Computational investigation of oxygen reduction and proton pumping in *cbb*$_3$-type Cytochrome *c* Oxidases

Vivek Sharma

Institute of Biotechnology
and
Division of Biochemistry
Department of Biosciences
Faculty of Biological and Environmental Sciences
and
Viikki Doctoral Programme in Molecular Biosciences
University of Helsinki

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki in Auditorium 1041 of Biocenter 2, Viikinkaari 5, Helsinki, on 4th May 2012, at 12 noon.

Helsinki 2012

Supervisors
Prof. Mårten Wikström
Institute of Biotechnology,
University of Helsinki, Helsinki,
FINLAND.

Docent Liisa Laakkonen
General Microbiology, Dept. of Biosciences,
University of Helsinki, Helsinki,
FINLAND.

Thesis Advisory Committee
Prof. Arto Annila
Institute of Biotechnology,
University of Helsinki,
Helsinki,
FINLAND.

Prof. Adrian Goldman
Institute of Biotechnology,
University of Helsinki,
Helsinki,
FINLAND.

Pre-examiners
Prof. Ulf Ryde
Dept. of Theoretical Chemistry,
Lund University,
Lund,
SWEDEN.

Dr. Pia Ädelroth
Dept. of Biochemistry and
Biophysics, Stockholm
University, Stockholm,
SWEDEN.

Opponent
Dr. Gerhard Hummer
NIDDK,
National Institutes of Health, USA.

Custos
Prof. Kari Keinänen
Division of Biochemistry,
Department of Biosciences,
University of Helsinki, Helsinki,
FINLAND.

Front Cover: The figure illustrates the coupling of free-energy of oxygen reduction to proton pumping across the membrane. This function is performed by cytochrome *c* oxidase and the electrochemical gradient thus generated is eventually utilized in the synthesis of ATP.

ISBN 978-952-10-7734-0 (paperback)
ISBN 978-952-10-7735-7 (PDF, http://ethesis.helsinki.fi)
ISSN 1799-7372

Unigrafia
Helsinki 2012

In loving memory of my grandfather

LIST OF CONTENTS

List of Original Publications
Abbreviations and symbols
Abstract
1 Introduction....1
1.1 Oxidative phosphorylation....1
1.2 Respiratory complexes....2
2 Respiratory complex IV....4
2.1 Molecular oxygen....4
2.2 Cytochrome *c* oxidase (C*c*O)....5
2.3 Family classification of heme-copper oxidases (HCOs)....7
2.4 Evolutionary origins of HCOs....7
2.5 Structural and functional aspects of A-type oxidases....8
2.5.1 Subunit I....9
2.5.1.1 Active-site....9
2.5.1.2 Proton channels....10
2.5.2 Subunit II....11
2.5.3 Mechanism of oxygen reduction and proton pumping....12
2.6 Nuclear and mitochondrial coded subunits: Regulation....15
2.7 *cbb*$_3$- or C-type oxidases....16
2.8 Nitric oxide reductases (NORs)....19
3 Multiple sequence alignments and homology modeling....20
3.1 Sequence alignments....20
3.2 Structural alignments....21
3.3 Protein structure prediction....23
3.3.1 Homology modeling....23
3.3.1.1 MODELER: Model building and validation....24
3.3.1.2 MODELER: Problems and loop modeling....26
4 Density functional theory (DFT)....27
4.1 Early ideas in DFT....27
4.2 Formulation of DFT....27
4.2.1 The Schrödinger equation and the Self-consistent field (SCF) method....27
4.2.2 Kohn-Sham (KS) orbitals....29
4.3 Basis sets....29
4.4 DFT: Exchange Correlation (XC) functionals....30
4.4.1 Local (spin) density approximation (LSDA)....30
4.4.2 Generalized gradient approximation (GGA) or meta-GGA functionals....31
4.4.3 Hybrid functionals....31
4.5 Functionality in DFT....31
4.5.1 Coulombic term....31
4.5.2 Solvation effects in DFT....32
4.6 Performance of DFT....32
4.6.1 Geometry optimization....32
4.6.2 Enzyme mechanism modeling....33
4.6.3 Problems associated with DFT....33
5 Continuum electrostatics....35
5.1 pK_a calculations: general theory....35
5.2 Poisson-Boltzmann equation....37
5.3 Strength and weakness of continuum electrostatics method....37
6 Molecular Dynamics simulations....39
6.1 General theory....39
6.1.1 Integration schemes....41

6.2 Statistical ensembles and thermodynamics.....42
6.3 Molecular dynamics: Methods for constant temperature and pressure.....43
6.3.1 Constant temperature.....43
6.3.2 Constant pressure.....44
6.4 MD simulations on biomolecules.....44
7 Aims and methods.....46
7.1 Aims.....46
7.2 Methods.....46
7.2.1 Multiple sequence alignment and homology modeling.....46
7.2.2 Continuum electrostatic calculations.....47
7.2.3 Density functional theory calculations.....48
7.2.4 Molecular Dynamics simulations.....50
8 Results and discussion.....52
8.1 Homology model of *cbb*$_3$-oxidase and its comparison with the crystal structure.....52
8.2 Quantum chemical modeling of the active-sites.....54
8.3 Proton-loading site: pK_a calculations.....58
8.4 Water-networks in *cbb*$_3$-oxidase: MD simulations and pK_a calculations.....59
9 Conclusions and future directions.....63
Bibliography.....64
Acknowledgements.....74

List of publications and contributions from the author of the thesis

I. **Vivek Sharma**, Anne Puustinen, Mårten Wikström and Liisa Laakkonen. Sequence analysis of the *cbb*$_3$-oxidases and an atomic model for the *Rhodobacter sphaeroides* enzyme. 2006, *Biochemistry*, 45 (18), 5754-5765.

AP, MW and LL designed the research. VS performed most of the research work, and wrote initial text along with LL, with subsequent refinements from AP and MW.

II. **Vivek Sharma**, Mårten Wikström and Liisa Laakkonen. Modeling the active-site structure of the *cbb*$_3$-type oxidase from *Rhodobacter sphaeroides*. 2008, *Biochemistry*, 47 (14), 4221-4227.

Refinements to models were proposed by MW, based on the experimental data. VS performed the work, with inputs from LL. VS wrote initial text, with further refinements from LL and MW.

III. **Vivek Sharma**, Mårten Wikström and Ville R. I. Kaila. Redox-coupled proton transfer in the active site of cytochrome *cbb*$_3$. 2010, *Biochimica et Biophysica Acta*, 1797 (8), 1512-1520.

Original idea by VRIK and MW. VS did most of the calculations, with some auxiliary calculations by VRIK. VS wrote major part of the text with further refinements from VRIK and MW.

IV. Ville R. I. Kaila, **Vivek Sharma**, Mårten Wikström. The identity of the transient proton loading site of the proton-pumping mechanism of cytochrome *c* oxidase. *Biochimica et Biophysica Acta*, 1807 (1), 80-84.

Original idea by VRIK and MW. VS did the electrostatic calculations, wrote part of the methods, and results and discussion sections.

V. **Vivek Sharma**, Mårten Wikström and Ville R. I. Kaila. Stabilization of the peroxy intermediate in the oxygen splitting reaction of cytochrome *cbb*$_3$. *Biochimica et Biophysica Acta*, 1807 (7), 813-818.

Original idea by VS and VRIK. VS did all the calculations, and wrote major part of the text with refinements from VRIK and MW. VS is also the corresponding author in the paper.

VI. **Vivek Sharma**, Mårten Wikström and Ville R. I. Kaila. Dynamic water networks in cytochrome *cbb*$_3$ oxidase. *Biochimica et Biophysica Acta* (accepted).

Original idea by VS. VS did all the calculations, and wrote major part of the text with refinements from VRIK and MW. VS is also the corresponding author in the paper.

Abbreviations and symbols

pmf	proton motive force, $\Delta\mu^{H^+} = (RT/nF)\Delta pH + \Delta\Psi$
hυ	quanta of energy (h = 6.63x10^{-34} Js, υ is frequency)
ATP	adenosine triphosphate
Q	ubiquinone (or Coenzyme Q_{10}), a two electron carrier
QH_2	ubiquinol (two electron reduced ubiquinone)
F_1F_O-ATP synthase	ATP synthase. F_O - membrane bound subunit; F_1 - hydrophilic subunit
AMP	adenosine monophosphate
ADP	adenosine diphosphate
P_i	inorganic phosphate
NADH	nicotinamide adenine dinucleotide, a two electron carrier
$FADH_2$	flavin adenine dinucleotide, a two electron carrier
TM	transmembrane (helices)
$\Delta\Psi$	potential difference across membrane
ETC	electron transport chain
C*c*O	cytochrome *c* oxidase (or complex IV)
ROS	reactive oxygen species
BNC	binuclear center
PDB	protein data bank
EPR	electron paramagnetic resonance
BLAST	basic local alignment search tool
NMR	nuclear magnetic resonance
DFT	density functional theory
HF	Hartree Fock
XC	exchange correlation
SVP	split-valence *plus* polarization (basis set)
TZVP	triple-zeta valence *plus* polarization (basis set)
GGA	generalized gradient approximation
BP86	Becke (exchange) and Perdew (correlation) density functional
B3LYP	Becke (exchange) and Lee, Young, Parr (correlation) density functional
TPSS	Tao, Perdew, Staroverov, Scuseria density functional
MC-SCF	multi configurational – self consistent field
PB	Poisson-Boltzmann (equation)
MC	Monte-Carlo (method)
MD	molecular dynamics (method)
LJ	Lennard-Jones (interaction)
HCO	heme-copper oxidase
NOR	nitric oxide reductase
rmsd	root mean square deviation
ESP	electrostatic potential
HOMO	highest occupied molecular orbital
LUMO	lowest unoccupied molecular orbital

NOTES:

a) The *cbb*$_3$-type oxidases are also referred to as the C-type oxidases in the thesis.

b) Amino acid numbers are written in format; CI:Arg323, where first letter corresponds to the enzyme subfamily (A, B or C-type); second roman numeral corresponds to the subunit number (I, II, III...*etc.*); last term after colon refers to the name and number of amino acid residue.

