. Although, the K-channel supplies protons only in the reductive phase (O→E→R), electrometric studies of the K319M mutant (bovine numbering) indicate that the electron transfer to BNC is coupled to charge transfer in the K-channel. Mechanistic studies suggest that in the oxidative phase, K-channel plays a passive role as a dielectric well with Lys-319 flipping up towards BNC and its positive charge density facilitating the reduction of BNC [22, 255]. In mutation studies of the K-channel, reduction of the fully oxidized state was dramatically slowed down, whereas, no inhibition was observed in oxidation of the fully reduced state, suggesting that the role of K-channel functions as a conduit for chemical protons in the reductive half of the catalytic cycle [304, 306–308]. MD studies of the K-channel have already highlighted the role of hydrogen bonded water channels and persistent water molecules in the K-channel

[309]. Recent computational voltage studies have provided further evidence of the Lys-319 up-flip assignment from electrometric studies [310]. Continuum electrostatics calculations based on the crystallographic structures suggest that Lys-319 is protonated in both, the fully oxidized and fully reduced forms of C*c*O [311], whereas, continuum electrostatics pK_a calculations based on MD structures suggest that Lys-319 exists in its deprotonated form throughout the catalytic cycle except in cases of uncompensated charge at BNC (P_R, $O_{H,R}$) where Lys-319 is protonated [312]. DFT studies also suggest that Tyr-244 may play a key role in the K-channel activation, and act as the gating element by modulating its pK_a [67, 118, 131, 313]. Although, some mechanistic aspects of the K-channel are much better understood than before, the activation mechanism of the K-channel, and how the chemical protons are transferred along the channel remain largely unanswered [82].

In this study, $P_M \rightarrow P_R$ and $O_H \rightarrow O_{H,R}$ transitions were studied to elucidate the role of Lys-319 residue, water and protein dynamics within the K-channel, and the energetic basis of activation of the K-channel. To this end, classical molecular dynamics (MD) simulations that provide valuable insights into the protein and water dynamics were employed. The structures from MD simulations have been subjected to continuum electrostatics calculations that provide pK_a estimations for catalytic residues inside the enzyme. Quantum chemical DFT cluster models were employed to obtain accurate energetic description of the proton transfer process along the channel, while hybrid quantum mechanics/molecular mechanics (QM/MM) free energy sampling were employed to estimate free energy landscapes of the proton transfer process in a native-like environment.

7.2 Models and methods

7.2.1 MD simulations

Starting configuration for molecular dynamics (MD) simulation was based on X-ray structure of bovine C*c*O (PDB ID: 1V54) [50]. Subunits I and II of C*c*O were embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane, TIP3P water molecules, and a 100 mM NaCl concentration (28 Na^+, 9 Cl^-), in total 88,837 atoms. Residues Glu-242, Asp-364 and Lys-319 of subunit I were modeled in their protonated states, with 4-6 water molecules added the non-polar cavity above Glu-242, as suggested in previous studies. The study was conducted with heme a modeled in the oxidized state, and the BNC in the P_M (a_3^{IV}=O^{2-}/Cu^{II}-OH^- TyrO$^\bullet$), P_R (a_3^{IV}=O^{2-}/Cu^{II}-OH^- TyrO$^-$), O_H (a_3^{III}-OH^-/Cu^I-OH_2 TyrO$^\bullet$) and $O_{H,R}$ (a_3^{III}-OH^-/Cu^I-OH_2 TyrO$^-$) states, based on DFT parameterization of the redox sites [276] and further in-house parameterization for the $O_{H,R}$ state. 500 ns MD simulations were performed for all four states using the CHARMM36 force field [227], a 2 fs integration step, with the Particle Mesh Ewald (PME) treatment [229] for long-range electrostatics. The MD simulations were performed with NAMD 2.9 [277] along with VMD [278] for visualization.

7.2.2 pK_a calculations

Continuum electrostatics pK_a calculations were performed on 100 structures obtained at every 5 ns from each of the 500 ns MD simulations trajectories of P_M, P_R, O_H and $O_{H,R}$ states. All water molecules and lipids from the MD structures were removed. The removed water and lipid molecules were modeled implicitly with dielectric constants(ϵ) of 80 and 2, respectively. The protein was considered explicitly with partial charges from the CHARMM36 [227] parametrization and was embedded in a low dielectric constant ($\epsilon = 4$). To estimate the protonation probabilities, linearized Poisson-Boltzmann equation was solved in MEAD [235]. Monte-Carlo pH titration of the 2^N titrable residues in the protein was performed using Karlsberg [243].

7.2.3 QM Cluster Models

Quantum chemical cluster models were constructed with *ca.* 220-230 atoms, based on a 100 ns relaxed MD structure of the P_R and $O_{H,R}$ states. The models included heme a_3, Cu_B, and residues Tyr-244, His-290, His-291, His-240, Thr-316, Lys-319 as well as 17-20 water molecules forming a hydrogen bonded network. The amino acids were cut at the $C\beta$ positions, which were fixed during the structure optimization to account for protein strain. The structures were optimized at the dispersion corrected BP86-D3 level [203, 205, 214] using the multipole accelerated resolution of identity approximation (RI-MARIJ) [287], the m4 integration grid, and def2-SVP/def2-TZVP (Fe, Cu) basis sets [186]. The protein environment was modeled using a polarizable dielectric medium with a dielectric constant set to 4 in the COSMO model [288]. Single point energy calculations of the optimized structures were performed using the B3LYP-D3 functional [212, 214], with def2-TZVP (Fe, Cu) and def2-SVP basis sets [186]. The DFT calculations were performed with TURBOMOLE 6.6 [289].

7.2.4 Hybrid QM/MM simulations

Hybrid quantum mechanics/molecular mechanics (QM/MM) calculations were performed, based on the MD relaxed structures for all the redox states. The QM region, comprising *ca.* 120-130 atoms, included the Cu_B, and residues Tyr-244, His-290, His-291, His-240, Thr-316, Lys-319 as well as 12-15 water molecules. The QM region was described with the dispersion corrected B3LYP-D3 functional [212, 214] and def2-SVP basis set and def2-SVP/def2-TZVP(Cu) basis sets[186]. The remaining system (*ca.* 88600 atoms) was treated classically with classical CHARMM36 force field [227]. The QM/MM systems were minimized for 100 steps, followed by 5 ps of QM/MM molecular dynamics simulations at T=310 K, and using an integration step of 1 fs. The CHARMM/TURBOMOLE interface [289, 291] was employed for hybrid QM/MM calculations. For QM/MM reaction path optimizations, harmonic restraints were imposed on the linear combination of all water O-H bond-distances between Lys-319 and Tyr-244 (Q), where, $Q = \sum_i (r^i_{\mathrm{OH}} - r^{i+1}_{\mathrm{H...O}})$ with a force constant of k = 500 kcal mol^{-1} Å^{-2}. QM/MM umbrella sampling simulations were employed using harmonic potentials with force constant of k = 100 kcal mol^{-1} Å^{-2} along the reaction coordinate (Q) obtained from reaction path optimizations. The PMFs were computed using the weighted histogram analysis method [233]. The QM region was treated at dispersion corrected BP86-D3 level [203, 205, 214] using multipole accelerated resolution of identity approximation (RI-MARIJ) [287] and def2-SVP/def2-TZVP (Cu) basis sets [186].

7.3 Results and discussion

7.3.1 Lys-319 and water dynamics from MD simulations

Lys-319 in protonated form

The MD simulations suggest that water molecules in the K-channel provide hydrogen bonded protonic connectivity between Lys-319 and Tyr-244 via Thr-316. In the P_M state, the Lys-319 side-chain points down, away from Tyr-244 and towards the N-side of the membrane for the complete 500 ns simulation whereas, in the P_R state, the side-chain rapidly flips up towards Tyr-244 at the start of the simulation (Fig. 7.2A, 7.2B). A similar distribution of water occupancy in the K-channel with an average of two and four water molecules between Lys-319 and Tyr-244 was found in the P_M and P_R states, respectively (Fig. 7.3A, 7.3B). This increase of K-channel hydration upon reduction of BNC is accompanied with the up-flip of Lys-319, accounting for the transfer of charge in electrometry experiments by Lepp et al. [255]. Interestingly, in the reductive half of the redox cycle, where experiments suggest that the K-channel is employed for proton uptake of chemical protons, a two-state distribution of the Lys-319 side-chain conformation was observed for both, the O_H and the $O_{H,R}$ states with a very high frequency of flipping (Fig. 7.2A, 7.2B). In the $O_{H,R}$ state, an average of seven water molecules above Lys-319 were observed, whereas in the O_H state, a bimodal distribution was observed, with three and one water molecules on average corresponding to the “up” or “down” conformations of Lys-319 (Fig. 7.3A, 7.3B).

This implies that in the one-electron reduced BNC states, P_R and $O_{H,R}$, the K-channel is well hydrated (4-7 water molecules), regardless of the side-chain conformation of Lys-319. In contrast, with no electron at the BNC (P_M and O_H states), the cavity is hydrated or

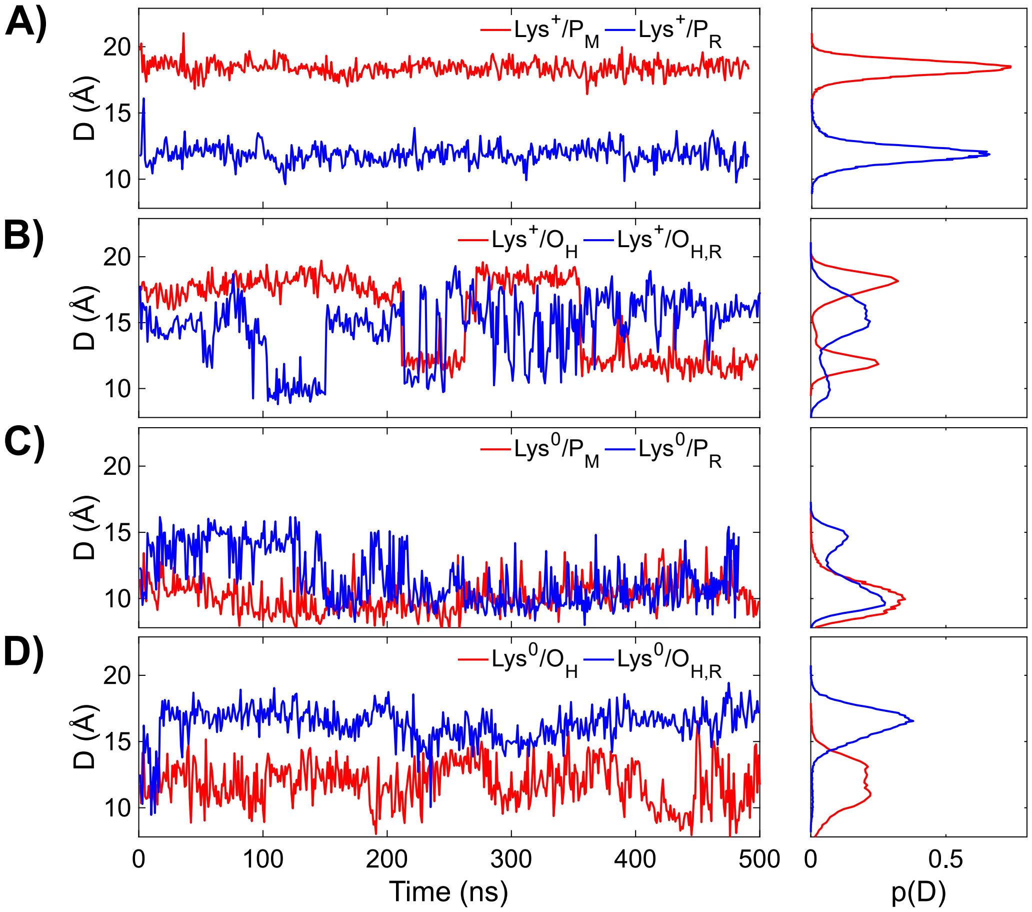

Figure 7.2: Dynamics of the Lys-319 side chain. The figure shows distances between Lys-319 head group and oxygen atom in the side-chain of Tyr-244 (D) from molecular dynamics simulations of P_M, P_R, O_H and $O_{H,R}$ states, respectively. Lys-319 was modelled in its protonated (Lys$^+$; panels A,B) and deprotonated (Lys^0; panels C,D) forms. Right panel shows probability density of the D for the respective catalytic states, and Lys-319 protonation states. The ”up” and ”down” conformations of Lys-319 correspond to small and large values of D, respectively.

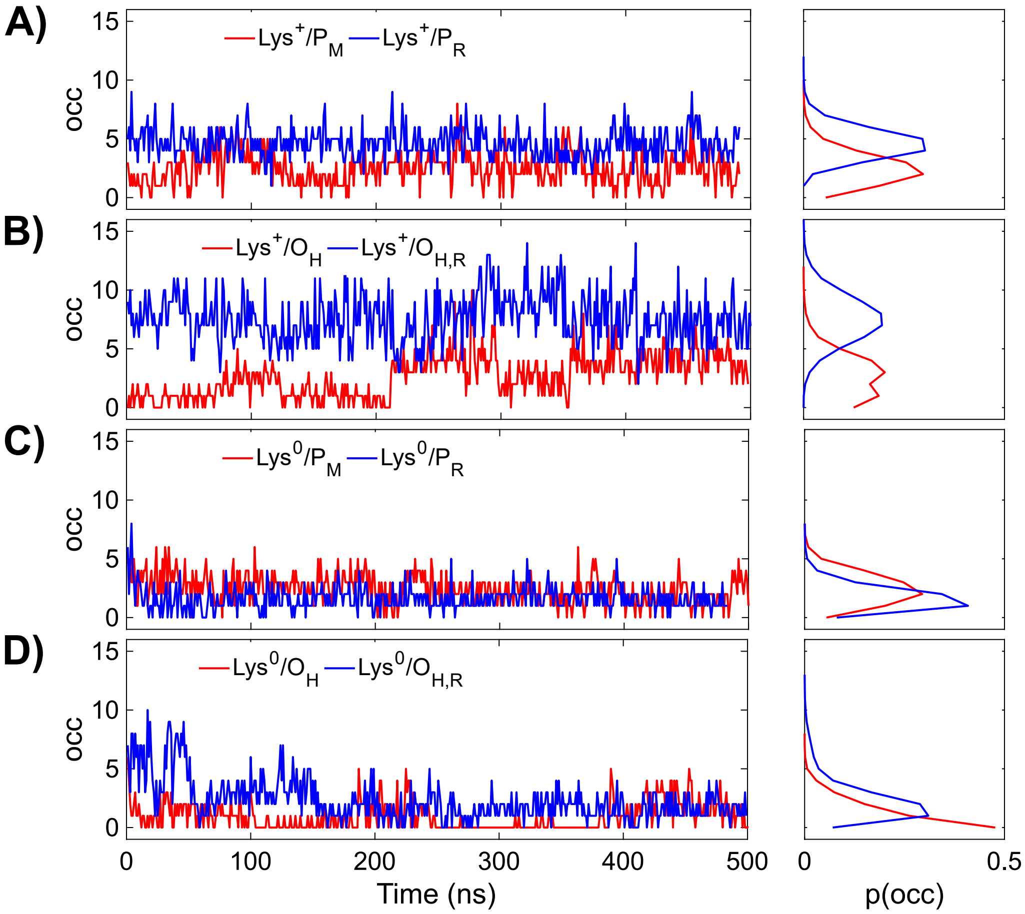

Figure 7.3: Water content in the K-channel between Lys-319 and Tyr-244. Water occupancy (occ) in the K-channel above Lys-319 from molecular dynamics simulations of P_M, P_R, O_H and $\text{O}_\text{H,R}$ states. Lys-319 was modelled in its protonated (Lys^+; panels A,B) and deprotonated (Lys^0; panels C,D) forms. Right panel shows probability density of the water content for the respective catalytic states, and Lys-319 protonation states.

dehydrated with Lys-319 in "up" or "down" configurations, respectively. Additionally, the bimodal Lys-319 side chain distributions in the reductive phase point to a lowering of side-chain barrier for Lys-319 flip, independent of the redox state of the BNC in O_H and $\text{O}_\text{H,R}$ states (Fig.7.4) from > 9.3 kcal mol^{-1} in P_M and P_R states (simulation timescales) to 1–2.5 kcal mol^{-1} in O_H and $\text{O}_\text{H,R}$ states. This could be attributed to structural alterations in the protein environment around Lys-319 as one progresses along the reaction cycle as previously also suggested by examining X-ray structures of fully reduced and fully oxidized C*c*O [314]. To probe such structural effects, helical motions around the K-channel in the 500 ns MD trajectories were analyzed. Bending of the surrounding helices in the O_H and $\text{O}_\text{H,R}$ states was observed in the MD simulation, making the K-channel wider and pulling in more water as compared to the P_M and P_R states (Fig. 7.5). To quantify the structural alterations in the helices surrounding the K-channel, helix RMSD (using backbone atoms) was evaluated for the MD trajectories of all the simulated states with Lys-319 in its protonated form (Fig. 7.6). Helix RMSD plot also suggest greater alterations around the K-channel in the O_H and $\text{O}_\text{H,R}$ states.