Abstract

Heme-copper oxidases terminate the respiratory chain in many eukaryotes and prokaryotes as the final electron acceptors. They catalyze the reduction of molecular oxygen to water, and conserve the free-energy by proton pumping across the inner mitochondrial membrane or plasma membrane of bacteria. This leads to the generation of an electrochemical gradient across the membrane, which is utilized in the synthesis of ATP. The catalytic mechanism of oxidase is a complex coupling of electrons and protons, which has been studied with the help of numerous biophysical and biochemical methods. The superfamily of oxidases is classified into three different subfamilies; A-, B- and C-type. The A- and B-type oxidases have been studied in great depth, whereas relatively less is known about the molecular mechanism of distinct C-type (or *cbb*$_3$-type) oxidases. The latter enzymes, which are known to possess unusually high oxygen affinity relative to the former class of enzymes, also share little sequence or structural similarity with the A- and B-type oxidases. In the work presented in this thesis, C-type oxidases have been studied using a variety of computational procedures, such as homology modeling, molecular dynamics simulations, density functional theory calculations and continuum electrostatics.

Homology models of the C-type oxidase correctly predicts the side-chain orientation of the cross-linked tyrosine and a proton-channel. The active-site region is also modelled with high accuracy in the models, which are subsequently used in the DFT calculations. With the help of these calculations it is proposed that the different orientation of the cross-linked tyrosine, and a strong hydrogen bond in the proximal side of the high-spin heme are responsible for the higher apparent oxygen affinity and a more rhombic EPR signal in the C-type oxidases. Furthermore, the pK_a profiles of two amino acid residues, which are located close to the active-site, suggest a strong electron-proton coupling and a unique proton pumping route. Molecular dynamics simulations on the two-subunit C-type oxidase allowed for the first time to observe redox state dependent water-chain formation in the protein interior, which can be utilized for the redox coupled proton transfer.

Chapter 1

Introduction

Life on earth first appeared around 3.8 billion years ago in the form of a single-celled organism, which contained essential genetic and energy-generating machinery. Since then the basic form of life has continuously evolved into the more complex multi-cellular organisms of today. Under increasing levels of atmospheric oxygen and a growth in energy demands, many organisms evolved in such a way that they switched their fermentation- or light-based energy production to highly efficient oxygen-based energy production, that is aerobic respiration.

Bioenergetics is a field of biochemistry that is concerned with the energy production in living organisms, which is fundamental to the growth and development of any organism. Bioenergeticists endeavors to explain how energy stored in light ($h\upsilon$) or in chemical bonds of small organic compounds is converted into a form usable in cells, such as Adenosine triphosphate (ATP). In this chapter a concise introduction of the synthesis of ATP through aerobic respiration or oxidative phosphorylation is given.

1.1 Oxidative phosphorylation

Adenosine triphosphate (ATP) is the energy currency of the cell. Hydrolysis of ATP yields adenosine diphosphate (ADP) and a molecule of inorganic phosphate (P_i). Due to excessive resonance stabilization of the orthophosphate group, the reaction is exergonic ($\Delta G°$ and $\Delta G<0$), and releases energy *ca.* 7 kcal/mol under standard conditions. The free energy available is then spent in performing many metabolic biochemical reactions in the cell [1]. Hydrolysis of ATP also occurs to yield one molecule of adenosine monophosphate (AMP) and two molecules of orthophosphates. This exergonic reaction is also utilized to power biochemical reactions such as protein synthesis [1]. ATP is predominantly produced in all aerobic organisms by glycolysis, Krebs cycle and oxidative phosphorylation. Relative to the amount of ATP produced during glycolysis and Krebs cycle, oxidative phosphorylation turns out to be the major source of energy in the cell [1].

In oxidative phosphorylation the reduced co-enzymes NADH and $FADH_2$, which are produced during glycolysis and Kreb's cycle, are oxidized. The electrons then traverse through a number of transmembrane (TM) protein complexes before being ultimately consumed in the reduction of molecular oxygen. Electron transport that takes place in oxidative phosphorylation is coupled to proton pumping across the inner mitochondrial membrane or plasma membrane of

bacteria, thereby generating an electrochemical gradient ($\Delta\mu$) or proton-motive force (*pmf*). The $\Delta\mu$ is then utilized by F_1F_O-ATP synthase to synthesize ATP in the cytoplasm or matrix of mitochondria (Figure 1.1). The complete array of enzymes in the membrane is also commonly called the electron transport chain (ETC) or the respiratory chain (Figure 1.1).

To quantify the ATP production during oxidative phosphorylation, H^+/ATP or P/O ratios are usually examined. The H^+/ATP ratio corresponds to the number of protons pumped across the membrane per ATP molecule synthesized, whereas the P/O ratio corresponds to the number of ATP molecules synthesized or P_i utilized, relative to the oxygen reduction (2 electrons per O atom). According to recent estimates H^+/ATP is 10/3, whereas P/O is 2.3 or 1.4, depending upon if NADH or succinate is the electron donor [7].

The generation of electrochemical gradient due to the movement of electrons and protons from the opposite sides of membrane was first suggested in 1961 by Peter Mitchell in his Chemiosmotic hypothesis [8]. According to Mitchell, the gradient is subsequently used in producing ATP through an ATP synthesis machinery [8]. After initial reluctance, the hypothesis was formally accepted. However, experimental data accumulated over the years suggest that respiratory complexes I and IV do not follow Mitchell's redox-loop mechanism (section 1.2, Figure 1.1) [9,10].

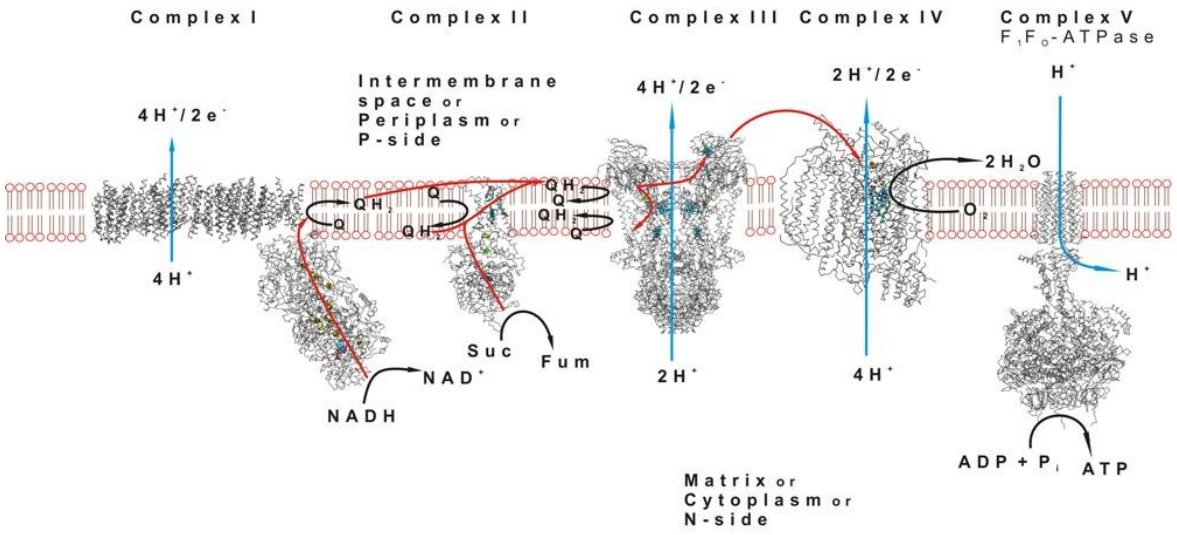

Figure 1.1 Schematic view of the electron transport chain (ETC) in mitochondria. Complex I (PDB id: 3M9S) [2]), complex II (PDB id: 1ZOY [3]), complex III (PDB id: 1KYO [4]), complex IV (PDB id: 1OCC [5]) and complex V (PDB id: 2XND [6]) are shown embedded in the lipid bilayer. Blue vertical arrows show the proton translocation across the membrane by the respiratory complexes, and red arrows show the electron transfer path. Other ancillary reactions are shown with black arrows. Suc, succinate; Fum, fumarate.

1.2 Respiratory complexes

In the following section a brief introduction to each member of the respiratory chain is given.

Complex I is the first enzyme in the respiratory chain to receive electrons from NADH. Electrons then traverse through an array of metal-centers present in the hydrophilic arm of the enzyme before finally being consumed in the reduction of ubiquinone to ubiquinol (Figure 1.1). The movement of electrons and reduction of ubiquinone is coupled to proton-pumping across the membrane at a ratio of $4H^+/2e^-$ [11]. The recently solved crystal structure of complex I from *Thermus thermophilus* and yeast *Yarrowia lipolytica* has shed some light on the possible coupling between the electron transfer and proton translocation [2,12].

Complex II is the second member of ETC. It oxidizes succinate into fumarate, and electrons are eventually used in the reduction of ubiquinone to ubiquinol (Figure 1.1). A cofactor (FAD), chain of iron-sulphur clusters and a heme group are responsible for the intra-protein electron transfer. The enzyme is unique for two reasons; *a)* besides having a role in ETC, it is also active in the Krebs cycle, and *b)* it is the only enzyme in ETC, in which intra-protein electron transfer is not coupled to any proton translocation [1].

Complex III functionality is based on a two-step Q-cycle mechanism [13,14], in which one of the two electrons from the oxidized ubiquinol is transferred to the mobile electron carrier cytochrome *c*, whereas the other one reduces a bound ubiquinone molecule. Subsequently, in a similar next step, another ubiquinol molecule is oxidized, which reduces another cytochrome *c*, and the bound ubisemiquinone radical (Figure 1.1). Both these steps are coupled to proton translocation, which is stoichiometrically described as $2H^+/2e^-$.

Since Complex IV is the main topic of the thesis work, it is described in detail in Chapter 2, though functioning of complex V is briefly discussed here. Complex V is a biological rotor that is responsible for the coupling of $\Delta\mu$ to ATP synthesis [15]. The movement of protons down-hill allows the rotation of membrane bound *c*-rotor (F_O), which in turn rotates the central stalk located inside the hydrophilic domain (F_1). This causes a conformational change in the nucleotide binding sites located in the F_1 domain leading to the generation of ATP (Figure 1.1). Complex V is a reversible machine, such that, ATP hydrolysis can be utilized to pump protons across the membrane. Recent crystal structures have shown variations in the subunit composition of *c*-rotor, which has once again turned the debate onto the previously suggested P/O ratios [6].

ETC harnesses the energy available from the movement of electrons from low-potential (320 mV of $NADH/NAD^+$) to high-potential (800 mV of O_2/H_2O), and couples it with proton translocation. ETC in combination with ATP-synthase is one of the most efficient pathways available in aerobic organisms whereas, anaerobic organisms make use of other metabolic pathways, such as fermentation. Moreover, instead of molecular oxygen as the terminal electron acceptor, other substrates such as sulphates or nitrates are used in the anaerobic respiration [1].

Chapter 2

Respiratory complex IV

Complex IV or Cytochrome *c* Oxidase (C*c*O) safely converts molecular oxygen (O_2) into water and couples the free energy of the reaction to proton translocation across the membrane. It consumes nearly all of the O_2 available in cells, and contributes up to ~40% of the total *pmf* generated by the respiratory chain. In this chapter the structure and function of this 'nanomachine' are described in detail, along with an introduction to the distinct *cbb*$_3$-type C*c*Os, which are the main topic of the thesis.

2.1 Molecular Oxygen

Molecular oxygen (O_2) constitutes 21% (by volume) of earth's atmospheric gaseous composition, and appears as a single most important molecule for the aerobic metabolism in prokaryotes and eukaryotes. In the respiratory chain, O_2 is converted into water and its high oxidizing potential is utilized to generate energy in the form of ATP. O_2 exists in the triplet ground state (*S=1*), which is responsible for its paramagnetic nature and also its stability in the atmosphere despite its very high oxidizing potential. In order to overcome 'spin-forbiddenness' of the reaction of O_2 with any molecular species, conversion of the triplet state into an *open-shell* singlet state (*S=0*), which is higher in energy by *ca.* 22 kcal/mol [16], is necessary. Various metal-containing protein systems (usually *open-shell*) overcome this limiting step by activating O_2 through an electron transfer to it [17], and this mechanism is often explained in terms of spin-orbit coupling [18].

The four electron reduction of O_2 to two molecules of water can be a concerted or a stepwise reaction [19], where formation of the first species O_2^{-*} (superoxide) in the latter mechanism is thermodynamically unfavorable relative to the H_2 oxidation (E_m° = 0 mV) (Figure 2.1). In addition to the 'spin-forbiddenness', negative redox potential of the first reaction is also responsible for the atmospheric inertness of O_2. However, after the formation of one electron reduced species (O_2^{-*}), the subsequent one-electron reduction steps are relatively spontaneous (Figure 2.1), and E_m° of the overall reaction from oxygen to water is *ca.* 800 mV. Addition of one electron at a time in a stepwise reaction also produces reactive oxygen species (ROS), such as HO^* or O_2^{-*}, which are very strong oxidants [20].

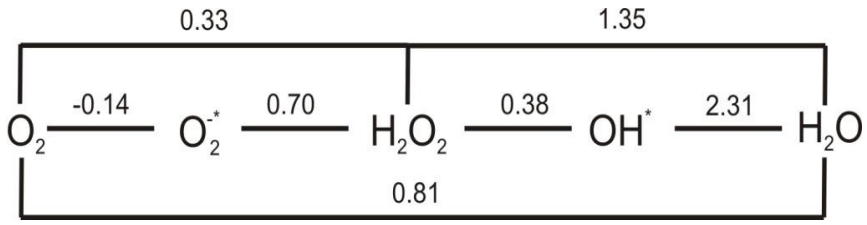

Figure 2.1 Stepwise one-electron reduction of O_2 to water. The standard redox potentials (at 25 °C and pH=7) associated with each reaction are shown in Volts. The figure is based on references 21 and 22.

2.2 Cytochrome *c* oxidase (C*c*O)

Complex IV or cytochrome *c* oxidase (C*c*O) catalyzes the reduction of O_2 to water (Equation 2.1).

$$O_2 + 8H_N^+ + 4e^- \longrightarrow 2H_2O + 4H_P^+ \qquad 2.1$$

The reaction is exergonic and free-energy available is utilized for proton-pumping across the inner-mitochondrial or bacterial membrane [21,23]. The four electrons required for reaction are supplied by the soluble (or membrane bound) electron carrier cytochrome *c*. The protons are transferred through well-defined proton channels that comprise polar amino acid residues and water molecules (Figure 1.1 and 2.2A) [21,23]. The substrate oxygen (O_2) diffuses through the membrane (Figure 1.1), and reaction occurs at a highly conserved active site formed by a bimetallic center (Fe_{a3}/Cu_B) (Figure 2.2B). The time length of the reaction is in milliseconds, and is performed with high efficiency and no ROS formation. For each electron and proton consumed for oxygen reduction, a proton is pumped to the P-side of the membrane (H_P^+). The $\Delta\mu$ thus contributes to the synthesis of ATP. In the sections below a general overview of the classification of oxidase superfamily is given, along with an emphasis on the structure and mechanism of oxygen reduction and proton pumping in C*c*O.

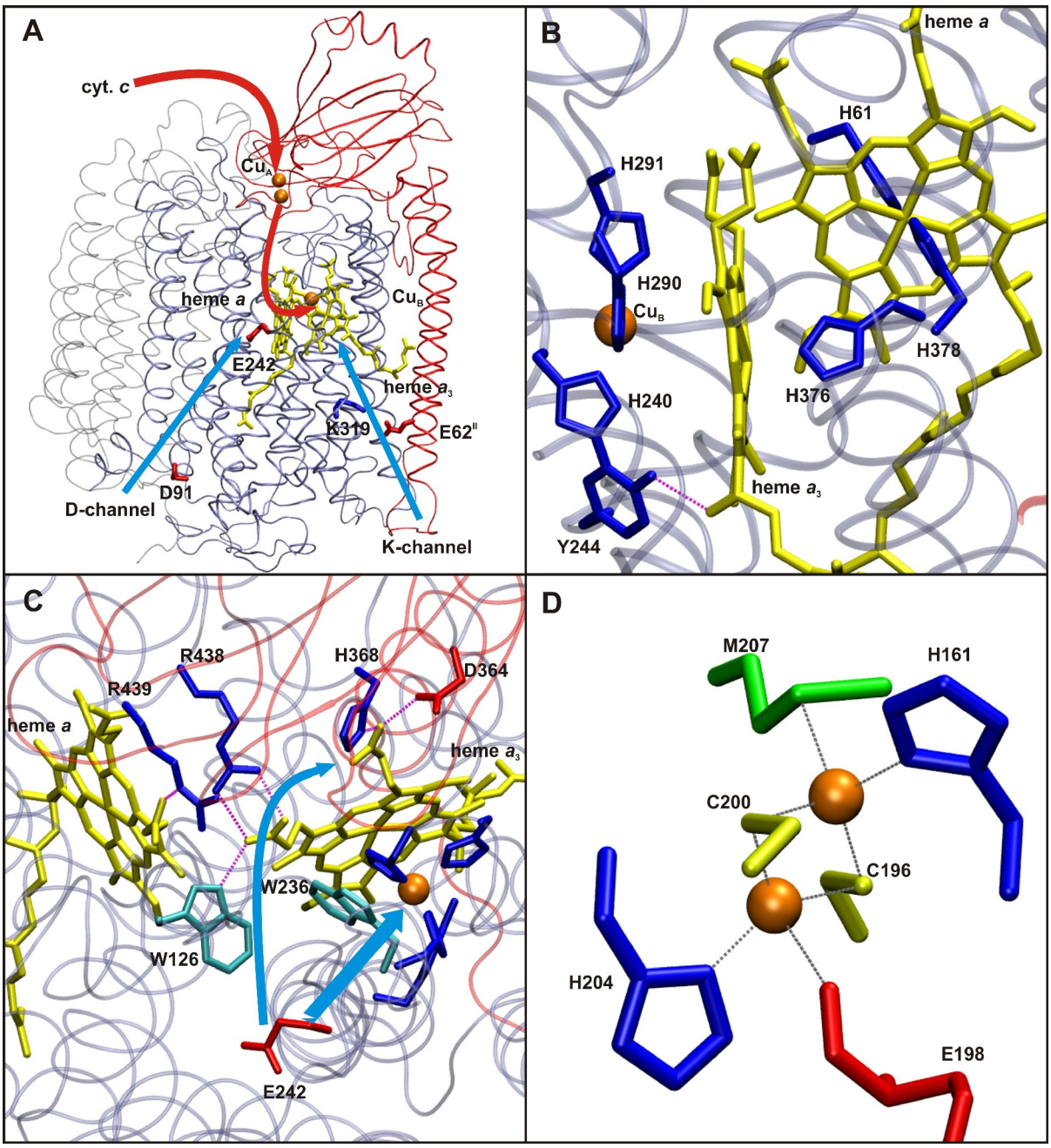

Figure 2.2 A) Structure of the A-type C*c*O from bovine heart mitochondria [50]. PDB id: 2EIJ. Subunit I (ice-blue), II (red) and III (gray, translucent) are shown in ribbon representations. Two proton channels and the electron transfer path are displayed with blue and red arrows, respectively. Few critical amino acid residues are also shown along with the metal centers. B) Binuclear center (BNC) formed by the high-spin heme (a_3) and Cu_B, and the electron queuing low-spin heme (*a*), are shown in licorice representations. Hemes are in yellow color and Cu_B in orange. The histidine ligands of metals centers are shown in blue color. C) View from above the propionates showing the propionate-protein interactions. The proton transfer paths, either to the BNC for oxygen reduction or to the proton-loading site (PLS) for pumping, are shown by the blue arrows. D) Dimetallic purple-copper site and ligands. Selected hydrogen bonds are shown with purple dotted lines. Ligation to Cu_A site is shown with gray dotted lines. The amino acid numbering corresponds to the A-type oxidase from *B. taurus*.