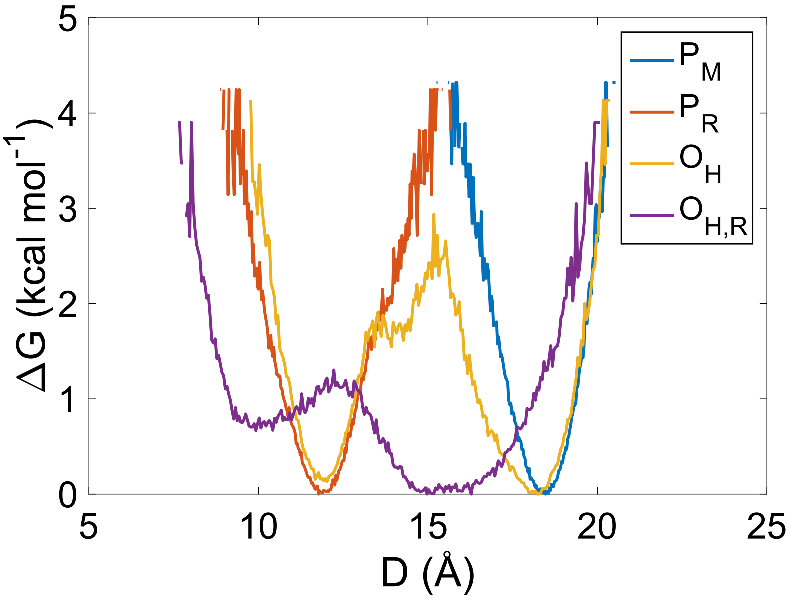

Figure 7.4: Free energy profiles for Lys-319 flipping dynamics as a function of the distance between Lys-319 head group and oxygen atom in the side-chain of Tyr-244 (D), from molecular dynamics simulations of P_M, P_R, OH and $O_{H,R}$ states with Lys-319 in its protonated form.

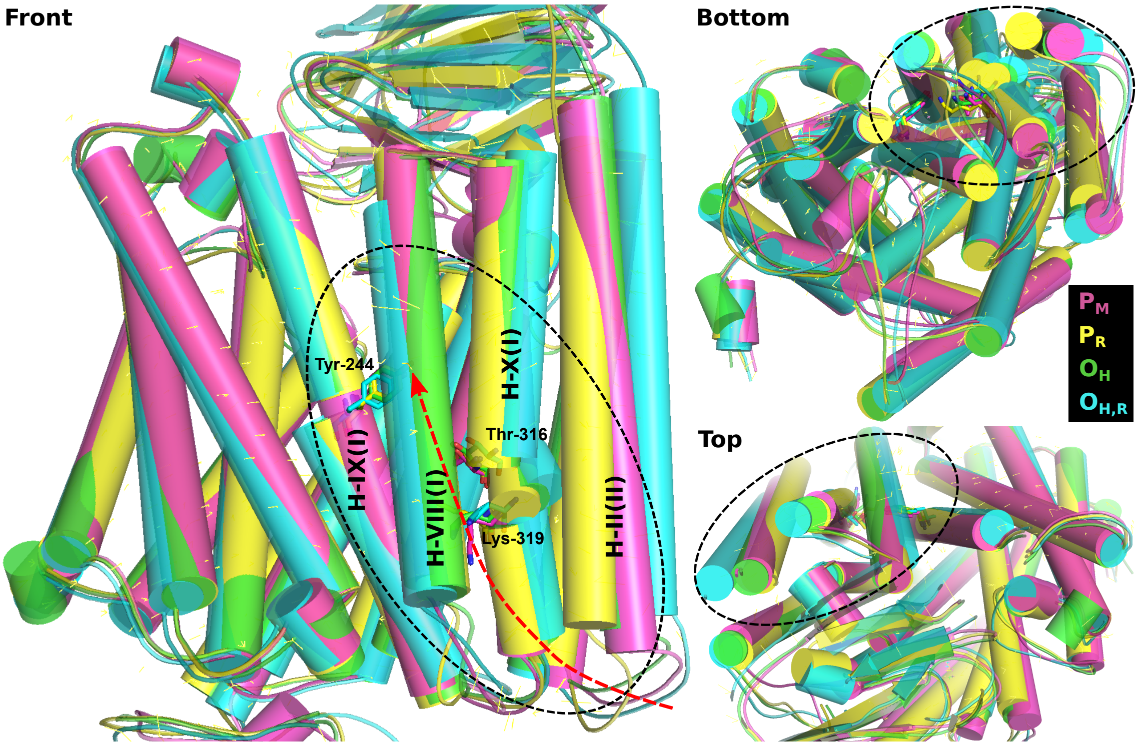

Figure 7.5: Averaged helix tilt in 500 ns MD simulations of P_M, P_R, O_H and $O_{H,R}$ states with Lys-319 in its protonated form. Helices VIII, IX, and X from subunit I and helix II from subunit II, which surround the K-channel, are labeled as H-VIII(I), H-IX(I), H-X(I), and H-II(II), respectively.

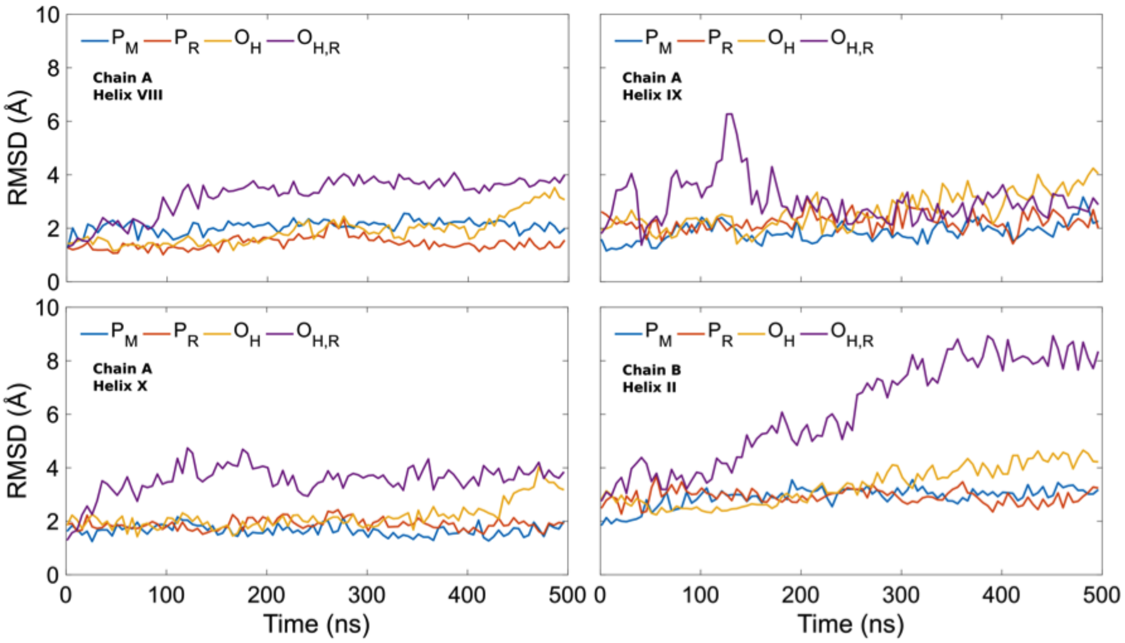

Figure 7.6: RMSD for helices VIII, IX, and X from subunit I and helix II from subunit II, which surround the K-channel, from 500 ns MD simulations of P_M, P_R, O_H and $O_{H,R}$ states with Lys-319 in its protonated form.

Lys-319 in deprotonated form

To also analyze the dynamics of a deprotonated Lys-319, 500 ns MD simulations were performed with Lys-319 in its neutral form in P_M, P_R, O_H and $O_{H,R}$ states starting with a 500 ns relaxed structure in P_R state with Lys-319 in flipped up state. In the simulations, the neutral Lys-319 flips down readily in one-electron reduced states (P_R and $O_{H,R}$), whereas there is an increase in the up-flipped population as compared to protonated Lys-319 simulations for states without an electron at the BNC (P_M and O_H) (Fig. 7.3C, 7.3D). There is also a reduction in water content across all four states with a sharper reduction in case of one-electron reduced states (P_R and $O_{H,R}$) (Fig. 7.3C, 7.3D). The emptying of cavity could indicate deactivation of the K-channel. Deprotonation of Lys-319, followed by a down-flip and quick reprotonation, akin to the Glu-242 ”valve” mechanism [99, 106, 149] could be a plausible mechanism for the role of Lys-319 residue in the K-channel.

7.3.2 Water-mediated protonic connectivities from MD simulations

In order to quantify the hydrogen bonded pathways, which are prerequisite to the proton transfer processes, the longest non-bonded connectivity along the shortest path connecting proton donor and proton acceptor, ζ, was evaluated, as a measure of the slowest or rate-limiting step in prospective proton transfer reactions (see section 5.2.2). Good hydrogen bonded connectivity was observed from Lys-319→Tyr-244 in P_R, O_H, and $O_{H,R}$ states but no connectivity was observed in the P_M state (Fig. 7.7). For the case of Glu-242→Cu_B, the connectivity exists only in P_M and P_R states, whereas, there is no connectivity for O_H and $O_{H,R}$ states (Fig. 7.7). For analogous one-electron BNC reduced states, P_R and $O_{H,R}$, the distribution along the D-channel paths (Glu-242 → Cu_B) and the K-channel paths (Lys-319 → Tyr-244) strongly differ. The ζ distributions suggests that there is always water connectivity to BNC via K-channel, but only in the oxidative phase for Glu-242 → BNC. In the oxidative phase, when both D- and K-channels have connectivity to BNC, the chemical proton is transferred from the kinetically favored D-channel, whereas, in the

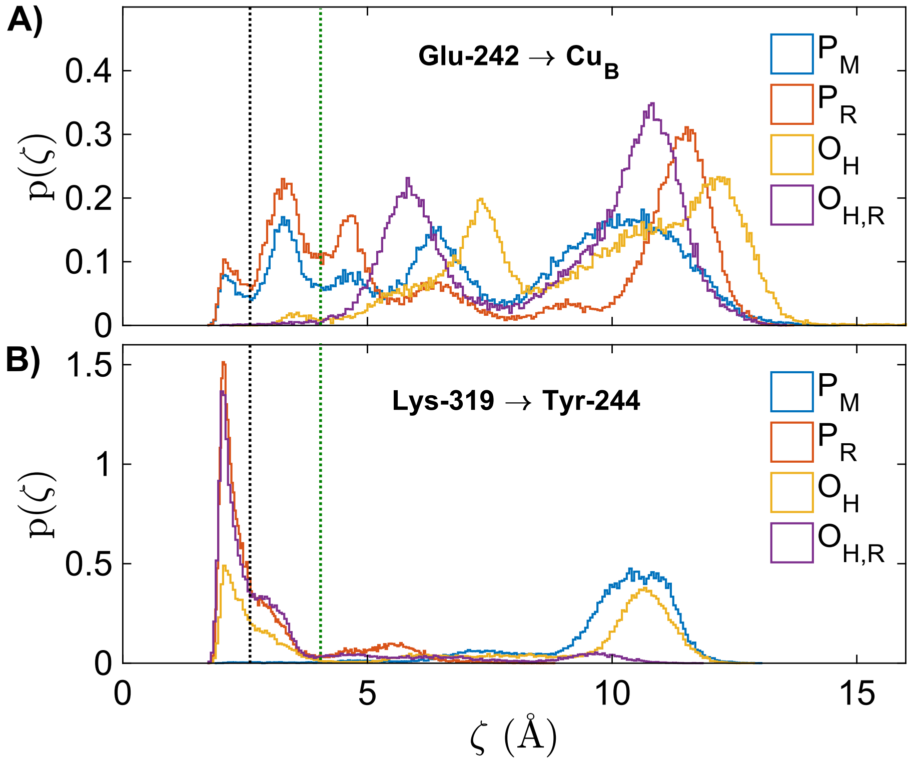

Figure 7.7: Longest non–bonded connectivity along the shortest path connecting proton donor and proton acceptor, ζ, for proton transfer to BNC. ζ was evaluated for protonic connectivity between A) Glu-242 (D-channel), and B) Lys-244 (K-channel) to the BNC for P_M, P_R, O_H, and $O_{H,R}$ states with Glu-242 and Lys-319 modeled in their protonated forms. ζ density to the left of the black dotted line ($\zeta < 2.5$) indicates hydrogen bonded configurations and the density to the left of $\zeta < 4.0$ (green dotted) indicates water configurations which may provide protonic connectivities (see section 5.3.1).

reductive phase with no connectivity from D-channel, the chemical proton can only be taken from the K-channel, thus triggering its activation. Moreover, the high probability density of forming hydrogen bonded pathways from Lys-319 to Tyr-244 in the O_H is due to high levels of hydration. The high density for $\zeta < 4.0$Å in the case of O_H state is when Lys-319 is flipped down, facing away from the active site (Fig. 7.8). Thus, the ζ distributions indicate that the absence of water-mediated protonic connectivity from the D-channel towards BNC in the reductive phase could activate the K-channel for proton transfer in the reductive phase. The high amount of water influx might indicate the activation of K-channel, but the proton transfer is kinetically favored only in the $O_{H,R}$ state.

7.3.3 Proton transfer energetics from DFT calculations

To probe the energetics of activation and gating of the K-channel, cluster models of *ca.* 120-160 atoms for P_R and $O_{H,R}$ states were prepared from relaxed MD structures for quantum chemical density functional theory (DFT) treatment. The proton affinity of Tyr-244 was examined in P_R and $O_{H,R}$ states by transferring a proton from Lys-319 to Tyr-244 via its tight hydrogen bonded water network, followed by reoptimization of the structures. The structures were based on MD simulations and were selected for optimum hydrogen bonded connectivity using the ζ measure, as described before. In the P_R and $O_{H,R}$ states, Tyr-244 residue had proton affinities of 3.3 and 4.8 kcal mol^{-1}, respectively. This indicates better

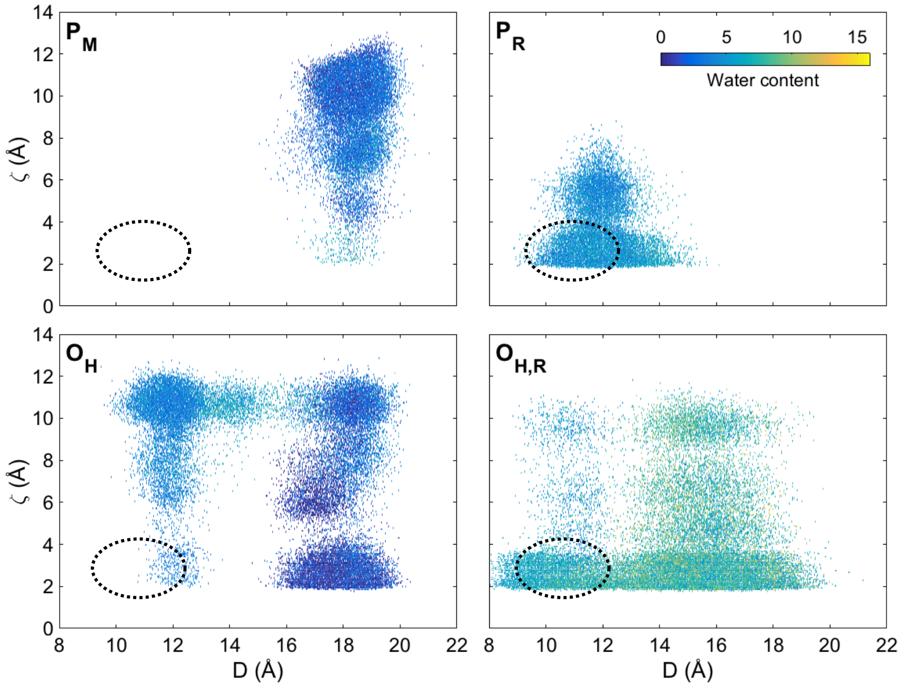

Figure 7.8: Distance vs ζ color-coded with water occupancy in the K-channel. This figure shows the distance between Lys-319 head group and oxygen atom in the side-chain of Tyr-244, D, against the longest non–bonded connectivity along the shortest path connecting the proton donor and the proton acceptor, ζ. Water occupancy is color-coded with blue for no water in the K-channel above Lys-319 head group and yellow for overfilled K-channel (top right panel; see color bar). Each data point corresponds the three properties described above and represents a snapshot along 500 ns MD trajectories of P_M, P_R, O_H and $O_{H,R}$ states of C*c*O. Dotted circles indicate regions on the plot suitable for proton transfer from the K-channel from Lys-319 to Tyr-244.

stabilization the proton on Tyr-244 in $O_{H,R}$ state as compared to P_R state, consistent with FTIR experiments [68]. The intrinsic proton affinity of all protonable sites in BNC (heme a_3, Cu_B and Tyr-244) was also calculated in the P_R and $O_{H,R}$ states. The intrinsic proton affinity of heme a_3, Cu_B and Tyr-244 in the P_R were found to be 14.2, 23.0 and 18.6 kcal mol^{-1}, respectively. Whereas, in the $O_{H,R}$ state, intrinsic proton affinities of 21.2 and 20.3 kcal mol^{-1} were obtained for heme a_3 and Tyr-244, respectively. The high proton affinity of Cu_B in the oxidative half, combined with ζ analysis of pathways suggests a possible explanation as to why D-channel is used for protonation of BNC in the oxidative phase. In the $O_{H,R}$ state, the elevated proton affinity of Tyr-244 coupled to the excellent hydrogen bonded connectivity between Lys-319 and Tyr-244 and no protonic connectivity between Glu-242 and Tyr-244, could explain the switch from the D-channel to the K-channel as the proton source for BNC protonation, going from oxidative to reductive phase of the reaction cycle (Fig. 7.1B).

7.3.4 Proton transfer to Tyr-244 (QM/MM)

QM/MM reaction path optimizations in a native-like environment were performed to probe the energetics of the proton transfer from Lys-319 to Tyr-244. The barriers for the proton transfer in the P_R and the $O_{H,R}$ states were in the range of 21 -24 kcal mol^{-1} with greater stabilization for protonated Tyr-244 in $O_{H,R}$ state. In order to estimate the free energy profile for proton-transfer, QM(DFT)/MM-MD umbrella sampling calculations were performed along reaction coordinates based on the structures obtained from QM/MM reaction path optimizations. The umbrella sampling yielded free energy barriers of 14.0 and 12.3 kcal mol-1 and free energies of 12.1 and 8.9 kcal mol-1 for the proton transfer from Lys-319 to Tyr-244 in P_R and $O_{H,R}$ states, respectively (Fig. 7.9). The free energy barriers of 14.0 and 12.3 kcal mol-1 would correspond to timescales of *ca.* 100 μs and 1 ms using Eyring's equation with a standard pre-exponential factor, for the P_R and the $O_{H,R}$ states, respectively, which is consistent with experimentally derived timescale of *ca.* 13 kcal mol-1 [315]. This 2 kcal mol-1 lowering of the barrier and additional 4 kcal mol-1 stabilization of the proton at Tyr-244, upon moving between isoelectric states, P_R to $O_{H,R}$ could play a key role, in addition to the kinetic hindrance from the D-channel, as observed in ζ analysis (Fig. 7.7). A possible reason for the increase in the proton affinity in the $O_H \rightarrow O_{H,R}$ step is that Cu_B site in the O_H state exits in form: [Cu^{II}-H_2O TyrO$^\bullet$ $\leftrightarrow$ Cu^I-H_2O TyrO$^-$], with the electron equilibrated between Cu^{II}/Cu^I and TyrO$^\bullet$/TyrO$^-$, which stabilizes into TyrO$^-$ upon reduction of the BNC, as suggested in previous DFT calculations [67].

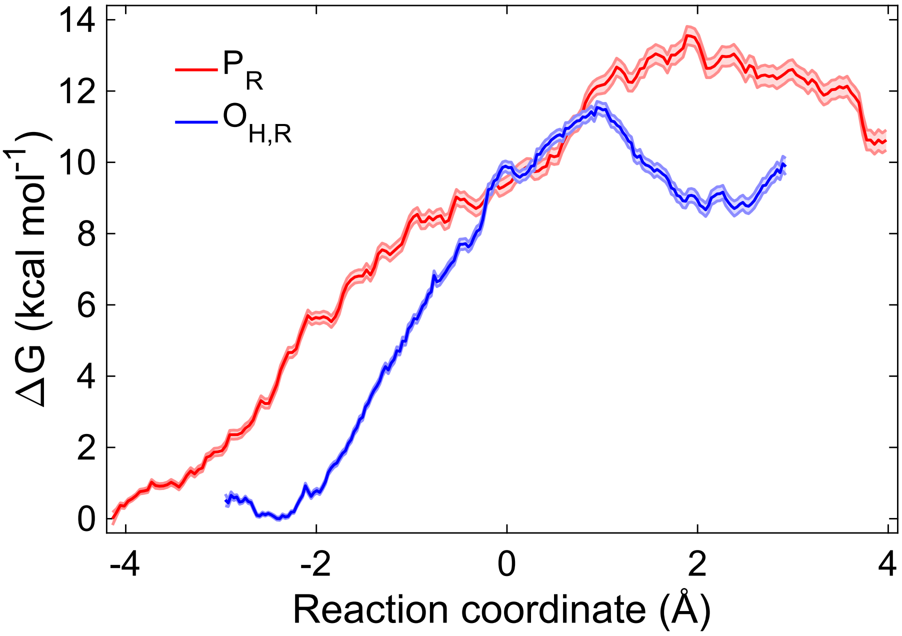

Figure 7.9: Free energy profiles of proton transfer. QM/MM-MD free energy profiles for proton transfer from Lys-319 to Tyr-244 in P_R and $O_{H,R}$ states. Lighter traces indicate the error bounds for the free energy profiles. The profiles were obtained using Umbrella sampling technique for free energy evaluation at with reaction coordinate, $Q = \sum_i (r^i_{\mathrm{OH}} - r^{i+1}_{\mathrm{H...O}})$, where negative coordinate values on the left indicate protonated Lys-319 and positive values on the right represent a proton transferred to Tyr-319.The calculations were performed BP86-D3 level [203, 205, 214] using multipole accelerated resolution of identity approximation (RI-MARIJ) [287] and def2-SVP/def2-TZVP (Cu) basis sets [186].

7.3.5 Lys-319 pK_a calculations (continuum electrostatics)

Continuum electrostatics models were employed to calculate pK_a values of Lys-319 in the P_M, P_R, O_H and $O_{H,R}$ states, based on structures from 500 ns MD trajectories. Although the pK_a values exhibit a large variability, a clear trend emerges in O_H state: Lys-319 is deprotonated when Lys-319 is in "up" state suggesting a proton release from Lys-319 to the water molecules in the hydrated channel (Fig. 7.10). Previous pK_a studies [312] had suggested that Lys-319 exists in a deprotonated form in all states except the P_R state. The pK_a calculations in this study suggest that Lys-319 remains predominantly protonated throughout the redox cycle. Additionally, in order to decouple the dependence of structure and redox states' dependence on pK_a values, the redox states of the P_M and P_R states, and, the O_H and $O_{H,R}$ states, were exchanged, respectively. Swapping the redox states led to the pK_a values shift by more than *ca.* 5 pK units with Lys-319 in the "up" conformation compared to only by *ca.* 1 pK units in the "down" conformation (Fig. 7.11), further illustrating extreme sensitivity of classical pK_a values obtained.

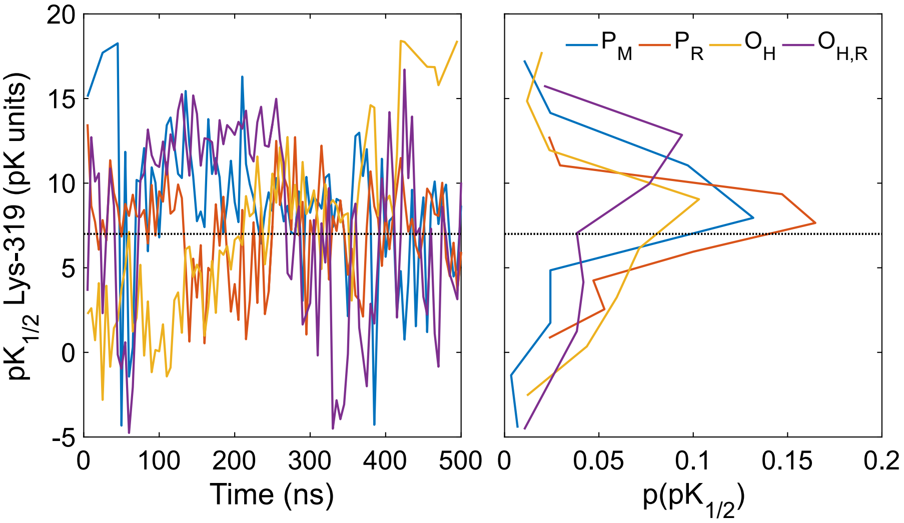

Figure 7.10: Lys-319 pK_a from MD trajectories. pK_a of Lys-319 residue was evaluated for snapshots, every 5 ns, from 500 ns MD trajectories of P_M, P_R, O_H, $O_{H,R}$ states using continuum electrostatics. Left panel shows pK_a evolution in time and right panel shows distribution of Lys-319 pK_a over the trajectory.

The pK_a calculations suggest that a flipped-up Lys-319, along with hydrogen bonded protonic connectivity to BNC and an elevated proton affinity of Tyr-244 are all prerequisites for proton-transfer from K-channel to the BNC. This work suggests that the $O_{H,R}$ state satisfies these conditions at classical and quantum chemical level.

7.4 Summary

Classical MD simulations suggest that Lys-319 flips "up" in the states with BNC reduced, in accordance with electrometric measurements. Structures from MD simulations were analyzed, and the energetics and dynamics of the proton transfer in K-channel were examined using quantum chemical calculations in order to explain why the K-channel is employed in the reductive phase, while the enzyme uses the D-channel to uptake chemical protons in the oxidative half. The calculations suggest that K-channel is extensively more hydrated and also undergoes structural changes in the reductive phase (O states: O_H and $O_{H,R}$ states) as compared to the oxidative phase (P states: P_M and P_R states). Pathway analysis (ζ) shows that while there is a prominent hydrogen bonded connectivity from both, the D- and the K-channels to the BNC in the P-states, the connectivity from the D-channel is lost in the O-states, thus causing a switch from the D-channel to the K-channel for proton delivery to the BNC. Proton affinities from QM calculations and QM/MM free energy profiles for the proton transfer also indicate an increase in the proton affinity and

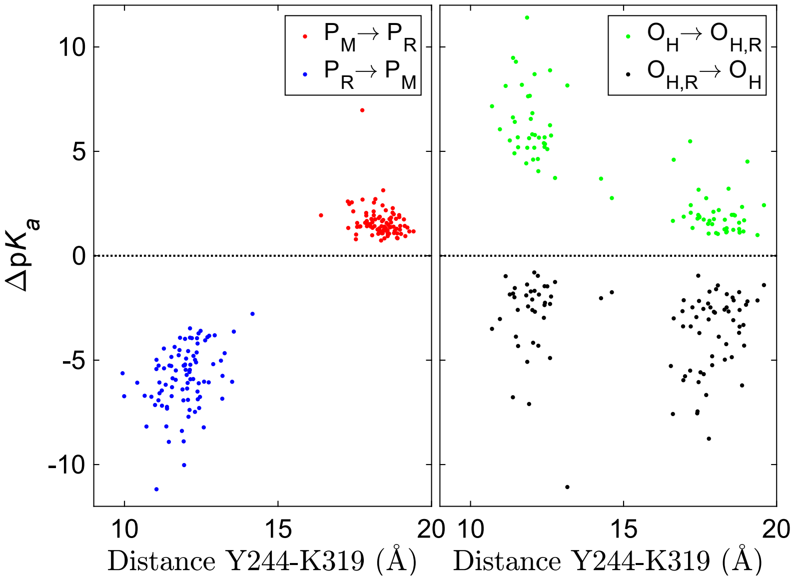

Figure 7.11: Lys-319 pK_a perturbation. pK_as of Lys-319 were evaluated using continuum electrostatics by swapping consecutive redox states($P_M \leftrightarrow P_R$, $O_H \leftrightarrow O_{H,R}$) for the 5 ns snapshot structures obtained from 500 ns of MD simulations for each state. For eg., $P_M \rightarrow P_R$ indicates the change in pK_a of Lys-319 for structures from MD simulation of P_M state, as a result of changing electrostatic charges to P_R state. Distance Y244-K319 label indicates the distance between the head groups of Tyr-244 and Lys-319 residues for the evaluated MD snapshot.

a reduction of the proton transfer barrier upon formation of the $O_{H,R}$ state as compared to the P_R state, favoring transfer of a proton to the BNC in the $O_{H,R}$ state.