2.3 Family classification of heme-copper oxidases (HCOs)

All oxygen reductases, which contain heme-copper binuclear center are collectively known as HCOs. Due to variations in their architecture and functionality, they are often classified into different subfamilies. The most common scheme for classification is based on the similarity of sequences and other functional components present in the catalytic subunit [24,25,26]. This predominantly led to three different HCO subfamilies; A-, B- and C-type, though more elaborated classifications have also been defined [25]. However, a recent study has argued against the further classification of HCO superfamily [27].

The majority of research work has been performed on the A-type oxidases, which are found in all three domains of life. The overall architecture of the enzyme varies between the prokaryotes and eukaryotes. The eukaryotic enzyme consists of additional regulatory subunits in addition to the homologous core subunits [28]. A-type oxidases are usually identified by their characteristic two proton channels (Figure 2.2A) and a conserved motif -G**H**P**E**V**Y**- in subunit I [26]. Due to the large amount of literature available on A-type oxidases, their mechanism of oxygen reduction and proton pumping is described in detail in the subsequent sections.

The second subclass comprise B-type oxidases, which are predominantly found in archaea [24,25], and possess different structural and functional properties as compared to the A-type oxidases [29,30,31]. For instance, recent structural and experimental data on a B-type oxidase from *Thermus thermophilus* suggest that there are critical differences in the proton channel architecture and in the mechanism of proton pumping with respect to the A-type oxidases [32,33]. Another fundamental difference between the B-type and A-type oxidases is that, in addition to oxygen reduction, B-type oxidases catalyze the reduction of nitric oxide (NO) to nitrous oxide (N_2O), albeit at lower rates [34].

The third subclass is formed by C-type oxidases, which are discussed in detail in the sections below.

2.4 Evolutionary origins of HCOs

Several phylogenetic studies have raised the issue of evolution of HCOs. The earliest work by Castresana and Saraste suggested that the denitrification process (production of N_2) preceded the aerobic respiration (reduction of O_2) [35,36]. They proposed that C-type oxidases, which share relatively high structural similarity with the nitric oxide reductases (NORs) were the earliest HCOs, and existed even before the divergence of current forms of life. In their work, both A- and B- types were suggested to be the recent oxidases, which evolved as the oxygen concentration in atmosphere increased, whereas C-type oxidases became limited to specific branches due to large-scale gene losses [35,36]. This hypothesis gains considerable support from

the experiments in which C-type oxidases are found to perform nearly efficient NO reduction [37]. Moreover, based on sequence analysis and catalytic efficiency it was suggested that the C-type oxidases are evolutionarily primitive as compared to the more complex and efficient A-type oxidases [24,26]. However, recent thorough surveys of completed genomes have found a contrasting picture [38]. Since there are no archaeal prototypes for the C-type oxidases, and they are restricted to the proteobacteria kingdom only, such large-scale gene losses are unlikely to occur over the given timescales, and all three subfamilies have different evolutionary origins [38, 39]. Further studies have also argued against the above hypothesis, and it was suggested that A-type oxidases are primitive oxidases and were present even before the divergence of archaea and bacteria [40]. It has also been proposed that B-type and C-type C*c*Os are the results of horizontal gene transfers (HGT), such that few B-type oxidases were transferred from archaea to bacteria, whereas C-type oxidases from exclusive proteobacterial family to other bacterial families [40].

While there are no definite answers to the evolutionary history of oxidases, more thorough work is required in order to accept one of the hypothesis. A recent study reported by Pereira and co-workers led to the development of a database that consists of all the members of HCO superfamily (http://www.evocell.org/hco) [41]. These authors reiterated that the classification of HCOs in three subfamilies (A, B and C-type) is conclusive [27,41].

2.5 Structural and functional aspects of A-type oxidases

With the advancement in techniques of membrane protein purification and crystallization, structures of A-type C*c*Os first appeared in the mid-nineties [5,42]. Currently, there are approximately 50 structures available in the Protein Data Bank (PDB), that comprise oxidases from different organisms, in different redox or ligand-bound states, and at varying resolutions. These also include members from different subfamilies or closely related enzymes [30,43,44]. Although crystal structures have supported the previous mechanistic models, which were based on pure biophysical and biochemical experiments, their analyses have been unable to elucidate the underlying molecular mechanism of C*c*Os [45,46].

The first structure of A-type C*c*O obtained from bovine heart mitochondria was solved in 1995 [47]. It is a multi subunit enzyme in which subunits I-III forms the catalytic core. These three subunits are also structurally and functionally conserved in all the bacterial C*c*Os [42,48]. The other subunits, which are unique to the mitochondrial oxidases possess regulatory functions [28] (section 2.6). The catalytic core is essential for the full activity of the oxidase (eq. 2.1), however, a two subunit enzyme (I and II) has also been successfully reconstituted into the liposomes with near-native catalytic activity [49]. Figure 2.2A shows the overall architecture of the three subunit mitochondrial enzyme.

2.5.1 Subunit I

2.5.1.1 Active-site

Subunit I is the largest subunit present in the enzyme and contains 12 TM helices arranged in a three-fold symmetry [5,42]. It harbors the active-site of oxygen reduction, which consists of a binuclear center (BNC) that comprises heme a_3 and Cu_B (Figure 2.2B). In the fully oxidized 'resting' state of the enzyme, heme a_3 is in high-spin (*S=5/2*) configuration and is ligated by a histidine ligand (AI:His376) on the proximal side (Figure 2.2B), and probably with a water (H_2O) on the other [51]. Under turnover conditions, it most likely remains in the high-spin configuration, although a hydroxyl (OH^-) ligates from the distal side [52,53,54]. The propionate groups of the heme are held intact by the hydrogen bonding to conserved proteinaceous residues, and are thought to be critical in the proton pumping mechanism of C*c*O [21] (Figure 2.2C). In particular, the A-propionate has been suggested to form the proton-loading site (PLS) from where the pumped protons are eventually released to the P-side [21].

The nearby metal center on the distal side of heme is Cu_B, and both metal sites are anti-ferromagnetically coupled (*S=2*) in the fully oxidized state through a bridging ligand [55,56]. Based on the earlier crystal structures it was suggested that the two metal centers may have water/hydroxyl as non-proteinaceous ligands, and with a hydrogen bond between them, anti-ferromagnetic coupling would be achieved that would cause no signal in the conventional EPR experiments [57,58]. Continuum electrostatic calculations also supported this observation [59]. However, high resolution structures and spectroscopic analyses obtained recently ascribes the electron-density between the metals to a peroxide-dianion (O^--O^-) bridge in the resting state of the enzyme [55,60]. Furthermore, structural and computational analysis of the BNC even led to the suggestions that bound ligand may either be a dioxygen molecule or a superoxide formed due to the X-ray radiation [54].

The copper center Cu_B is ligated by three histidines; AI:His240, AI:His290 and AI:His291, one of which makes a covalent attachment to the conserved tyrosine (AI:Tyr244) (Figure 2.2B). The so-called cross-linked tyrosine is a unique feature known to exist in all the members of HCOs. In the A-type oxidase it is located in the helix VI of subunit I, and its phenolic oxygen hydrogen bonds with the hydroxyl of ethyl-farnesyl chain of heme a_3 (Figure 2.2A and 2.2B). Both, the hydrogen bond and the covalent linkage modulates the electron and proton affinities of the tyrosine and the BNC as a whole [61]. It has been suggested that tyrosine plays a crucial role in the initial steps of oxygen reduction chemistry by supplying a proton and an electron to the oxygenous ligand of heme a_3 (see section 2.5.3) [62,63]. The fourth ligation site of Cu_B is presumably occupied by a hydroxyl or a water molecule in the oxidized and reduced case, respectively [59]. However, its true identity in turnover conditions still remains

unresolved [54,64,65,66].

Subunit I also contains an electron queuing site, heme *a*, from which the electrons are transferred to the BNC one at a time (Figure 2.2A and 2.2B). The heme is low-spin (*S=1/2*), and is ligated by two absolutely conserved histidine ligands, AI:His61 and AI:His378 from helices II and X, respectively (Figure 2.2B). Similar to heme a_3, strong ionic interactions exist between the heme-propionate groups and protein residues, thereby modulating its redox potential (Figure 2.2C). It is known for long that this heme plays an important role in the proton-pumping mechanism of C*c*O [67,68,69,70], and many computational studies also support this notion (section 2.5.3) [71,72,73].