8 Conclusions

C*c*O is a redox driven proton pumping enzyme that harnesses the energy of O_2 reduction to water, and utilizes it to pump protons across the mitochondrial inner membrane in eukaryotes or the cytoplasmic membrane in aerobic bacteria. Despite decades of structural, biochemical, and biophysical research, the exact molecular mechanism employed by C*c*O to catalyze the coupled proton-electron transfer reactions remains elusive.

In this dissertation, multi-scale molecular simulations were employed to study water dynamics in the active site, to probe mechanistic aspects of the pumping machinery and to examine energetics and dynamics of the proton-electron coupled reactions in C*c*O. Dynamics of bovine C*c*O, embedded in a native-like membrane-ion-water environment was studied on μs timescales in different redox states, mimicking key reaction steps in the catalytic cycle of C*c*O by classical full atomistic molecular dynamics simulations. The Poisson-Boltzmann (PB) equation was evaluated using continuum electrostatics model and Monte Carlo sampling to probe protonation changes of key catalytic residues in C*c*O. DFT calculations at both QM cluster and QM/MM levels, were used to explore the proton transfer energetics and to examine functional couplings of key intermediates involved the proton pumping machinery of C*c*O.

Dijkstra's algorithm was implemented to analyze water dynamics along the protons transfer paths in the active site of C*c*O. Redox sensitive water pathways were observed, with enhanced connectivity to the pump site and the oxygen reduction site in states with heme *a* reduced and with BNC reduced, respectively. The pathway analysis suggested that short-circuit in the pumping reaction might not be prevented by the water-orientational effects alone, since a statistically relevant number of pathways connecting to the oxygen reduction site were also observed prior to the reduction of the BNC. Electric field effects in the active were also examined and are likely to provide the physical basis for water orientations in the active as observed in the MD simulations.

The identity of the transient proton-loading site (PLS) and proton transfer to the PLS were examined using QM and QM/MM methods based on equilibrated structures from MD simulations of different redox states of the enzyme. The QM and QM/MM calculations suggested that a structurally conserved water cluster above the D-propionate group of heme a_3 might serve as a transient proton acceptor site before the proton is ejected across the membrane. The proton affinity of the identified water cluster was also found to be sensitive to the redox states of the enzyme. Barriers for proton transfer to the PLS were found to be modulated by the redox state of the heme *a* site, with lowered barriers when heme *a* is reduced and increased barriers upon transfer of the electron from heme *a* to the BNC. Thus, heme *a* can serve as the gating element, facilitating PLS protonation upon heme *a* reduction, and preventing the leakage of the protons from the PLS towards the thermodynamically favored direction, once the electron has been transferred to the BNC. IR spectra were also computed from QM/MM MD trajectories and a loss in the characteristic "dangling-waters" signal was observed upon protonation of the PLS, with striking similarities to water signals observed in the light-driven proton pump, bacteriorhodopsin.

In the third part of the thesis, the role and activation of the proton-conducting K-channel was examined using classical molecular dynamics simulations and PB continuum methods, as well as QM and QM/MM methodologies, modeling the dynamics of C*c*O in the oxidative and reductive halves of its catalytic cycle. In the MD simulations, Lys-319, a conserved residue in the K-channel, was observed to flip "up" towards the active site upon reduction of the BNC. Also, structural changes and additional influx of water in the K-channel was observed in the reductive phase, O→E transition. The QM calculations suggested an increase in the proton affinity of Tyr-244 for states with a reduced BNC in the reductive phase, possibly due to an electronic equilibration between $\mathrm{Cu^{II/I}}$ and $\mathrm{TyrO^{\bullet/-}}$ forms of the BNC. The elevation of the proton affinity of Tyr-244, along with lack of water-mediated proton transfer connectivities from Glu-242 to the BNC in the reductive phase, and lowered free energy barriers of proton transfer along the K-channel based on QM/MM free energy calculations, indicate possible mechanisms that trigger the activation of the K-channel in the reductive phase of the catalytic cycle.

In future work, it would be important to understand mechanistic differences between different members of the HCO superfamily in order to derive a minimal machinery necessary to drive the redox-driven proton pumping.

This work has shown that multi-scale molecular simulations can provide valuable insight into the complex catalytic machinery of the trans-membrane enzyme, cytochrome *c* oxidase. By studying the structure, dynamics, and energetics of this enzyme on a broad range of time-scales, a better mechanistic understanding, complementary to many experiments has been obtained. The results of the protonated water cluster can stimulate new experiments, particularly FTIR spectroscopic studies, and the proposed activation mechanism for the K-channel might provide insights into designing new site-directed mutagenesis experiments.

Publications

Publications pertinent to this thesis:

- A Protonated Water Cluster as a Transient Proton-Loading Site in Cytochrome *c* Oxidase. Supekar, S., Gamiz-Hernandez, A. P., Kaila V. R. I., Angewandte Chemie International Edition (2016), 128(39), 12119-12123.
- Activation of the K-channel in Cytochrome *c* Oxidase. Supekar, S., Kaila, V. R. I. (manuscript)

Other publications :

- Distribution of residence time of water around DNA base pairs: governing factors and the origin of heterogeneity. Saha, D.*, Supekar, S.*, & Mukherjee, A., The Journal of Physical Chemistry B (2015), 119(34), 11371-11381.
- Conformational Selection Mechanism of Dimethylarginine by the Tudor Domain of the Human Survival Motor Neuron Protein. Supekar, S., Papageorgiou, A. C., Gemmecker, G., Peltzer, R., Johansson, M.P., Tripsianes K., Sattler, M., and Kaila, V. R. I. (Angewandte Chemie 2017; accepted; doi: 10.1002/anie.201708233 and 10.1002/ange.201708233)

Bibliography

[1] P. M. Hurley, *How old is the earth* (Anchor books doubleday, 1959).

[2] S. M. Awramik, “The oldest records of photosynthesis”, Photosynth. Res. **33**, 75–89 (1992).

[3] E. A. Bell, P. Boehnke, T. M. Harrison, and W. L. Mao, “Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon”, Proc. Natl. Acad. Sci. U.S.A. **112**, 14518–14521 (2015).

[4] R. Dawkins, *The selfish gene* (Oxford University Press, 1976).

[5] H. G. Hill and J. A. Nuth, “The catalytic potential of cosmic dust: implications for prebiotic chemistry in the solar nebula and other protoplanetary systems”, Astrobiology **3**, 291–304 (2003).

[6] H. Clemmey and N. Badham, “Oxygen in the precambrian atmosphere: an evaluation of the geological evidence”, Geology **10**, 141–146 (1982).

[7] V. Sojo, B. Herschy, A. Whicher, E. Camprubí, and N. Lane, “The origin of life in alkaline hydrothermal vents”, Astrobiology **16**, 181–197 (2016).

[8] T. W. Lyons, C. T. Reinhard, and N. J. Planavsky, “The rise of oxygen in earth’s early ocean and atmosphere”, Nature **506**, 307–315 (2014).

[9] D. G. Nicholls, *Bioenergetics* (Academic press, 2013).

[10] W. F. Martin, S. Garg, and V. Zimorski, “Endosymbiotic theories for eukaryote origin”, Phil. Trans. R. Soc. B **370**, 20140330 (2015).

[11] M. Saraste, “Oxidative phosphorylation at the *fin de siècle*”, Science **283**, 1488–1493 (1999).

[12] D. C. Wallace, W. Fan, and V. Procaccio, “Mitochondrial energetics and therapeutics”, Annu. Rev. Pathol. **5**, 297–348 (2010).

[13] U. Brandt, “Energy converting NADH: quinone oxidoreductase (complex I)”, Annu. Rev. Biochem. **75**, 69–92 (2006).

[14] M. Wikström, “Two protons are pumped from the mitochondrial matrix per electron transferred between NADH and ubiquinone”, FEBS Lett. **169**, 300–304 (1984).

[15] R. G. Efremov, R. Baradaran, and L. A. Sazanov, “The architecture of respiratory complex I”, Nature **465**, 441–445 (2010).

[16] G. Cecchini, “Function and structure of complex II of the respiratory chain”, Annu. Rev. Biochem. **72**, 77–109 (2003).

[17] P. Mitchell, “Possible molecular mechanisms of the protonmotive function of cytochrome systems”, J. Theor. Biol. **62**, 327–367 (1976).

[18] A. R. Crofts, “Proton-coupled electron transfer at the Q_o-site of the bc_1 complex controls the rate of ubihydroquinone oxidation”, Biochim. Biophys. Acta, Bioenerg. **1655**, 77–92 (2004).

[19] A. R. Crofts, S. Hong, C. Wilson, R. Burton, D. Victoria, C. Harrison, and K. Schulten, “The mechanism of ubihydroquinone oxidation at the Q_o-site of the cytochrome bc_1 complex”, Biochim. Biophys. Acta, Bioenerg. **1827**, 1362–1377 (2013).

[20] S. Ferguson-Miller and G. T. Babcock, “Heme/copper terminal oxidases”, Chem. Rev. **96**, 2889–2908 (1996).

[21] P. Brzezinski and R. B. Gennis, “Cytochrome *c* oxidase: exciting progress and remaining mysteries”, J. Bioenerg. Biomembr. **40**, 521–531 (2008).

[22] V. R. I. Kaila, M. I. Verkhovsky, and M. Wikström, “Proton-coupled electron transfer in cytochrome oxidase”, Chem. Rev. **110**, 7062–81 (2010).

[23] S. Yoshikawa and A. Shimada, “Reaction mechanism of cytochrome *c* oxidase”, Chem. Rev **115**, 1936–1989 (2015).

[24] P Mitchell, “Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism.”, Nature **191**, 144–148 (1961).

[25] K. Okazaki and G. Hummer, “Phosphate release coupled to rotary motion of F_1-ATPase.”, Proc. Natl. Acad. Sci. U.S.A. **110**, 16468–16473 (2013).

[26] J. E. Walker, “The ATP synthase: the understood, the uncertain and the unknown.”, Biochem. Soc. Trans. **41**, 1–16 (2013).

[27] D. G. Nicholls and R. M. Locke, “Thermogenic mechanisms in brown fat.”, Physiol. Rev. **64**, 1–64 (1984).

[28] M. P. Murphy, “Slip and leak in mitochondrial oxidative phosphorylation.”, Biochim. Biophys. Acta **977**, 123–141 (1989).

[29] C. W. Hoganson, M. A. Pressler, D. A. Proshlyakov, and G. T. Babcock, “From water to oxygen and back again: mechanistic similarities in the enzymatic redox conversions between water and dioxygen.”, Biochim. Biophys. Acta **1365**, 170–174 (1998).

[30] F. Westheimer, “The mechanisms of chromic acid oxidations.”, Chem. Rev. **45**, 419–451 (1949).

[31] J. M. Mayer, “Understanding hydrogen atom transfer: from bond strengths to marcus theory”, Acc. Chem. Res. **44**, 36–46 (2010).

[32] S. Hammes-Schiffer, “Proton-coupled electron transfer: moving together and charging forward”, J. Am. Chem. Soc. **137**, PMID: 26110700, 8860–8871 (2015).

[33] M. Verkhovskaya and D. A. Bloch, “Energy-converting respiratory complex I: on the way to the molecular mechanism of the proton pump.”, Int. J. Biochem. Cell Biol. **45**, 491–511 (2013).

[34] A. Stuchebrukhov, “Long-distance electron tunneling in proteins: a new challenge for time-resolved spectroscopy”, Laser Phys. **20**, 125–138 (2010).

[35] C. C. Moser, P. L. Dutton, et al., “Nature of biological electron transfer”, Nature **355**, 796 (1992).

[36] V. P. Skulachev, P. C. Hinkle, and P. Mitchell, *Chemiosmotic proton circuits in biological membranes* (Addison-Wesley, 1981).

[37] M. K. Wikström, “Proton pump coupled to cytochrome *c* oxidase in mitochondria”, Nature **266**, 271–3 (1977).

[38] P. Brzezinski and G. Larsson, “Redox-driven proton pumping by heme-copper oxidases”, Biochim. Biophys. Acta, Bioenerg. **1605**, 1–13 (2003).

[39] F. L. Sousa, R. J. Alves, J. B. Pereira-Leal, M. Teixeira, and M. M. Pereira, "A bioinformatics classifier and database for heme-copper oxygen reductases.", PLoS One **6**, e19117 (2011).

[40] F. L. Sousa, R. J. Alves, M. A. Ribeiro, J. B. Pereira-Leal, M. Teixeira, and M. M. Pereira, "The superfamily of heme–copper oxygen reductases: types and evolutionary considerations", Biochim. Biophys. Acta, Bioenerg. **1817**, 629–637 (2012).

[41] H.-Y. Chang, J. Hemp, Y. Chen, J. A. Fee, and R. B. Gennis, "The cytochrome ba_3 oxygen reductase from *Thermus thermophilus* uses a single input channel for proton delivery to the active site and for proton pumping.", Proc. Natl. Acad. Sci. U.S.A. **106**, 16169–16173 (2009).

[42] J. Hemp, H. Han, J. H. Roh, S. Kaplan, T. J. Martinez, and R. B. Gennis, "Comparative genomics and site-directed mutagenesis support the existence of only one input channel for protons in the C-family (cbb_3 oxidase) of heme- copper oxygen reductases", Biochemistry **46**, 9963–9972 (2007).

[43] H. J. Lee, J. Reimann, Y. Huang, and P. Adelroth, "Functional proton transfer pathways in the heme-copper oxidase superfamily.", Biochim. Biophys. Acta **1817**, 537–544 (2012).

[44] V. Srinivasan, C. Rajendran, F. L. Sousa, A. M. Melo, L. M. Saraiva, M. M. Pereira, M. Santana, M. Teixeira, and H. Michel, "Structure at 1.3 Å resolution of *Rhodothermus marinus* caa_3 cytochrome *c* domain", J. Mol. Biol. **345**, 1047–1057 (2005).

[45] T. Hino, Y. Matsumoto, S. Nagano, H. Sugimoto, Y. Fukumori, T. Murata, S. Iwata, and Y. Shiro, "Structural basis of biological N_2O generation by bacterial nitric oxide reductase", Science **330**, 1666–1670 (2010).

[46] T. Soulimane, G. Buse, G. P. Bourenkov, H. D. Bartunik, R. Huber, and M. E. Than, "Structure and mechanism of the aberrant ba_3-cytochrome *c* oxidase from *Thermus thermophilus*", EMBO J. **19**, 1766–1776 (2000).

[47] T Tsukihara, H Aoyama, E Yamashita, T Tomizaki, H Yamaguchi, K Shinzawa-Itoh, R Nakashima, R Yaono, and S Yoshikawa, "The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å", Science **272**, 1136–1144 (1996).

[48] V. Rauhamäki, M. Baumann, R. Soliymani, A. Puustinen, and M. Wikström, "Identification of a histidine-tyrosine cross-link in the active site of the cbb_3-type cytochrome *c* oxidase from *Rhodobacter sphaeroides*", Proc. Natl. Acad. Sci. U.S.A. **103**, 16135–16140 (2006).

[49] S. Yoshikawa, K. Shinzawa-Itoh, R. Nakashima, R. Yaono, E. Yamashita, N. Inoue, M. Yao, M. J. Fei, C. P. Libeu, T. Mizushima, et al., "Redox-coupled crystal structural changes in bovine heart cytochrome *c* oxidase", Science **280**, 1723–1729 (1998).

[50] T. Tsukihara, K. Shimokata, Y. Katayama, H. Shimada, K. Muramoto, H. Aoyama, M. Mochizuki, K. Shinzawa-Itoh, E. Yamashita, M. Yao, Y. Ishimura, and S. Yoshikawa, "The low-spin heme of cytochrome *c* oxidase as the driving element of the proton-pumping process", Proc. Natl. Acad. Sci. U. S. A. **100**, 15304–9 (2003).

[51] C Ostermeier, A Harrenga, U Ermler, and H Michel, "Structure at 2.7 å resolution of the *Paracoccus denitrificans* two-subunit cytochrome *c* oxidase complexed with an antibody FV fragment.", Proc. Natl. Acad. Sci. U.S.A. **94**, 10547–10553 (1997).

[52] J. Koepke, E. Olkhova, H. Angerer, H. Müller, G. Peng, and H. Michel, "High resolution crystal structure of *Paracoccus denitrificans* cytochrome *c* oxidase: new insights into the active site and the proton transfer pathways", Biochim. Biophys. Acta, Bioenerg. **1787**, 635–645 (2009).

[53] J. Liu, L. Qin, and S. Ferguson-Miller, "Crystallographic and online spectral evidence for role of conformational change and conserved water in cytochrome oxidase proton pump", Proc. Natl. Acad. Sci. U. S. A. **108**, 1284–9 (2011).

[54] M. Svensson-Ek, J. Abramson, G. Larsson, S. Tornroth, P. Brzezinski, and S. Iwata, "The X-ray crystal structures of wild-type and EQ(I-286) mutant cytochrome *c* oxidases from *Rhodobacter sphaeroides*", J. Mol. Biol. **321**, 329–39 (2002).

[55] J. A. Lyons, D. Aragão, O. Slattery, A. V. Pisliakov, T. Soulimane, and M. Caffrey, "Structural insights into electron transfer in caa_3-type cytochrome oxidase", Nature **487**, 514–518 (2012).

[56] B. Liu, Y. Zhang, J. T. Sage, S. M. Soltis, T. Doukov, Y. Chen, C. D. Stout, and J. A. Fee, "Structural changes that occur upon photolysis of the $\mathrm{Fe(II)}_{a3}$–CO complex in the cytochrome ba_3-oxidase of *Thermus thermophilus*: a combined X-ray crystallographic and infrared spectral study demonstrates co binding to Cu_B", Biochim. Biophys. Acta, Bioenerg. **1817**, 658–665 (2012).

[57] S. Buschmann, E. Warkentin, H. Xie, J. D. Langer, U. Ermler, and H. Michel, "The structure of cbb_3 cytochrome oxidase provides insights into proton pumping", Science **329**, 327–330 (2010).

[58] B. Kadenbach, M. Hüttemann, S. Arnold, I. Lee, and E. Bender, "Mitochondrial energy metabolism is regulated via nuclear-coded subunits of cytochrome *c* oxidase", Free Radical Biol. Med. **29**, 211–221 (2000).

[59] R. A. Capaldi, "Structure and function of cytochrome *c* oxidase.", Annu. Rev. Biochem. **59**, 569–596 (1990).

[60] R. A. Capaldi, "Structure and assembly of cytochrome *c* oxidase.", Arch. Biochem. Biophys. **280**, 252–262 (1990).

[61] D. Pierron, D. E. Wildman, M. Hüttemann, G. C. Markondapatnaikuni, S. Aras, and L. I. Grossman, "Cytochrome *c* oxidase: evolution of control via nuclear subunit addition.", Biochim. Biophys. Acta **1817**, 590–597 (2012).

[62] S. Yoshikawa, K. Muramoto, K. Shinzawa-Itoh, and M. Mochizuki, "Structural studies on bovine heart cytochrome *c* oxidase.", Biochim. Biophys. Acta **1817**, 579–589 (2012).

[63] M. F. Tweedle and L. J. Wilson, "Electronic state of heme in cytochrome oxidase III. the magnetic susceptibility of beef heart cytochrome oxidase and some of its derivatives from 7-200 K. Direct evidence for an antiferromagnetically coupled Fe(III)/Cu(II) pair.", J. Biol. Chem. **253**, 8065–8071 (1978).

[64] G. T. Babcock, P. M. Callahan, M. R. Ondrias, and I Salmeen, "Coordination geometries and vibrational properties of cytochromes *a* and a_3 in cytochrome oxidase from soret excitation raman spectroscopy.", Biochemistry **20**, 959–966 (1981).

[65] G Buse, T Soulimane, M Dewor, H. E. Meyer, and M Blüggel, "Evidence for a copper-coordinated histidine-tyrosine cross-link in the active site of cytochrome oxidase.", Protein Sci. **8**, 985–990 (1999).

[66] J. P. Hosler, S Ferguson-Miller, M. W. Calhoun, J. W. Thomas, J Hill, L Lemieux, J Ma, C Georgiou, J Fetter, and J Shapleigh, "Insight into the active-site structure and function of cytochrome oxidase by analysis of site-directed mutants of bacterial cytochrome aa_3 and cytochrome *bo*.", J. Bioenerg. Biomembr. **25**, 121–136 (1993).

[67] V. R. Kaila, M. P. Johansson, D. Sundholm, L. Laakkonen, and M. Wikström, “The chemistry of the Cu_B site in cytochrome *c* oxidase and the importance of its unique His–Tyr bond”, Biochim. Biophys. Acta, Bioenerg. **1787**, 221–233 (2009).

[68] E. A. Gorbikova, M. Wikström, and M. I. Verkhovsky, “The protonation state of the cross-linked tyrosine during the catalytic cycle of cytochrome *c* oxidase.”, J. Biol. Chem. **283**, 34907–34912 (2008).

[69] G. T. Babcock, L. E. Vickery, and G Palmer, “Electronic state of heme in cytochrome oxidase. I. Magnetic circular dichroism of the isolated enzyme and its derivatives.”, J. Biol. Chem. **251**, 7907–7919 (1976).

[70] D. A. Proshlyakov, M. A. Pressler, and G. T. Babcock, “Dioxygen activation and bond cleavage by mixed-valence cytochrome *c* oxidase.”, Proc. Natl. Acad. Sci. U.S.A. **95**, 8020–8025 (1998).

[71] G. T. Babcock, “How oxygen is activated and reduced in respiration.”, Proc. Natl. Acad. Sci. U.S.A. **96**, 12971–12973 (1999).

[72] M Wikström, K Krab, and M Saraste, “Proton-translocating cytochrome complexes.”, Annu. Rev. Biochem. **50**, 623–655 (1981).

[73] A. J. Thomson, C Greenwood, P. M. Gadsby, J Peterson, D. G. Eglinton, B. C. Hill, and P Nicholls, “The structure of the cytochrome a_3-Cu_B site of mammalian cytochrome *c* oxidase as probed by MCD and EPR spectroscopy.”, J. Inorg. Biochem. **23**, 187–197 (1985).

[74] K. M. McCauley, J. M. Vrtis, J. Dupont, and W. A. van der Donk, “Insights into the fu nctional role of the tyrosine-histidine linkage in cytochrome *c* oxidase”, J. Am. Chem. Soc. **122**, 2403–2404 (2000).

[75] S Larsson, B Källebring, P Wittung, and B. G. Malmström, “The Cu_A center of cytochrome-*c* oxidase: electronic structure and spectra of models compared to the properties of Cu_A domains.”, Proc. Natl. Acad. Sci. U.S.A. **92**, 7167–7171 (1995).

[76] P. M. Kroneck, W. E. Antholine, D. H. Kastrau, G Buse, G. C. Steffens, and W. G. Zumft, “Multifrequency EPR evidence for a bimetallic center at the Cu_A site in cytochrome *c* oxidase.”, FEBS Lett. **268**, 274–276 (1990).

[77] M. Brunori, G. Antonini, F. Malatest, P. Sarti, and M. T. Wilson, “Cytochrome-*c* oxidase”, Eur. J. Biochem. **169**, 1–8 (1987).

[78] D. A. Mills and J. P. Hosler, “Slow proton transfer through the pathways for pumped protons in cytochrome *c* oxidase induces suicide inactivation of the enzyme”, Biochemistry **44**, 4656–4666 (2005).

[79] K. S. Alnajjar, J. Hosler, and L. Prochaska, “Role of the N-terminus of subunit III in proton uptake in cytochrome *c* oxidase of *Rhodobacter sphaeroides*.”, Biochemistry **53**, 496–504 (2014).

[80] L. Varanasi and J. P. Hosler, “Subunit III-depleted cytochrome *c* oxidase provides insight into the process of proton uptake by proteins.”, Biochim. Biophys. Acta **1817**, 545–551 (2012).

[81] J. P. Hosler, “The influence of subunit III of cytochrome *c* oxidase on the D pathway, the proton exit pathway and mechanism-based inactivation in subunit I.”, Biochim. Biophys. Acta **1655**, 332–339 (2004).

[82] M. Wikström, V. Sharma, V. R. I. Kaila, J. P. Hosler, and G. Hummer, “New perspectives on proton pumping in cellular respiration”, Chem. Rev. **115**, 2196–221 (2015).

[83] X. Zheng, D. M. Medvedev, J. Swanson, and A. A. Stuchebrukhov, "Computer simulation of water in cytochrome *c* oxidase.", Biochim. Biophys. Acta **1557**, 99–107 (2003).

[84] D. M. Popović and A. A. Stuchebrukhov, "Proton exit channels in bovine cytochrome *c* oxidase.", J. Phys. Chem. B **109**, 1999–2006 (2005).

[85] B. Schmidt, J. McCracken, and S. Ferguson-Miller, "A discrete water exit pathway in the membrane protein cytochrome *c* oxidase.", Proc. Natl. Acad. Sci. U.S.A. **100**, 15539–15542 (2003).

[86] R. Sugitani and A. A. Stuchebrukhov, "Molecular dynamics simulation of water in cytochrome *c* oxidase reveals two water exit pathways and the mechanism of transport", Biochim. Biophys. Acta **1787**, 1140–50 (2009).

[87] N. Yano, K. Muramoto, A. Shimada, S. Takemura, J. Baba, H. Fujisawa, M. Mochizuki, K. Shinzawa-Itoh, E. Yamashita, T. Tsukihara, and S. Yoshikawa, "The Mg^{2+}-containing water cluster of mammalian cytochrome *c* oxidase collects four pumping proton equivalents in each catalytic cycle.", J. Biol. Chem. **291**, 23882–23894 (2016).

[88] V. Rauhamäki and M. Wikström, "The causes of reduced proton-pumping efficiency in type B and C respiratory heme-copper oxidases, and in some mutated variants of type A.", Biochim. Biophys. Acta **1837**, 999–1003 (2014).

[89] C. J. T. Grotthuss, "Mémoire-sur la décomposition de l'eau et des corps qu'elle tient en dissolution á l'aide the l'électricité galvanique", Ann. Chim. (Paris) **LVIII**, 54–74 (1806).

[90] L Onsager, "The motion of ions: principles and concepts.", Science **166**, 1359–1364 (1969).

[91] J. F. Nagle and S Tristram-Nagle, "Hydrogen bonded chain mechanisms for proton conduction and proton pumping.", J. Membr. Biol. **74**, 1–14 (1983).

[92] S. Cukierman, "Et tu, Grotthuss! and other unfinished stories.", Biochim. Biophys. Acta **1757**, 876–885 (2006).

[93] D. Marx, "Proton transfer 200 years after von Grotthuss: insights from ab initio simulations.", Chemphyschem **7**, 1848–1870 (2006).

[94] P. R. Rich and A. Maréchal, "Functions of the hydrophilic channels in protonmotive cytochrome *c* oxidase.", J. R. Soc. Interface **10**, 20130183 (2013).

[95] S. Iwata, C. Ostermeier, B. Ludwig, and H. Michel, "Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*", Nature **376**, 660–9 (1995).

[96] I. Belevich, D. A. Bloch, N. Belevich, M. Wikström, and M. I. Verkhovsky, "Exploring the proton pump mechanism of cytochrome *c* oxidase in real time", Proc. Natl. Acad. Sci. U. S. A. **104**, 2685–90 (2007).

[97] R. B. Gennis, "Multiple proton-conducting pathways in cytochrome oxidase and a proposed role for the active-site tyrosine", Biochim. Biophys. Acta, Bioenerg. **1365**, 241–248 (1998).

[98] I. Belevich, M. I. Verkhovsky, and M. Wikström, "Proton-coupled electron transfer drives the proton pump of cytochrome *c* oxidase", Nature **440**, 829–32 (2006).

[99] V. R. I. Kaila, M. I. Verkhovsky, G. Hummer, and M. Wikström, "Glutamic acid 242 is a valve in the proton pump of cytochrome *c* oxidase", Proc. Natl. Acad. Sci. U. S. A. **105**, 6255–9 (2008).

[100] S. Jünemann, B. Meunier, N. Fisher, and P. R. Rich, "Effects of mutation of the conserved glutamic acid-286 in subunit I of cytochrome *c* oxidase from *Rhodobacter sphaeroides*", Biochemistry **38**, 5248–5255 (1999).

[101] A. A. Konstantinov, S. Siletsky, D. Mitchell, A. Kaulen, and R. B. Gennis, "The roles of the two proton input channels in cytochrome *c* oxidase from *Rhodobacter sphaeroides* probed by the effects of site-directed mutations on time-resolved electrogenic intraprotein proton transfer", Proc. Natl. Acad. Sci. U.S.A. **94**, 9085–9090 (1997).

[102] P. Ädelroth, M. Svensson Ek, D. M. Mitchell, R. B. Gennis, and P. Brzezinski, "Glutamate 286 in cytochrome aa_3 from *Rhodobacter sphaeroides* is involved in proton uptake during the reaction of the fully-reduced enzyme with dioxygen", Biochemistry **36**, 13824–13829 (1997).