2.5.1.2 Proton channels

The movement of electrons and protons in oxidase is a tightly coupled process. The electrons are transferred through the metal centers present in the enzyme, whereas protons are transferred to the BNC for oxygen reduction chemistry ('chemical protons') or to the P-side of the membrane for pumping ('pumped protons') through well-defined proton channels. Two such channels, D-channel and K-channel, are named after conserved amino acid residues, aspartate (D) and lysine (K) (Figure 2.2A).

The D-channel is structurally conserved in the A-type oxidases, where AI:D91 forms the proton-uptake site (Figure 2.2A). Mutation of AI:D91 in bacterial oxidases either to an alanine or an asparagine renders the enzyme with no pumping and with a very little catalytic activity [74]. From AI:D91 onwards the D-channel continues inside the protein through an array of conserved amino acid residues and water molecules. Many of these residues have also been mutated, and are known to be functionally important [75,76,77,78,79,80,81,82]. The channel eventually terminates at another highly conserved acidic residue; AI:Glu242 (Figure 2.2A and 2.2C). This residue is also conserved in all the A-type oxidases. The proton transfer path from AI:Glu242 onwards, *i.e.* either to the P-side for pumping or to the BNC for oxygen reduction was unclear in the earlier crystal structures (Figure 2.2C). However, computational investigations suggested that dynamic water molecules might be present in the region above AI:Glu242, and could assist in proton transfer [83,84,85,86]. This proposal gained considerable support from the recent crystal structures in which a water molecule was indeed observed in the region [87]. The absence of water molecules in the region above glutamate is presumably due to the non-polar nature of the cavity, which causes high conformational entropy in the water molecules, consequently they fail to resolve with the conventional X-ray diffraction studies. AI:Glu242 has also been a subject of many experimental and computational studies [88,89,90,91,92,93,94,95]. Its side-chain which points towards the N-side ('down') in most of the crystal structures, undergoes rotational

isomerization during molecular dynamics (MD) simulations to form an 'up' conformation, from which it could release protons either to the BNC or to the P-side [92,94]. Recent crystal structure has revealed that AI:Glu242 can indeed form an 'up' conformation [96], thereby supporting observations from the MD simulations. Based on biophysical experiments on the enzyme under steady state conditions, the pK_a of glutamic acid has been found to be relatively high (>9), and it has been suggested to remain protonated throughout the catalytic cycle due to a protonic equilibrium with the N-side (fast reprotonation) [97,98,99]. The fast isomerization of the deprotonated AI:Glu242 to its 'down' position has also been suggested to be critical in preventing proton-leaks [72,93]. Data from various experiments suggest that all pumped protons are taken from the D-channel [99,100], and that it also supplies two protons to the BNC for oxygen reduction [82].

The other channel is the K-channel, which is based on a conserved lysine (AI:Lys319) (Figure 2.2A), and transfers at most two protons to the BNC with no involvement in proton pumping [82,100]. The mutation of lysine to methionine has shown diminished electron transfer activity to the BNC, which suggests a critical role of this channel in proton transfer to the BNC [82]. The mouth of the channel is formed by a conserved acidic residue from subunit II, AII:Glu62 [101], whereas it terminates at another highly conserved and crucial residue, the cross-linked tyrosine (Figure 2.2A). Due to lack of water molecules in the region from lysine to tyrosine, it has been difficult to envisage a continuous proton transfer path. However, MD simulations on water-loaded crystal structure have shown the formation of stable water-chains, which would allow efficient proton-transfer in this channel [102]. A strong electrostatic interaction also exist between the lysine and the proton-uptake site, which may be important in redox-dependent proton transfer through this pathway [103].

2.5.2 Subunit II

Subunit II consists of two TM helices, and a β-sheet structure, which lies above subunit I in the hydrophilic domain of the membrane (Figure 2.2A). It harbors the purple-copper site, a dimetallic copper center (Cu_A), which receives electrons from the mobile electron-donor cytochrome *c* and donates them to the low-spin heme *a* (Figure 2.2D and 2.2A). It has been suggested that a charged surface near the Cu_A site is responsible for the tight but transient binding of cytochrome *c*, which would enable efficient electron transfer between the two [104]. A conserved tryptophan (AII:W104) has also been found to be critical in the bacterial oxidases for the electron transfer from cytochrome *c* to Cu_A [105]. As discussed in the previous section, a proton-uptake site (AII:Glu62) is also located in one of the TM helices of this subunit, which enables an uninterrupted supply of protons to the BNC [101].

2.5.3 Mechanism of oxygen reduction and proton pumping

The catalytic cycle of A-type C*c*O is shown in Figure 2.3, followed by a discussion.

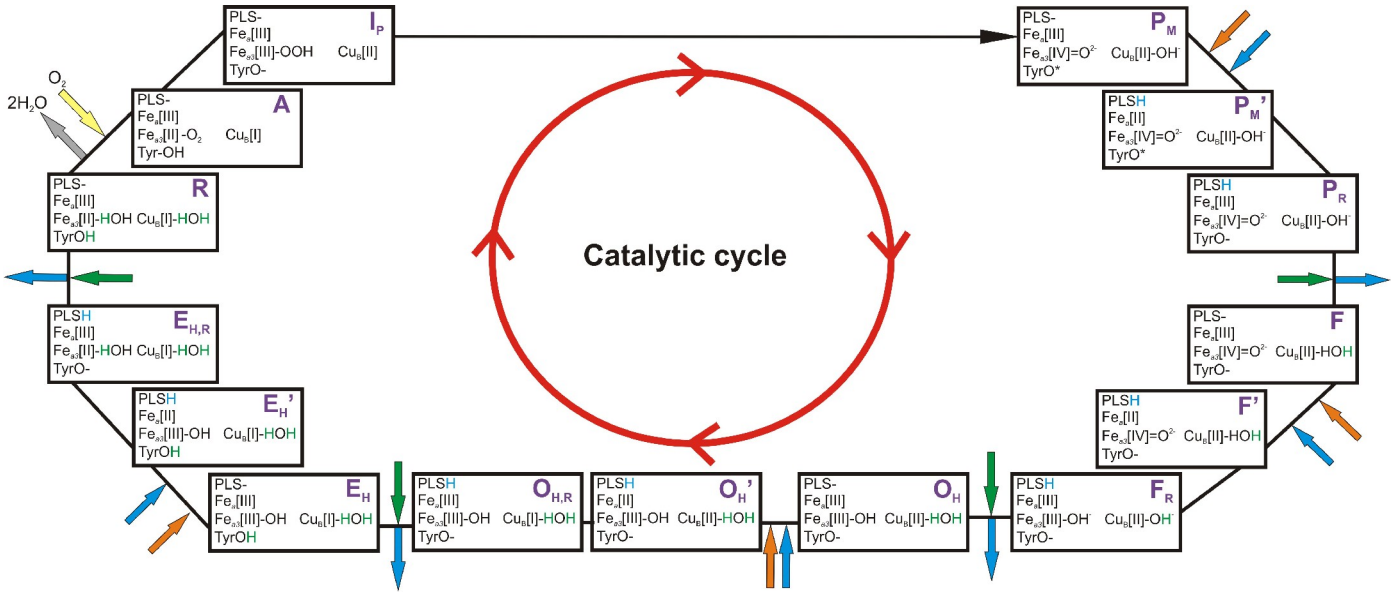

Figure 2.3 Proposed catalytic cycle in the A-type C*c*Os. Various intermediate states are shown, with changes in oxidation states of low and high spin hemes, cross-linked tyrosine and Cu_B. Putative proton acceptors at BNC are shown along with the PLS protonated or deprotonated. Inwardly directed orange arrows show the electron input to heme *a* (from Cu_A). Inwardly and outwardly directed blue arrows show the uptake and release of proton from the PLS, respectively. Green arrows show the proton transfer to BNC. The first two protons to the BNC are supplied by the D-channel, whereas the last two protons are supplied by the K-channel (all green arrows). Two water molecules escape from the cavity in the **R** state, and dioxygen binds to form the state **A**. In each box the first Fe atom on the second line belongs to heme *a* and second Fe atom in the third line belongs to heme a_3.

The reaction cycle shown in Figure 2.3 begins from a mixed-valence (MV) form of the enzyme (**R**) in which BNC is reduced, heme *a* and Cu_A are oxidized and PLS is deprotonated. Formation of MV state from the fully oxidized state ($\mathbf{O_H}$) is coupled to the uptake of two protons from solution, and electroneutrality of the BNC is maintained [106,107,108,109,110]. Molecular O_2 binds weakly to the reduced heme iron (Fe[II]), either replacing the bound CO or under physiological conditions, a weakly bound water molecule. Since experiments are usually performed on the CO poised enzyme, the CO is flash photolysed in the presence of oxygen and state **A** is formed [111], which can be described by an admixture of Fe[II]-O_2/Cu_B[I] or Fe[III]-O_2^{-*}/Cu_B[I] [112] (see below). In the latter arrangement, an electron transfer from the heme iron has activated the oxygen for further scission [17]. Based on spectroscopy measurements and chemically constructed active-site models of C*c*O, it has been suggested that the dioxygen molecule ligates only the heme iron, and does not form a 'μ-peroxo' bridge between the two metals Fe and Cu_B [113,114]. However, the role of Cu_B[I] in enhancing oxygen binding to the heme has also been implicated [114]. The state **A** has been found EPR (electron paramagnetic resonance) silent [112], which further suggests that either the low-spin heme iron is ligated with

a *singlet* oxygen as in Fe[II]-O_2/Cu_B[I], or in the case of Fe[III]-$O_2^{-/*}$/Cu_B[I], low-spin heme iron is anti-ferromagnetically coupled with a *doublet* oxygen. Quantum chemical studies also suggest a *open-shell singlet* ground-state for the intermediate **A** [115,116,117]. Additionally, analogous studies on myoglobin like systems, in which there is no adjacent copper site, also advocate similar electronic structure of the heme-oxygen system [118,119,120].