[103] I. Belevich, E. Gorbikova, N. P. Belevich, V. Rauhamäki, M. Wikström, and M. I. Verkhovsky, "Initiation of the proton pump of cytochrome *c* oxidase", Proc. Natl. Acad. Sci. U.S.A. **107**, 18469–18474 (2010).

[104] M. Wikström and M. I. Verkhovsky, "The D-channel of cytochrome oxidase: An alternative view", Biochim. Biophys. Acta, Bioenerg. **1807**, 1273–1278 (2011).

[105] S. A. Siletsky, A. S. Pawate, K. Weiss, R. B. Gennis, and A. A. Konstantinov, "Transmembrane charge separation during the ferryl-oxo → oxidized transition in a nonpumping mutant of cytochrome *c* oxidase", J. Biol. Chem. **279**, 52558–65 (2004).

[106] V. R. I. Kaila, M. I. Verkhovsky, G. Hummer, and M. Wikström, "Mechanism and energetics by which glutamic acid 242 prevents leaks in cytochrome *c* oxidase", Biochim. Biophys. Acta **1787**, 1205–14 (2009).

[107] U. Pfitzner, K. Hoffmeier, A. Harrenga, A. Kannt, H. Michel, E. Bamberg, O.-M. Richter, and B. Ludwig, "Tracing the D-pathway in reconstituted site-directed mutants of cytochrome *c* oxidase from *Paracoccus denitrificans*", Biochemistry **39**, 6756–6762 (2000).

[108] J. Zhu, H. Han, A. Pawate, and R. B. Gennis, "Decoupling mutations in the D-channel of the aa_3-type cytochrome *c* oxidase from *Rhodobacter sphaeroides* suggest that a continuous hydrogen-bonded chain of waters is essential for proton pumping", Biochemistry **49**, 4476 (2010).

[109] D. Han, A. Namslauer, A. Pawate, J. E. Morgan, S. Nagy, A. S. Vakkasoglu, P. Brzezinski, and R. B. Gennis, "Replacing Asn207 by aspartate at the neck of the D channel in the aa_3-type cytochrome *c* oxidase from *Rhodobacter sphaeroides* results in decoupling the proton pump", Biochemistry **45**, 14064 (2006).

[110] G. Gilderson, L. Salomonsson, A. Aagaard, J. Gray, P. Brzezinski, and J. Hosler, "Subunit III of cytochrome *c* oxidase of *Rhodobacter sphaeroides* is required to maintain rapid proton uptake through the D pathway at physiologic pH", Biochemistry **42**, 7400–7409 (2003).

[111] P. Brzezinski and A.-L. Johansson, "Variable proton-pumping stoichiometry in structural variants of cytochrome *c* oxidase", Biochim. Biophys. Acta, Bioenerg. **1797**, 710–723 (2010).

[112] A Kannt, C. R. Lancaster, and H Michel, "The coupling of electron transfer and proton translocation: electrostatic calculations on *Paracoccus denitrificans* cytochrome *c* oxidase.", Biophys. J. **74**, 708–721 (1998).

[113] C. Hiser, L. Buhrow, J. Liu, L. Kuhn, and S. Ferguson-Miller, “A conserved amphipathic ligand binding region influences K-path-dependent activity of cytochrome *c* oxidase.”, Biochemistry **52**, 1385–1396 (2013).

[114] J. W. Thomas, L. J. Lemieux, J. O. Alben, and R. B. Gennis, “Site-directed mutagenesis of highly conserved residues in helix VIII of subunit I of the cytochrome *bo* ubiquinol oxidase from *Escherichia coli*: an amphipathic transmembrane helix that may be important in conveying protons to the binuclear center.”, Biochemistry **32**, 11173–11180 (1993).

[115] J. P. Hosler, J. P. Shapleigh, D. M. Mitchell, Y Kim, M. A. Pressler, C Georgiou, G. T. Babcock, J. O. Alben, S Ferguson-Miller, and R. B. Gennis, “Polar residues in helix VIII of subunit I of cytochrome *c* oxidase influence the activity and the structure of the active site.”, Biochemistry **35**, 10776–10783 (1996).

[116] M Wikström, A Jasaitis, C Backgren, A Puustinen, and M. I. Verkhovsky, “The role of the D- and K-pathways of proton transfer in the function of the haem-copper oxidases.”, Biochim. Biophys. Acta **1459**, 514–520 (2000).

[117] I. Belevich and M. I. Verkhovsky, “Molecular mechanism of proton translocation by cytochrome *c* oxidase.”, Antioxid. Redox Signaling **10**, 1–29 (2008).

[118] M. R. A. Blomberg, “Mechanism of oxygen reduction in cytochrome c oxidase and the role of the active site tyrosine.”, Biochemistry **55**, 489–500 (2016).

[119] H. Y. Chang, S. K. Choi, A. S. Vakkasoglu, Y. Chen, J. Hemp, J. A. Fee, and R. B. Gennis, “Exploring the proton pump and exit pathway for pumped protons in cytochrome ba_3 from *Thermus thermophilus*”, Proc. Natl. Acad. Sci. U.S.A. **109**, 5259–64 (2012).

[120] K. Shimokata, Y. Katayama, H. Murayama, M. Suematsu, T. Tsukihara, K. Muramoto, H. Aoyama, S. Yoshikawa, and H. Shimada, “The proton pumping pathway of bovine heart cytochrome *c* oxidase.”, Proc. Natl. Acad. Sci. U.S.A. **104**, 4200–4205 (2007).

[121] P. G. Furtmüller, M. Zederbauer, W. Jantschko, J. Helm, M. Bogner, C. Jakopitsch, and C. Obinger, “Active site structure and catalytic mechanisms of human peroxidases.”, Arch. Biochem. Biophys. **445**, 199–213 (2006).

[122] R. Aminaka, K. Shimokata, M. Itoh, Y. Katayama, T. Tsukihara, S. Yoshikawa, and H. Shimada, “Mutagenesis studies on D-pathway function of bovine heart cytochrome *c* oxidase”, Biochim. Biophys. Acta, Bioenerg. **1797**, 93–94 (2010).

[123] I. Hofacker and K. Schulten, “Oxygen and proton pathways in cytochrome *c* oxidase”, Proteins: Struct., Funct., Genet. **30**, 100–107 (1998).

[124] K. Shinzawa-Itoh, H. Aoyama, K. Muramoto, H. Terada, T. Kurauchi, Y. Tadehara, A. Yamasaki, T. Sugimura, S. Kurono, K. Tsujimoto, et al., “Structures and physiological roles of 13 integral lipids of bovine heart cytochrome *c* oxidase”, EMBO J. **26**, 1713–1725 (2007).

[125] S. Riistama, A. Puustinen, M. I. Verkhovsky, J. E. Morgan, and M. Wikström, “Binding of O_2 and its reduction are both retarded by replacement of valine 279 by isoleucine in cytochrome *c* oxidase from *Paracoccus denitrificans*”, Biochemistry **39**, 6365–6372 (2000).

[126] E. A. Gorbikova, I. Belevich, M. Wikström, and M. I. Verkhovsky, “The proton donor for O-O bond scission by cytochrome *c* oxidase.”, Proc. Natl. Acad. Sci. U.S.A. **105**, 10733–10737 (2008).

[127] B. Chance, C. Saronio, and J. Leigh, "Functional intermediates in the reaction of membrane-bound cytochrome oxidase with oxygen.", J. Biol. Chem. **250**, 9226–9237 (1975).

[128] B. Chance, C. Saronio, and J. Leigh, "Functional intermediates in reaction of cytochrome oxidase with oxygen", Proc. Natl. Acad. Sci. U.S.A. **72**, 1635–1640 (1975).

[129] M. I. Verkhovsky, J. E. Morgan, A. Puustinen, and M. Wikstrom, "Kinetic trapping of oxygen in cell respiration", Nature **380**, 268 (1996).

[130] S. Han, Y.-c. Ching, and D. L. Rousseau, "Ferryl and hydroxy intermediates in the reaction of oxygen with reduced cytochrome *c* oxidase", Nature **348**, 89–90 (1990).

[131] V. Sharma, K. D. Karlin, and M. Wikström, "Computational study of the activated o_h state in the catalytic mechanism of cytochrome *c* oxidase", Proc. Natl. Acad. Sci. U.S.A. **110**, 16844–16849 (2013).

[132] E. Fadda, C. H. Yu, and R. Pomes, "Electrostatic control of proton pumping in cytochrome *c* oxidase", Biochim. Biophys. Acta **1777**, 277–84 (2008).

[133] R. M. Nyquist, D. Heitbrink, C. Bolwien, R. B. Gennis, and J. Heberle, "Direct observation of protonation reactions during the catalytic cycle of cytochrome *c* oxidase", Proc. Natl. Acad. Sci. U. S. A. **100**, 8715–20 (2003).

[134] T. Egawa, S.-R. Yeh, and D. L. Rousseau, "Redox-controlled proton gating in bovine cytochrome *c* oxidase", PLoS One **8**, e63669 (2013).

[135] P. Rich, "Towards an understanding of the chemistry of oxygen reduction", Functional Plant Biology **22**, 479–486 (1995).

[136] M. Iwaki and P. R. Rich, "An IR study of protonation changes associated with heme-heme electron transfer in bovine cytochrome *c* oxidase", J. Am. Chem. Soc. **129**, 2923–9 (2007).

[137] A. Marechal and P. R. Rich, "Water molecule reorganization in cytochrome *c* oxidase revealed by FTIR spectroscopy", Proc. Natl. Acad. Sci. U. S. A. **108**, 8634–8 (2011).

[138] V. R. I. Kaila, V. Sharma, and M. Wikström, "The identity of the transient proton loading site of the proton-pumping mechanism of cytochrome *c* oxidase", Biochim. Biophys. Acta **1807**, 80–4 (2011).

[139] P. E. M. Siegbahn, M. R. A. Blomberg, and M. L. Blomberg, "Theoretical study of the energetics of proton pumping and oxygen reduction in cytochrome oxidase", J. Phys. Chem. B (2003).

[140] W.-G. H. Du and L. Noodleman, "Broken symmetry DFT calculations/analysis for oxidized and reduced dinuclear center in cytochrome c oxidase: relating structures, protonation states, energies, and mössbauer properties in ba_3 *Thermus thermophilus*", Inorg. Chem. **54**, 7272 (2015).

[141] J. Quenneville, D. M. Popović, and A. A. Stuchebrukhov, "Combined DFT and electrostatics study of the proton pumping mechanism in cytochrome *c* oxidase", Biochim. Biophys. Acta, Bioenerg. **1757**, 1035–1046 (2006).

[142] J. Lu and M. R. Gunner, "Characterizing the proton loading site in cytochrome *c* oxidase", Proc. Natl. Acad. Sci. U. S. A. **111**, 12414–9 (2014).

[143] J. E. Morgan, M. I. Verkhovsky, and M. Wikström, "The histidine cycle: a new model for proton translocation in the respiratory heme-copper oxidases", J. Bioenerg. Biomembr. **26**, 599–608 (1994).

[144] R. Mitchell and P. R. Rich, "Proton uptake by cytochrome *c* oxidase on reduction and on ligand binding", Biochim. Biophys. Acta, Bioenerg. **1186**, 19–26 (1994).

[145] H. Michel, “The mechanism of proton pumping by cytochrome *c* oxidase”, Proc. Natl. Acad. Sci. U.S.A. **95**, 12819–12824 (1998).

[146] M. Wikström, M. I. Verkhovsky, and G. Hummer, “Water-gated mechanism of proton translocation by cytochrome *c* oxidase”, Biochim. Biophys. Acta **1604**, 61–5 (2003).

[147] Y. C. Kim, M. Wikström, and G. Hummer, “Kinetic gating of the proton pump in cytochrome *c* oxidase”, Proc. Natl. Acad. Sci. U.S.A. **106**, 13707–13712 (2009).

[148] Y. C. Kim, M. Wikström, and G. Hummer, “Kinetic models of redox-coupled proton pumping”, Proc. Natl. Acad. Sci. U.S.A. **104**, 2169–2174 (2007).

[149] Y. C. Kim and G. Hummer, “Proton-pumping mechanism of cytochrome *c* oxidase: a kinetic master-equation approach”, Biochim. Biophys. Acta, Bioenerg. **1817**, 526–536 (2012).

[150] A.-L. Johansson, S. Chakrabarty, C. L. Berthold, M. Högbom, A. Warshel, and P. Brzezinski, “Proton-transport mechanisms in cytochrome *c* oxidase revealed by studies of kinetic isotope effects”, Biochim. Biophys. Acta, Bioenerg. **1807**, 1083–1094 (2011).

[151] S. Chakrabarty, I. Namslauer, P. Brzezinski, and A. Warshel, “Exploration of the cytochrome *c* oxidase pathway puzzle and examination of the origin of elusive mutational effects”, Biochim. Biophys. Acta **1807**, 413–26 (2011).

[152] A. V. Pisliakov, P. K. Sharma, Z. T. Chu, M. Haranczyk, and A. Warshel, “Electrostatic basis for the unidirectionality of the primary proton transfer in cytochrome *c* oxidase”, Proc. Natl. Acad. Sci. U. S. A. **105**, 7726–31 (2008).

[153] M. H. Olsson and A. Warshel, “Monte carlo simulations of proton pumps: on the working principles of the biological valve that controls proton pumping in cytochrome *c* oxidase”, Proc. Natl. Acad. Sci. U.S.A. **103**, 6500–6505 (2006).

[154] M. H. Olsson, P. E. Siegbahn, M. R. Blomberg, and A. Warshel, “Exploring pathways and barriers for coupled ET/PT in cytochrome *c* oxidase: a general framework for examining energetics and mechanistic alternatives”, Biochim. Biophys. Acta, Bioenerg. **1767**, 244–260 (2007).

[155] T. Yamashita and G. A. Voth, “Insights into the mechanism of proton transport in cytochrome *c* oxidase”, J. Am. Chem. Soc. **134**, 1147–52 (2012).

[156] J. Xu and G. A. Voth, “Redox-coupled proton pumping in cytochrome *c* oxidase: further insights from computer simulation”, Biochim. Biophys. Acta **1777**, 196–201 (2008).

[157] J. Xu, M. A. Sharpe, L. Qin, S. Ferguson-Miller, and G. A. Voth, “Storage of an excess proton in the hydrogen-bonded network of the D-pathway of cytochrome *c* oxidase: identification of a protonated water cluster”, J. Am. Chem. Soc. **129**, 2910–3 (2007).

[158] Y. Peng, J. M. Swanson, S.-g. Kang, R. Zhou, and G. A. Voth, “Hydrated excess protons can create their own water wires”, J. Phys. Chem. B **119**, 9212–9218 (2014).

[159] N. Ghosh, X. Prat-Resina, M. Gunner, and Q. Cui, “Microscopic pK_a analysis of Glu286 in cytochrome *c* oxidase (*Rhodobacter sphaeroides*): toward a calibrated molecular model”, Biochemistry **48**, 2468–2485 (2009).

[160] S. Yang and Q. Cui, “Glu286 rotation and water wire reorientation are unlikely the gating elements for proton pumping in cytochrome *c* oxidase”, Biophys. J. **101**, 61–69 (2011).

[161] P. Goyal, J. Lu, S. Yang, M. Gunner, and Q. Cui, “Changing hydration level in an internal cavity modulates the proton affinity of a key glutamate in cytochrome c oxidase”, Proc. Natl. Acad. Sci. U.S.A. **110**, 18886–18891 (2013).

[162] P. Goyal, S. Yang, and Q. Cui, “Microscopic basis for kinetic gating in cytochrome c oxidase: insights from QM/MM analysis”, Chem. Sci. **6**, 826–841 (2015).

[163] P. E. Siegbahn and M. R. Blomberg, “Proton pumping mechanism in cytochrome c oxidase”, J. Phys. Chem. A **112**, 12772–80 (2008).

[164] M. R. Blomberg and P. E. Siegbahn, “The mechanism for proton pumping in cytochrome c oxidase from an electrostatic and quantum chemical perspective”, Biochim. Biophys. Acta **1817**, 495–505 (2012).

[165] H. Hu and W. Yang, “Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods”, Annu. Rev. Phys. Chem. **59**, 573–601 (2008).

[166] P. A. M. Dirac, “Quantum mechanics of many-electron systems”, in Proc. r. soc. london, ser. a, Vol. 123, 792 (The Royal Society, 1929), pp. 714–733.

[167] E. Schrödinger, “An undulatory theory of the mechanics of atoms and molecules”, Phys. Rev. **28**, 1049 (1926).

[168] M. Born and R. Oppenheimer, “Zur quantentheorie der molekeln”, Ann. Phys. **389**, 457–484 (1927).

[169] M. Born and K. Huang, *Dynamical theory of crystal lattices* (Clarendon press, 1954).

[170] C. D. Snow, E. J. Sorin, Y. M. Rhee, and V. S. Pande, “How well can simulation predict protein folding kinetics and thermodynamics?”, Annu. Rev. Biophys. Biomol. Struct. **34**, 43–69 (2005).

[171] D. R. Hartree, “The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods”, in Math. proc. cambridge philos. soc. Vol. 24, 01 (Cambridge Univ Press, 1928), pp. 89–110.

[172] W. Pauli, “Über den zusammenhang des abschlusses der elektronengruppen im atom mit der komplexstruktur der spektren”, Z. Physik A-Hadron Nucl. **31**, 765–783 (1925).

[173] J. C. Slater, “The theory of complex spectra”, Phys. Rev. **34**, 1293 (1929).

[174] V. Fock, “Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems”, Zeitschrift für Physik A Hadrons and Nuclei **61**, 126–148 (1930).

[175] J. C. Slater, “Note on hartree’s method”, Phys. Rev. **35**, 210 (1930).

[176] C. C. J. Roothaan, “New developments in molecular orbital theory”, Rev. Mod. Phys. **23**, 69 (1951).

[177] G. G. Hall, “The molecular orbital theory of chemical valency. VIII. A method of calculating ionization potentials”, Proc. R. Soc. London Ser. A **205**, 541–552 (1951).

[178] P.-O. Löwdin, “Exchange, correlation, and spin effects in molecular and solid-state theory”, Rev. Mod. Phys. **34**, 80 (1962).

[179] M. Kolář, T. Kubař, and P. Hobza, “On the role of london dispersion forces in biomolecular structure determination.”, J. Phys. Chem. B **115**, 8038–8046 (2011).

[180] J. Vondrášek, L. Bendová, V. Klusák, and P. Hobza, “Unexpectedly strong energy stabilization inside the hydrophobic core of small protein rubredoxin mediated by aromatic residues: correlated ab initio quantum chemical calculations”, J. Am. Chem. Soc. **127**, 2615–2619 (2005).

[181] X. He, L. Fusti-Molnar, G. Cui, and K. M. Merz Jr, “Importance of dispersion and electron correlation in ab initio protein folding”, J. Phys. Chem. B **113**, 5290–5300 (2009).

[182] A. J. Cohen, P. Mori-Sánchez, and W. Yang, “Challenges for density functional theory”, Chem. Rev. **112**, 289–320 (2011).

[183] I. Guseinov, “New complete orthonormal sets of exponential-type orbitals and their application to translation of slater orbitals”, Int. J. Quantum Chem. **90**, 114–118 (2002).

[184] W. J. Hehre, R. F. Stewart, and J. A. Pople, “Self-consistent molecular-orbital methods. i. use of gaussian expansions of Slater-type atomic orbitals”, J. Chem. Phys. **51**, 2657–2664 (1969).

[185] R. Ditchfield, W. J. Hehre, and J. A. Pople, “Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules”, J. Chem. Phys. **54**, 724–728 (1971).

[186] F. Weigend and R. Ahlrichs, “Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy”, Phys. Chem. Chem. Phys. **7**, 3297–305 (2005).

[187] F. Jensen, *Introduction to computational chemistry* (John Wiley & Sons, 2006).

[188] C. Møller and M. S. Plesset, “Note on an approximation treatment for many-electron systems”, Phys. Rev. **46**, 618–622 (1934).

[189] F. Weigend and M. Häser, “RI-MP2: first derivatives and global consistency”, Theor. Chem. Acc. **97**, 331–340 (1997).

[190] R. J. Bartlett and M. Musiał, “Coupled-cluster theory in quantum chemistry”, Rev. Mod. Phys. **79**, 291 (2007).

[191] J. Čížek, “On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods, journal = J. Chem. Phys.”, **45**, 4256–4266 (1966).

[192] H. G. Kümmel, “A biography of the coupled cluster method”, Int. J. Mod Phys B **17**, 5311–5325 (2003).

[193] L. H. Thomas, “The calculation of atomic fields”, Math. Proc. Cambridge Philos. Soc. **23**, 542–548 (1927).

[194] E. Fermi, “Eine statistische methode zur bestimmung einiger eigenschaften des atoms und ihre anwendung auf die theorie des periodischen systems der elemente”, Z. Physik **48**, 73–79 (1928).

[195] P. A. M. Dirac, “Note on Exchange Phenomena in the Thomas Atom”, Proc. Cambridge Philos. Soc. **26**, 376 (1930).

[196] N. H. March, “Origins—the thomas-fermi theory”, in *Theory of the inhomogeneous electron gas*, edited by S. Lundqvist and N. H. March (Springer US, Boston, MA, 1983), pp. 1–77.

[197] P. Hohenberg and W. Kohn, “Inhomogeneous electron gas”, Phys. Rev. **136**, B864–B871 (1964).

[198] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects", Phys. Rev. **140**, A1133–A1138 (1965).

[199] U von Barth, "Basic density-functional theory—an overview", Phys. Scr. **2004**, 9 (2004).

[200] G. Oliver and J. Perdew, "Spin-density gradient expansion for the kinetic energy", Phys. Rev. A **20**, 397 (1979).

[201] D. M. Ceperley and B. Alder, "Ground state of the electron gas by a stochastic method", Phys. Rev. Lett. **45**, 566 (1980).

[202] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka, "Prescription for the design and selection of density functional approximations: more constraint satisfaction with fewer fits", J. Chem. Phys. **123**, 062201 (2005).

[203] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior", Phys. Rev. A **38**, 3098–3100 (1988).

[204] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density", Phys. Rev. B **37**, 785 (1988).

[205] J. P. Perdew, "Density-functional approximation for the correlation energy of the inhomogeneous electron gas", Phys. Rev. B: Condens. Matter Mater. Phys. **33**, 8822–8824 (1986).

[206] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, "Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation", Phys. Rev. B **46**, 6671 (1992).

[207] J. P. Perdew, M. Ernzerhof, and K. Burke, "Rationale for mixing exact exchange with density functional approximations", J. Chem. Phys. **105**, 9982–9985 (1996).

[208] A. St-Amant, W. D. Cornell, P. A. Kollman, and T. A. Halgren, "Calculation of molecular geometries, relative conformational energies, dipole moments, and molecular electrostatic potential fitted charges of small organic molecules of biochemical interest by density functional theory", J. Comput. Chem. **16**, 1483–1506 (1995).

[209] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, "Climbing the density functional ladder: nonempirical meta–generalized gradient approximation designed for molecules and solids", Phys. Rev. Lett. **91**, 146401 (2003).

[210] F. Furche and J. P. Perdew, "The performance of semilocal and hybrid density functionals in $3d$ transition-metal chemistry", J. Chem. Phys. **124**, 044103 (2006).

[211] A. D. Becke, "A new mixing of hartree-fock and local density-functional theories", J. Chem. Phys. **98**, 1372–1377 (1993).

[212] A. D. Becke, "Density functional thermochemistry. iii. the role of exact exchange", J. Chem. Phys. **98**, 5648–5652 (1993).

[213] P. Stephens, F. Devlin, C. Chabalowski, and M. J. Frisch, "Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields", J. Phys. Chem. **98**, 11623–11627 (1994).

[214] S. Grimme, "Semiempirical GGA-type density functional constructed with a long-range dispersion correction", J. Comput. Chem. **27**, 1787–99 (2006).

[215] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu", J. Chem. Phys. **132**, 154104 (2010).

[216] B. Alder and T. Wainwright, “Molecular dynamics by electronic computers”, Transport processes in statistical mechanics, 97–131 (1958).

[217] A. Rahman and F. H. Stillinger, “Molecular dynamics study of liquid water”, J. Chem. Phys. **55**, 3336–3359 (1971).

[218] F. H. Stillinger and A. Rahman, “Improved simulation of liquid water by molecular dynamics”, J. Chem. Phys. **60**, 1545–1557 (1974).

[219] J. A. McCammon, B. R. Gelin, and M. Karplus, “Dynamics of folded proteins.”, Nature **267**, 585 (1977).

[220] R. O. Dror, R. M. Dirks, J. Grossman, H. Xu, and D. E. Shaw, “Biomolecular simulation: a computational microscope for molecular biology”, Annu. Rev. Biophys. **41**, 429–452 (2012).

[221] I. Newton, *Philosophiæ naturalis principia mathematica* (1687).

[222] D Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, Sept. 1987), p. 288.

[223] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids* (Oxford university press, 1989).

[224] M. P. Allen et al., “Introduction to molecular dynamics simulation”, Computational soft matter: from synthetic polymers to proteins **23**, 1–28 (2004).

[225] W. F. van Gunsteren and H. J. Berendsen, “Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry”, Angew. Chem. Int. Ed. **29**, 992–1023 (1990).

[226] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, “All-atom empirical potential for molecular modeling and dynamics studies of proteins”, J. Phys. Chem. B **102**, 3586–616 (1998).

[227] R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig, and A. D. MacKerell Jr, “Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles”, J. Chem. Theory Comput. **8**, 3257–3273 (2012).

[228] D. M. York, T. A. Darden, and L. G. Pedersen, “The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods”, J. Chem. Phys. **99**, 8345–8348 (1993).

[229] T. Darden, D. York, and L. Pedersen, “Particle mesh Ewald: An N log (N) method for Ewald sums in large systems”, J. Chem. Phys. **98**, 10089–10092 (1993).

[230] P. P. Ewald, “Die berechnung optischer und elektrostatischer gitterpotentiale”, Ann. Phys. **369**, 253–287 (1921).

[231] N. Grønbech-Jensen and O. Farago, “Constant pressure and temperature discrete-time Langevin molecular dynamics”, J. Chem. Phys. **141**, 194108 (2014).

[232] G. M. Torrie and J. P. Valleau, “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”, J. Comput. Phys. **23**, 187–199 (1977).

[233] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, “The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method”, J. Comput. Chem. **13**, 1011–1021 (1992).

[234] Y. Song, E. Michonova-Alexova, and M. Gunner, “Calculated proton uptake on anaerobic reduction of cytochrome *c* oxidase: is the reaction electroneutral?”, Biochemistry **45**, 7959–7975 (2006).

[235] D. Bashford and K. Gerwert, “Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin”, J. Mol. Biol. **224**, 473–86 (1992).

[236] E. Bombarda and G. M. Ullmann, “Continuum electrostatic calculation on bovine rhodopsin: protonation and the effect of the membrane potential”, Photochem. Photobiol. (2017).

[237] D. G. Isom, C. A. Castañeda, B. R. Cannon, et al., “Large shifts in pK_a values of lysine residues buried inside a protein”, Proc. Natl. Acad. Sci. U.S.A. **108**, 5260–5265 (2011).

[238] A. Warshel, “Electrostatic basis of structure-function correlation in proteins”, Acc. Chem. Res. **14**, 284–290 (1981).

[239] D. Bashford and M. Karplus, “pK_a’s of ionizable groups in proteins: atomic detail from a continuum electrostatic model”, Biochemistry **29**, 10219–10225 (1990).

[240] G. M. Ullmann and E.-W. Knapp, “Electrostatic models for computing protonation and redox equilibria in proteins”, Eur. Biophys. J. **28**, 533–551 (1999).

[241] Y. Liu, A. H. Patel, S. K. Burger, and P. W. Ayers, “Benchmarking pK_a prediction methods for Lys115 in acetoacetate decarboxylase”, J. Mol. Model. **23**, 155 (2017).

[242] I. Sazanavets, A. Cawley, and J. Warwicker, “Electrostatics models for biology”, in (Springer, 2015), pp. 1–16.

[243] G. Kieseritzky and E.-W. Knapp, “Optimizing pK_a computation in proteins with pH adapted conformations”, Proteins: Struct., Funct., Bioinf. **71**, 1335–1348 (2008).

[244] A. Warshel and M. Levitt, “Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme”, J. Mol. Biol. **103**, 227–249 (1976).

[245] H. M. Senn and W. Thiel, “QM/MM methods for biomolecular systems”, Angew. Chem. Int. Ed. **48**, 1198–1229 (2009).

[246] H. Lin and D. G. Truhlar, “QM/MM: what have we learned, where are we, and where do we go from here?”, Theor. Chem. Acc. **117**, 185 (2007).

[247] M. W. van der Kamp and A. J. Mulholland, “Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology”, Biochemistry **52**, 2708–2728 (2013).

[248] M. Svensson, S. Humbel, R. D. Froese, T. Matsubara, S. Sieber, and K. Morokuma, “ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels- Alder reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ oxidative addition”, J. Phys. Chem. **100**, 19357–19363 (1996).

[249] K. E. Ranaghan and A. J. Mulholland, “QM/MM Methods for Simulating Enzyme Reactions”, Simulating Enzyme Reactivity: Computational Methods in Enzyme Catalysis, 377 (2016).

[250] M. J. Field, P. A. Bash, and M. Karplus, “A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations”, J. Comput. Chem. **11**, 700–733 (1990).

[251] D. M. Philipp and R. A. Friesner, “Mixed ab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide”, J. Comput. Chem. **20**, 1468–1494 (1999).

[252] R. Car and M. Parrinello, “Unified approach for molecular dynamics and density-functional theory”, Phys. Rev. Lett. **55**, 2471 (1985).