In the next 200 μs, an electron and hydrogen atom transfer takes place from the Cu_B and the cross-linked tyrosine, respectively, which results in the O-O bond cleavage. The state formed is called $\mathbf{P_M}$, and is characterized by Fe[IV]=O^{2-}, Cu_B[II]-OH^- and a neutral tyrosyl radical ($TyrO^*$) [121,122]. The cross-linked tyrosine plays a critical role at this stage of the reaction by providing an electron and a proton. Catalytic importance of this residue is known from its mutation to phenylalanine in the bacterial oxidases, which leaves the enzyme with no catalytic activity [123,124,125,126], whereas the chemical evidence for the existence of histidine-tyrosine cross-link comes from the mass spectrometric analysis [127]. Hence, all four electrons required for the reduction of O-O are consumed from the BNC itself, *i.e.* two from iron (Fe[II] → Fe[IV]), one from copper (Cu_B[I] → Cu_B[II]) and another one from the cross-linked tyrosine (Tyr → TyrO*). This reaction (**A** → $\mathbf{P_M}$) is spontaneous and irreversible, and does not lead to the formation of any biologically toxic ROS. The state $\mathbf{P_M}$ is also EPR silent, with possible anti-ferromagnetic coupling between the two metal-sites (Fe[IV]=O^{2-}:*S=2* and Cu_B[II]/TyrO*:*S=1*) [17,63]. The two hydrogen bonds, *a)* between the distal oxygen atom of oxo-ferryl species with the hydroxyl bound on Cu_B, and *b)* between the phenolic oxygen of tyrosine and hydroxyl of ethyl-farnesyl chain of heme a_3, could play critical roles in achieving such coupling. In all quantum chemical calculations performed on BNC models, such scenarios have been modeled and adequate energetics have been obtained [128,129]. The transition from **A** → $\mathbf{P_M}$ is not coupled to proton uptake or proton pumping, while the high oxidizing potential of molecular oxygen is transferred to the enzyme in the form of high oxidation states of iron of heme a_3 (Fe[IV]), copper (Cu_B[II]) and the tyrosine residue (TyrO*). Subsequent reduction of each center drives the proton-pumping mechanism in C*c*O. The four, one-electron reduction steps, $\mathbf{P_M}$ → **F**, **F** → $\mathbf{O_H}$, $\mathbf{O_H}$ → $\mathbf{E_H}$ and $\mathbf{E_H}$ → **R** are each coupled with a proton consumption at the BNC and proton pumping to the P-side via PLS (Figure 2.3).

As early as 1981 the intermediates of the O_2 splitting reaction were first described by Wikström, which consisted of **P** (peroxy/OOH^-) and **F** (oxo-ferryl) intermediates [130]. However, correct identities of these states became clear at a later stage when it turned out that in the **P** state, O=O bond is already broken, and it has an oxo-ferryl character [121,122]. Later on, experiments and subsequent modeling of the reaction from **A** → $\mathbf{P_M}$ also suggested a presence of 'peroxy' intermediate [131]. This state was called $\mathbf{I_P}$, and has not been observed experimentally

so far due to its high energy and lower occupancy [131]. Furthermore, based on quantum chemical calculations, the likely structure of $\mathbf{I_P}$ intermediate was found to be Fe[III]-OOH^-/Cu_B[II]/$TyrO^-$ (Figure 2.3) [128,129] and it was suggested that one electron each from the iron and copper and a proton from the cross-linked tyrosine has been transferred to the oxygenous ligand of heme to form the peroxide [128,129,131]. Even though the peroxy state $\mathbf{I_P}$ has never been observed experimentally in the mechanism of HCOs, its possible existence has been shown in the cytochrome *bd*-type oxidases [132].

Each of the four proton pumping steps shown in Figure 2.3 are assumed to proceed in a similar way. However, identities of the final electron and proton acceptors in each step are different. During each pump step the free-energy differences diminish, such that proton pumping in the $\mathbf{P_M \rightarrow F}$ and $\mathbf{F \rightarrow O_H}$ transitions is thermodynamically more favorable compared to the $\mathbf{O_H \rightarrow E_H}$ and $\mathbf{E_H \rightarrow R}$ transitions. With the help of fast electron injection in the $\mathbf{O_H}$ ('pulsed') state of the enzyme, combined with absorption spectroscopy and time-resolved electrometry, $\mathbf{O_H \rightarrow E_H}$ transition was studied in great depth and allowed the tracking of individual electron and proton transfer events in the proton pumping mechanism of A-type C*c*O for the first time [133,134]. This work led to the proposal of a pumping mechanism as described below, which would be valid for both bacterial and mitochondrial A-type oxidases [86,98,136].

A single electron injected into the enzyme (in $\mathbf{O_H}$ state) equilibrates in 10 μs between the initial electron acceptor Cu_A and electron queuing heme *a* in the ratio of 30:70. At this stage there is almost no reduction of BNC due to its low redox potential. The slower proton uptake is not observed in this step due to the measurement of fast electron transfer from Cu_A to heme *a*. Nevertheless, it is known that redox potential of heme *a* is pH dependent [135]. Next the electron equilibrates between the heme *a* and BNC in 150 μs, in the ratio of 40:60. The mid-point potential of the BNC is raised from its previous value most probably due to the loading of proton-loading site (PLS), which presumably lies closer to the BNC (or heme a_3). The following slower phase of 800 μs (or nearly a ms) in which electron completely moves to the Cu_B and leaves all other redox centers oxidized is due to the high redox potential of Cu_B [133,134]. This huge increase in potential can be ascribed to the protonation of BNC. A last phase with a smaller amplitude corresponding roughly to 2.3 ms is also observed, which is most likely the release of proton from the PLS to the P-side [133,134]. Analysis of the potential generation across the membrane identified two phases; 150 μs and 800 μs, which corresponds to the transfer of 'pumped' and 'chemistry' protons to the PLS and Cu_B, respectively. It is assumed that this pumping mechanism repeats four times, once for each transition.

Experimental data shows that heme *a* is the driving element in the pumping process of C*c*O with its equilibrium redox potential being pH dependent [67,68,69,135]. However, the

location of the site which would be protonated upon reduction of heme *a* remains unknown. Based on theoretical modeling of experimental parameters [136] and computational calculations [137,138], it has been proposed that A-propionate of heme a_3 is most probably the protonatable site upon reduction of heme *a*. It has also been proposed that PLS may not be just a single site but a cluster of residues on which the charge of a proton is smeared [98].

Assuming a tight coupling between the PLS and metal centers, the following sequence of events has been modeled in many different computational studies [72,86,93]. *a)* Upon reduction of heme *a*, PLS gets protonated through a proton-uptake from the solution ($\mathbf{P_M'}$, **F'**, $\mathbf{O_H'}$ and $\mathbf{E_H'}$ states in Figure 2.3), which in turn increases the redox potential of heme a_3/Cu_B system due to its close proximity to the latter. This allows electron transfer from heme *a* to the BNC ($\mathbf{P_R}$, $\mathbf{F_R}$, $\mathbf{O_{H,R}}$ and $\mathbf{E_{H,R}}$ states in Figure 2.3), which is further coupled to another proton-uptake from the N-side. The charge compensation at the BNC, pushes out the proton from the PLS to the P-side of the membrane [136] (Figure 2.3). One of the critical aspects of this mechanism, which is based on thorough experimental findings, is that the 'pumped' proton moves to the PLS prior to the 'chemistry' proton in the BNC. However, models have also been presented in which 'chemistry' proton is first transferred to the BNC [45].

2.6 Nuclear and mitochondrial coded subunits: Regulation

The A-type mitochondrial oxidases consist of three subunits coded by the mitochondrial DNA (subunits I-III) and 10 subunits by the nuclear genome (subunit IV-XIII). The mitochondrial coded subunits are also present in the homologous bacterial oxidases (section 2.3), whereas nuclear coded subunits are unique to the mitochondrial oxidases only, and play a critical role in the regulation of oxidase activity [139]. Specifically, the subunits IV and VI of A-type oxidase from *B. taurus* are known to modulate the activity of the oxidase, which depends upon the [ATP]/[ADP] ratio in the mitochondrial matrix [140]. The decrease in catalytic activity of oxidase upon higher [ATP]/[ADP] ratio is either achieved by reducing the H^+/e^- ratio or by decreasing the affinity for electron-donor cytochrome *c* [140]. Allosteric inhibitory mechanisms have also been suggested to cause the modulation in activity of C*c*O when an ATP molecule binds to one of the accessory subunits [141]. Furthermore, phosphorylation of threonines, serines and tyrosines located on the matrix-side of the enzyme also probably alters the catalytic activity through an allosteric inhibition mechanism [140].

Currently, mitochondrial C*c*O is a subject of numerous studies due to its implications in neuro-degenerative disorders, mitochondrial diseases and other pathological processes, and has been approached via multiple experimental techniques. Moreover, it is known to be the only enzyme in the mitochondrial respiratory chain that has tissue-specific isoforms in addition to

having developmental stage-specific isoforms [140]. This finding further highlights the medical importance of the enzyme.

2.7 *cbb*$_3$- or C-type oxidases

The third dominant subfamily in the HCO kingdom is formed by the C-type oxidases, which are predominantly found in proteobacteria [142]. These high affinity oxidases are expressed in low oxygen tensions, and convert free energy of oxygen reduction to proton pumping with the same stoichiometry as their A-type counterparts do [143]. However, recent studies provide a contrasting picture by questioning the pumping stoichiometry of 1 H^+/e^-, and suggesting it to be only 0.5 H^+/e^- [144].

Initial biochemical experiments identified that these enzymes are encoded by an operon called *fix*NOQP (or *cco*NOQP), which was named after the first isolation of these enzymes from the nitrogen-fixing bacteria [145,146,147]. This operon is always found upstream (towards the 5' end) to another operon called *cco/fix*GHIS, which codes for the proteins responsible for the assembly of oxidase [148]. The structural gene *cco/fix*NOQP encodes four subunits; N, O, P and Q. The largest subunit N shares the overall fold with the catalytic subunit of A-type oxidases (section 2.5.1). It consists of 12 TM helices and conserved histidine residues that ligate the metal centers (Figure 2.4A and 2.4B) [149]. The two other subunits (O and P) are unique to the C-type oxidases, and do not resemble any of the non-catalytic subunits of A- or B-type oxidases. The subunit O consists of a TM helix and a cytochrome *c*-type fold connected through a highly conserved linker (Figure 2.4A) [43]. The heme *c* has methionine and histidine axial ligands, together with the covalently bound cysteines. The cysteines and histidine ligand together forms the highly conserved **C**xx**CH** motif, commonly found in the *c*-type cytochromes [150]. The other non-catalytic subunit P consists of two cytochrome *c*-type folds that are fused together through a conserved linker (Figure 2.4A) [43]. It also contains two conserved **C**xx**CH** motifs, for each heme *c* present in the subunit. The smaller and less conserved subunit, Q, is not important for the catalytic activity or for the assembly of the enzyme, but it has been suggested that it performs a protective role under oxic conditions [151].

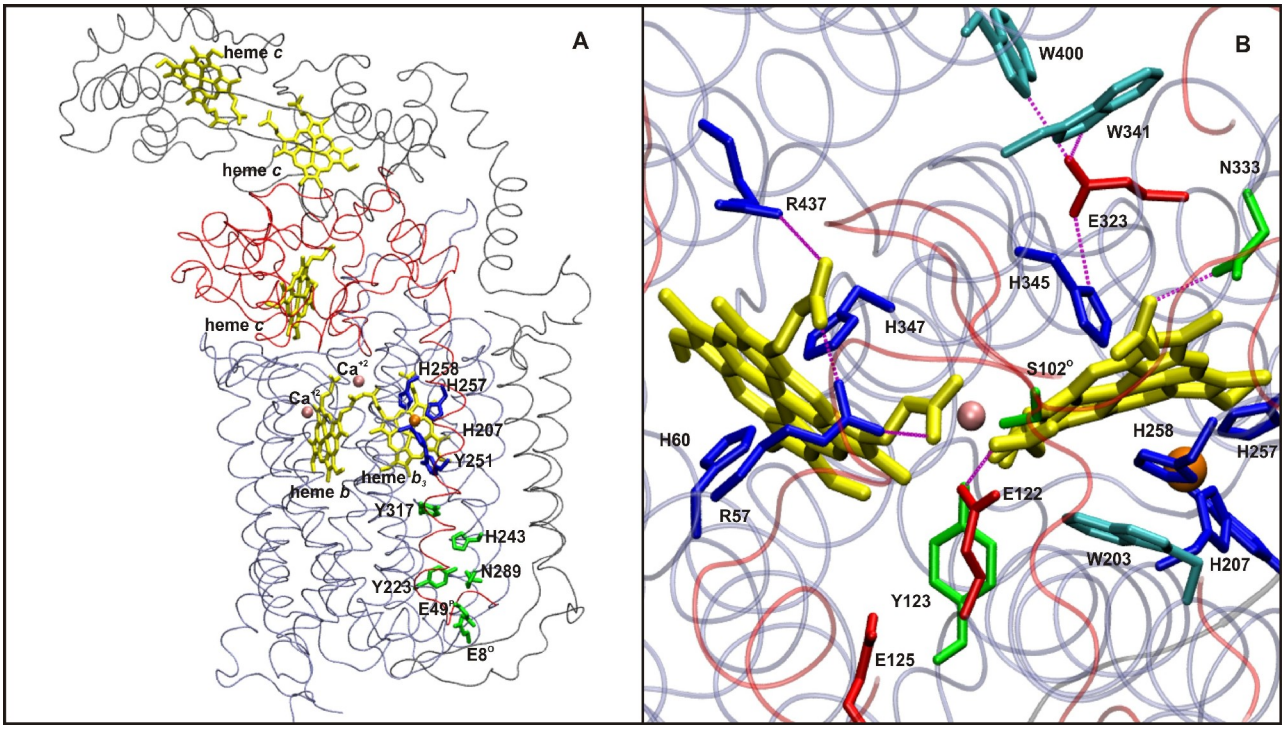

Figure 2.4 Crystal structure of C-type oxidase from *P. stutzeri*. PDB id 3MK7 [43]. A) Subunit I/N (ice-blue), II/O (red) and III/P (gray) are shown in ribbon representations, along with critical amino acid residues which form the putative proton channel. B) View from above the heme propionates show the metal ligands and other important residues surrounding the BNC. Metal-centers are shown in licorice representations; hemes in yellow, Cu_B in orange and calcium ions in pink. Selected hydrogen bonds are also shown as purple dotted lines. Amino acid numbering corresponds to the *cbb*$_3$-oxidase from *P. stutzeri*.

It was found based on biophysical experiments that the catalytic subunit N consists of a low-spin electron queuing heme *b*, and a BNC formed by high-spin heme b_3 and Cu_B [152], which are most likely arranged in a fashion that is similar to the A-type oxidases (section 2.5.1). Earlier studies also suggested critical differences in the active-sites of A- and C-type oxidases, especially as was found that binding of CO to heme b_3 is more linear in the C-type oxidases, which was suggestive of a more 'open' BNC cavity [153]. Furthermore, CO recombination (post-photolysis) kinetics was also observed to be faster in the C-type oxidases, and thereby emphasizing varying active-site architecture in the two subfamilies [154,155].

The crystal structure of C-type oxidase from *Pseudomonas stutzeri* has recently been solved [43]. Nevertheless, computational modeling and biochemical experiments performed prior to the crystal structure yielded valuable insights. Homology models of the *Rhodobacter sphaeroides* enzyme suggested that a conserved tyrosine from the helix VII could form a covalent attachment to the histidine ligand of Cu_B [126] (publication I,II). This role is performed by a tyrosine from helix VI in the A- and B-type oxidases. Later on, site-directed mutagenesis

studies, mass-spectrometric analysis and X-ray structure confirmed the presence of a histidine-tyrosine cross-link as predicted (Figure 2.4A and 2.4B) [43,156,157,158]. Similarly, the prediction of a proton channel in *cbb*$_3$-oxidase (publication I) was confirmed by the mutagenesis studies and X-ray structure (Figure 2.4A and 2.4B) [43,156]. Based on the EPR measurements on wild-type and mutant enzymes in different redox states, it was found that a glutamate residue hydrogen bonds to the histidine ligand of heme b_3 [159], alike to the scenario in the catalytic triad of cytochrome *c* peroxidases ('Fe-His-Asp') [160]. The recently solved crystal structure also confirms this prediction (Figure 2.4B) [43].

The crystal structure shows some critical structural differences when compared to the structure of A-type oxidases. First, distance between the hemes in the catalytic subunit is relatively shorter in the C-type oxidases, which would most likely alter the intra-heme electron transfer rates and the kinetics of oxygen trapping [43]. Second, the proteinaceous partners of heme propionate groups are also different in the C-type oxidases [43] (Figure 2.4B). In particular, the D-propionates of both hemes ligate to a single Ca^{++} ion, whereas in the A- or B-type oxidases, a pair of arginine residues neutralizes the negative charge of the propionates [42,43]. Furthermore, the proton channel identified in the location analogous to the K-channel of A-type oxidases is also structurally different (Figure 2.4A).

Based on mutagenesis studies it was suggested that B-type oxidases use a single channel for transferring both 'chemical' and 'pumped' protons [32]. Similar proposals have been made for the *cbb*$_3$-oxidases, which immediately leads to the question of evolution of proton channels in the three different subfamilies, and in particular, about the proton channels responsible for pumping. C-type oxidases also reduce NO [37], but do so without coupling the reaction to proton pumping, and it has even been found that this catalytic reaction is not electrogenic, due to the uptake of both protons and electrons from the same side of the membrane *i.e.* the P-side [161,162]. This also poses a question about the proton transfer path from the P-side, which could be utilized for proton pumping (in a reverse way). Furthermore, biophysical experiments found that the redox potential of heme b_3 is uniquely low (-60 mV), in sharp contrast to the E_m° of analogous heme in the A-type oxidases [135,159]. Though some structural reasons for this stark difference have been identified, it further raises questions about the differences in thermodynamics of electron transfer in the two enzyme subfamilies.

In many bacteria that belong to the *Helicobacter* genus and *Campylobacter* genus, the electron transport chain terminates at the C-type oxidases as the sole electron acceptors [163]. These bacteria are known to be responsible for many pathological conditions, which in turn presents a unique and novel scenario for the targeted drug design.