[253] M. E. Tuckerman, “Ab initio molecular dynamics: basic concepts, current trends and novel applications”, J. Phys.: Condens. Matter **14**, R1297 (2002).

[254] K. Faxén, G. Gilderson, P. Adelroth, and P. Brzezinski, “A mechanistic principle for proton pumping by cytochrome *c* oxidase”, Nature **437**, 286–9 (2005).

[255] H. Lepp, E. Svahn, K. Faxén, and P. Brzezinski, “Charge transfer in the K proton pathway linked to electron transfer to the catalytic site in cytochrome *c* oxidase”, Biochemistry **47**, 4929–35 (2008).

[256] F. Garczarek and K. Gerwert, “Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy”, Nature **439**, 109–12 (2006).

[257] P. Ball, “Water as an active constituent in cell biology”, Chem. Rev. **108**, 74–108 (2008).

[258] D. Saha, S. Supekar, and A. Mukherjee, “Distribution of residence time of water around DNA base pairs: governing factors and the origin of heterogeneity”, J. Phys. Chem. B **119**, 11371–11381 (2015).

[259] F. Sterpone, G. Stirnemann, and D. Laage, “Magnitude and molecular origin of water slowdown next to a protein”, J. Am. Chem. Soc. **134**, 4116–4119 (2012).

[260] B. Bagchi, “Water dynamics in the hydration layer around proteins and micelles”, Chem. Rev. **105**, 3197–3219 (2005).

[261] A. C. Fogarty and D. Laage, “Water dynamics in protein hydration shells: the molecular origins of the dynamical perturbation”, J. Phys. Chem. B **118**, 7715–7729 (2014).

[262] U. Ermler, G. Fritzsch, S. K. Buchanan, and H. Michel, “Structure of the photosynthetic reaction centre from Rhodobacter sphaeroides at 2.65 Å resolution: cofactors and protein-cofactor interactions”, Structure **2**, 925–936 (1994).

[263] R. Pomes and B. Roux, “Structure and dynamics of a proton wire: a theoretical study of H^+ translocation along the single-file water chain in the gramicidin A channel”, Biophys. J. **71**, 19–39 (1996).

[264] C. A. Wraight, “Chance and design—proton transfer in water, channels and bioenergetic proteins”, Biochim. Biophys. Acta, Bioenerg. **1757**, 886–912 (2006).

[265] A. E. García and G. Hummer, “Water penetration and escape in proteins”, Proteins: Struct., Funct., Bioinf. **38**, 261–272 (2000).

[266] S. Riistama, G. Hummer, A. Puustinen, R. B. Dyer, W. H. Woodruff, and M. Wikström, “Bound water in the proton translocation mechanism of the haem-copper oxidases”, FEBS Lett. **414**, 275–280 (1997).

[267] F. Zhu and K. Schulten, “Water and proton conduction through carbon nanotubes as models for biological channels”, Biophys. J. **85**, 236–244 (2003).

[268] B. F. Habenicht and S. J. Paddison, “Ab initio simulations of the effects of nanoscale confinement on proton transfer in hydrophobic environments”, J. Phys. Chem. B **115**, 10826–10835 (2011).

[269] L. D. Gelb, K. Gubbins, R Radhakrishnan, and M Sliwinska-Bartkowiak, “Phase separation in confined systems”, Rep. Prog. Phys. **62**, 1573 (1999).

[270] V. Sharma, G. Enkavi, I. Vattulainen, T. Rog, and M. Wikström, “Proton-coupled electron transfer and the role of water molecules in proton pumping by cytochrome *c* oxidase”, Proc. Natl. Acad. Sci. U. S. A. **112**, 2040–5 (2015).

[271] B. L. de Groot and H. Grubmüller, "The dynamics and energetics of water permeation and proton exclusion in aquaporins", Curr. Opin. Struct. Biol. **15**, 176–183 (2005).

[272] B. Ilan, E. Tajkhorshid, K. Schulten, and G. A. Voth, "The mechanism of proton exclusion in aquaporin channels", Proteins: Struct., Funct., Bioinf. **55**, 223–228 (2004).

[273] D. Wree, B. Wu, T. Zeuthen, and E. Beitz, "Requirement for asparagine in the aquaporin NPA sequence signature motifs for cation exclusion", FEBS J. **278**, 740–748 (2011).

[274] S. Vaitheeswaran, J. C. Rasaiah, and G. Hummer, "Electric field and temperature effects on water in the narrow nonpolar pores of carbon nanotubes", J. Chem. Phys. **121**, 7955–7965 (2004).

[275] M. Chaplin, "Water's hydrogen bond strength", arXiv preprint arXiv:0706.1355 (2007).

[276] M. P. Johansson, V. R. I. Kaila, and L. Laakkonen, "Charge parameterization of the metal centers in cytochrome *c* oxidase", J. Comput. Chem. **29**, 753–67 (2008).

[277] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable molecular dynamics with NAMD", J. Comput. Chem. **26**, 1781–802 (2005).

[278] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics", J. Mol. Graph **14**, 33–8, 27–8 (1996).

[279] E. W. Dijkstra, "A note on two problems in connexion with graphs", Numerische mathematik **1**, 269–271 (1959).

[280] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules", Nat. Struct. Mol. Biol. **9**, 646–652 (2002).

[281] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten, "Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality", J. Chem. Phys. **119**, 3559–3566 (2003).

[282] T. D. Romo and A. Grossfield, "Block covariance overlap method and convergence in molecular dynamics simulation", J. Chem. Theory Comput. **7**, 2464–2472 (2011).

[283] M. Wikström, C. Ribacka, M. Molin, L. Laakkonen, M. Verkhovsky, and A. Puustinen, "Gating of proton and water transfer in the respiratory enzyme cytochrome *c* oxidase", Proc. Natl. Acad. Sci. U. S. A. **102**, 10478–81 (2005).

[284] M. A. Sharpe, M. D. Krzyaniak, S. Xu, J. McCracken, and S. Ferguson-Miller, "EPR evidence of cyanide binding to the Mn(Mg) center of cytochrome *c* oxidase: support for Cu_A-Mg involvement in proton pumping", Biochemistry **48**, 328–35 (2009).

[285] D. M. Popovic and A. A. Stuchebrukhov, "Electrostatic study of the proton pumping mechanism in bovine heart cytochrome c oxidase", J. Am. Chem. Soc. **126**, 1858–71 (2004).

[286] P. E. M. Siegbahn and F. Himo, "The quantum chemical cluster approach for modeling enzyme reactions", WIREs Comput Mol Sci **1**, 323–336 (2011).

[287] M. Sierka, A. Hogekamp, and R. Ahlrichs, "Fast evaluation of the coulomb potential for electron densities using multipole accelerated resolution of identity approximation", J. Chem. Phys. **118**, 9136–9148 (2003).

[288] A. Klamt and G. Schüürmann, “COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient”, J. Chem. Soc., Perkin Trans. 2, 799–805 (1993).

[289] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, “Electronic structure calculations on workstation computers: the program system turbomole”, Chem. Phys. Lett. **162**, 165–169 (1989).

[290] H. L. Woodcock, M. Hodoscek, A. T. Gilbert, P. M. Gill, r. Schaefer H. F., and B. R. Brooks, “Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations”, J. Comput. Chem. **28**, 1485–502 (2007).

[291] S. Riahi and C. N. Rowley, “The CHARMM–TURBOMOLE interface for efficient and accurate QM/MM molecular dynamics, free energies, and excited state properties”, J. Comput. Chem. **35**, 2076–2086 (2014).

[292] M. Wikström and M. I. Verkhovsky, “Mechanism and energetics of proton translocation by the respiratory heme-copper oxidases”, Biochim. Biophys. Acta **1767**, 1200–14 (2007).

[293] P. E. Siegbahn and M. R. Blomberg, “Energy diagrams and mechanism for proton pumping in cytochrome *c* oxidase”, Biochim. Biophys. Acta **1767**, 1143–56 (2007).

[294] T. S. Zwier, “Chemistry. The structure of protonated water clusters”, Science **304**, 1119–20 (2004).

[295] J. M. Headrick, E. G. Diken, R. S. Walters, N. I. Hammer, R. A. Christie, J. Cui, E. M. Myshakin, M. A. Duncan, M. A. Johnson, and K. D. Jordan, “Spectral signatures of hydrated proton vibrations in water clusters”, Science **308**, 1765–9 (2005).

[296] P. Goyal, N. Ghosh, P. Phatak, M. Clemens, M. Gaus, M. Elstner, and Q. Cui, “Proton storage site in bacteriorhodopsin: new insights from quantum mechanics/molecular mechanics simulations of microscopic pK_a and infrared spectra”, J. Am. Chem. Soc. **133**, 14981–97 (2011).

[297] D. Riccardi, P. Schaefer, Y. Yang, H. Yu, N. Ghosh, X. Prat-Resina, P. König, G. Li, D. Xu, H. Guo, et al., *Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes*, 2006.

[298] V. R. I. Kaila and G. Hummer, “Energetics and dynamics of proton transfer reactions along short water wires”, Phys. Chem. Chem. Phys. **13**, 13207–13215 (2011).

[299] M. Karpefors, P. Ädelroth, A. Namslauer, Y. Zhen, and P. Brzezinski, “Formation of the “peroxy” intermediate in cytochrome *c* oxidase is associated with internal proton/hydrogen transfer”, Biochemistry **39**, 14664–14669 (2000).

[300] M. Karpefors, P. Ädelroth, and P. Brzezinski, “The onset of the deuterium isotope effect in cytochrome *c* oxidase”, Biochemistry **39**, 5045–5050 (2000).

[301] R. A. Kendall, T. H. Dunning Jr, and R. J. Harrison, “Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions”, J. Chem. Phys. **96**, 6796–6806 (1992).

[302] A. Jasaitis, M. I. Verkhovsky, J. E. Morgan, M. L. Verkhovskaya, and M. Wikström, “Assignment and charge translocation stoichiometries of the major electrogenic phases in the reaction of cytochrome *c* oxidase with dioxygen”, Biochemistry **38**, 2697–2706 (1999).

[303] P. R. Rich, B. Meunier, R. Mitchell, and A. J. Moody, “Coupling of charge and proton movement in cytochrome *c* oxidase”, Biochim. Biophys. Acta, Bioenerg. **1275**, 91–95 (1996).

[304] P. Ädelroth, R. B. Gennis, and P. Brzezinski, “Role of the pathway through K (I-362) in proton transfer in cytochrome *c* oxidase from *R. sphaeroides*”, Biochemistry **37**, 2470–2476 (1998).

[305] D. Bloch, I. Belevich, A. Jasaitis, C. Ribacka, A. Puustinen, M. I. Verkhovsky, and M. Wikström, “The catalytic cycle of cytochrome *c* oxidase is not the sum of its two halves”, Proc. Natl. Acad. Sci. U.S.A. **101**, 529–533 (2004).

[306] T. Vygodina, C Pecoraro, D Mitchell, R Gennis, and A. Konstantinov, “Mechanism of inhibition of electron transfer by amino acid replacement K362M in a proton channel of *Rhodobacter sphaeroides* cytochrome *c* oxidase”, Biochemistry **37**, 3053–3061 (1998).

[307] C. Pecoraro, R. B. Gennis, T. Vygodina, and A. Konstantinov, “Role of the K-channel in the pH-dependence of the reaction of cytochrome *c* oxidase with hydrogen peroxide”, Biochemistry **40**, 9695–9708 (2001).

[308] D. Riegler, L. Shroyer, C. Pokalsky, D. Zaslavsky, R. Gennis, and L. J. Prochaska, “Characterization of steady-state activities of cytochrome *c* oxidase at alkaline pH: mimicking the effect of K-channel mutations in the bovine enzyme”, Biochim. Biophys. Acta, Bioenerg. **1706**, 126–133 (2005).

[309] R. I. Cukier, “A molecular dynamics study of water chain formation in the proton-conducting K channel of cytochrome *c* oxidase”, Biochim. Biophys. Acta, Bioenerg. **1706**, 134–146 (2005).

[310] I. Kim and A. Warshel, “Analyzing the electrogenicity of cytochrome *c* oxidase”, Proc. Natl. Acad. Sci. U.S.A., 201608118 (2016).

[311] A. Tuukkanen, M. I. Verkhovsky, L. Laakkonen, and M. Wikström, “The K-pathway revisited: a computational study on cytochrome *c* oxidase”, Biochim. Biophys. Acta, Bioenerg. **1757**, 1117–1121 (2006).

[312] A. L. Woelke, G. Galstyan, and E.-W. Knapp, “Lysine 362 in cytochrome *c* oxidase regulates opening of the K-channel via changes in pK_a and conformation”, Biochim. Biophys. Acta, Bioenerg. **1837**, 1998–2003 (2014).

[313] V. Sharma and M. Wikström, “The role of the K-channel and the active-site tyrosine in the catalytic mechanism of cytochrome *c* oxidase”, Biochim. Biophys. Acta, Bioenerg. **1857**, 1111–1115 (2016).

[314] L. Qin, J. Liu, D. A. Mills, D. A. Proshlyakov, C. Hiser, and S. Ferguson-Miller, “Redox-dependent conformational changes in cytochrome *c* oxidase suggest a gating mechanism for proton uptake”, Biochemistry **48**, 5121–5130 (2009).

[315] M. R. Blomberg and P. E. Siegbahn, “How cytochrome *c* oxidase can pump four protons per oxygen molecule at high electrochemical gradient”, Biochim. Biophys. Acta, Bioenerg. **1847**, 364–376 (2015).

Acknowledgements

This dissertation would not have been possible without the help and support of many during the course of this work. Now that the day of submission is close, I would like to express, in writing, my gratitude to all those who have contributed, in both direct and indirect ways towards the completion of the dissertation.

First and foremost, I would like to thank my supervisor, Prof. Ville R. I. Kaila, for his excellent guidance, insightful discussions and for providing a great environment and computational resources for my research. Thank you for your patience, encouragement and belief from the beginning till the end. I am also really thankful for extensively detailed comments on the various drafts of the dissertation.

I would like to extend special thanks to Dr. Ana P. Gamiz-Hernandez for sharing your expertise of continuum electrostatics and especially QM/MM, which has been an integral part of this work. I would also like to thank Sophie Mader, Andrea Di Luca, Dr. Ina Bisha, Dr. Mikko Muuronen and Dr. Patrizia Saura and Dr. Gamiz-Hernandez again for proof-reading and providing feedback for parts of the dissertation. Furthermore, I would like to thank all the current and previous members of the group for all the scientific, and at times not so scientific discussions. It has been fun.

I would also like to extend my gratitude to Prof. Michael Sattler, Dr. Gerd Gemmecker and Dr. Mikael Johansson for collaboration and helpful discussions on the Tudor project.

Much appreciation is due to my friends in Munich: Alexej, Aradhana, Swati, Vishal and Padma for their support in their unique ways during this whole long drawn process. I would like to thank Sukanya for her support, inputs and for cheering me up in good and bad moments. Finally, I would like to express my deepest gratitude to my family, my parents and my brother, for their endless support and encouragement.