2.8 Nitric oxide reductases

Nitric oxide reductases (NORs) form another family of enzymes that are responsible for the reduction of nitric oxide (NO) to nitrous oxide (N_2O). These enzymes share local sequence similarities with the C-type oxidases (*ca.* 18% for the catalytic subunit) and are also capable of reducing O_2 to water, albeit at lower rates [164,165,166]. The NOR enzyme comprises two subunits, one heme *c* containing subunit NOR*c*, which is analogous to the subunit O of the C-type oxidases, and the main catalytic subunit which comprises heme *b* and b_3-types. Heme b_3 along with a non-heme Fe forms the EPR-silent dinuclear center in which the catalytic reaction takes place. The structure of the NOR*bc* complex from *Pseudomonas aeruginosa* has recently been solved at 2.7Å resolution [167], and shows high local structural similarity with the structure of C-type oxidase from *Pseudomonas stutzeri*. The dinuclear active site contains a non-heme Fe, which has an unusual glutamate residue ligation, in addition to the three histidines. The presence of a charged amino acid near the dinuclear center causes low redox potential of the heme b_3 [167]. It has also been proposed recently that heme b_3 is in low-spin configuration, contrary to the popular perception of it being high-spin [168]. The catalytic reaction leads to the formation of N-N bond in N_2O, which requires two electrons and two protons along with two substrate (NO) molecules. It has been observed that the reaction is not electrogenic, and both protons and electrons are taken from the P-side of the membrane [169,170]. The recently solved structure of enzyme also reveals that the proton uptake channels from the N-side, which are active in the C-type oxidases, are not conserved in NORs, whereas there are traces of proton pathways from the P-side [167].

Chapter 3

Multiple sequence alignments and homology modeling

Currently there are about 20,000 unique entries in the Protein Data Bank (PDB), calculated after removing redundant protein sequences with 50% or above similarity. On the other hand, the 'RefSeq' database (http://www.ncbi.nlm.nih.gov/RefSeq/) at the National Center for Biological Information (NCBI) lists 13,137,813 protein sequences obtained from nearly 16,000 organisms. Despite the advancements in the methodologies of genomics and proteomics, this shows that the protein structure data lags behind the sequence data. In order to overcome this huge gap, protein structure prediction methods play a crucial role, and are commonly used in the field of drug discovery or *de-novo* protein design. In this chapter a brief introduction to the methods of multiple sequence alignments and homology modeling is given.

3.1 Sequence alignments

Sequence alignments provide multiple solutions in the realm of protein or gene sequence analysis. They help in the identification of homologous sequences, their comparison, validation and phylogenetic analysis, and structure modeling. Classically, pairwise sequence comparisons were performed with the help of 'dot-plots' [171], where two sequences were aligned on a two-dimensional matrix and amino-acid matches were observed on the diagonal. However, such methods provided redundant comparisons, and gaps (insertions and deletions) were not taken into account explicitly, which resulted in suboptimal alignments.

Currently, pairwise sequence alignments are performed by using two methods: *global alignment* and *local alignment*. Global alignments are suited for two sequences that are similar in length and have relatively high similarity, whereas local alignments are more useful in identifying the regions of local similarity in two divergent sequences. Local alignments are often more difficult to perform.

Two sequences can be optimally aligned with the help of global alignment methodology implemented in the Needleman-Wunsch method [172]. This method is based on the 'dynamic programming' approach, in which maximum value of the alignment score is achieved based on the predefined substitution matrix, gap and gap-extension penalties. Although robust, this method is slow, particularly when large sequence data have to be analyzed. On the other hand, the Smith-Waterman algorithm, which is a variation from the Needleman-Wunsch algorithm, can be reliably used to obtain local alignments in the cases of low sequence similarity [173].

However, due to the slower performance of aforementioned methods, faster heuristic methods have been developed which are capable of treating larger sequence data [174]. These heuristic approaches are based upon the progressive alignment techniques in which pairwise alignments of more similar sequences are first generated, and then based on phylogenetic tree constructions more divergent sequences are aligned [174,175]. Since accuracy of an alignment falls with the increase in divergence between the sequences, other constraints can be used to generate initial pairwise alignments as accurately as possible. It has been suggested that sequence identity of less than 30% may lead to large errors in an alignment [175]. Therefore, newer methods have been proposed, which often yield alignments of high quality. Two such methods are, iterative refinement and consistence scoring [176,177,178,179]. In the iterative refinement procedure the alignment is broken down into smaller subsets and re-alignments are performed until the score is optimized. For the consistence scoring method, the initial scoring is performed by taking multiple sequences into account, rather than just two.

CLUSTAL [180] and T-COFFEE [179] are popular sequence alignment programs based on progressive alignment techniques.

3.2 Structural alignments

It is well-known that a protein structure can remain conserved during the evolution, even though variations may occur in its sequence. Structural alignments take into account the three dimensional structure of a protein and have been shown to greatly improve the quality of the sequence alignment [175], specially in the cases of distantly related proteins. Web-based methods such as Basic Local Alignment Search Tool (BLAST) [181] can also be adjusted to take into account the structural information during the sequence search. CLUSTAL and 3D-COFFEE [182] too provides a methodology that generates multiple sequence alignments by including structural information. The high gap and gap-extension penalties in structurally conserved regions do not allow insertions to occur, whereas in the loop regions such insertions and deletions occur freely due to lower sequence and structural homology.

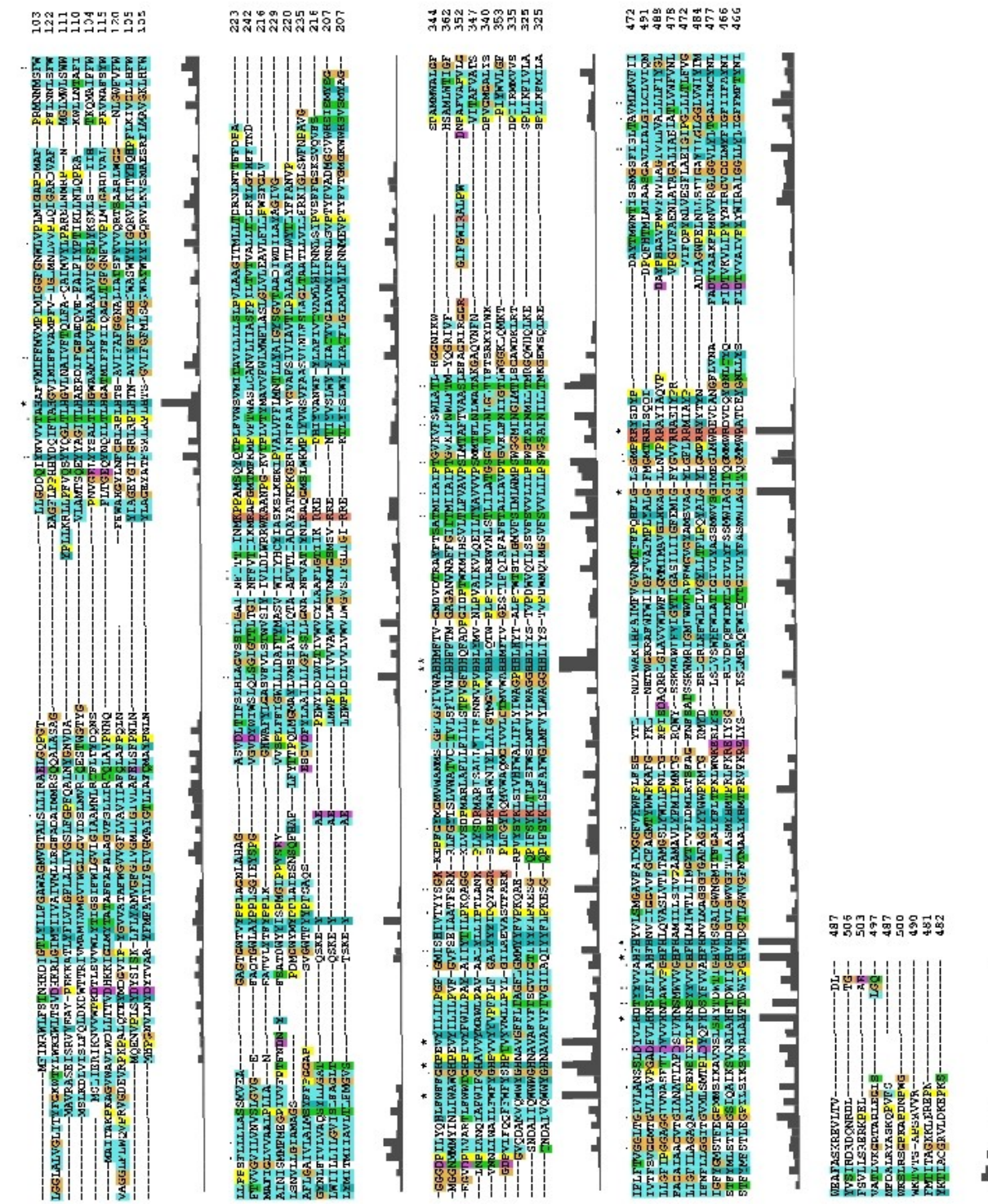

Figure 3.1 Figure showing multiple sequence alignment of subunit I from A, B and C-type oxidases. A-type oxidases from *Bos taurus* (B_t), *E. coli* (E_c), and *Sulfolobus acidocaldarius* (S_a); B-type from *T. thermophilus* (T_t) and *A. ambivalens* (A_a); and C-type from *R. sphaeroides* (R_s), *C. jejuni* (C_j) and *H. pylori* (H_p). Additionally, a *caa*$_3$-type oxidase from *T. thermophilus* is also shown in the alignment.

3.3 Protein Structure Prediction

The tertiary or quaternary structure of a protein determines its function. Hence, in order to understand its biochemical operations, a high resolution structure is often required. The most commonly used methods for protein structure determination are X-ray crystallography and Nuclear Magnetic Resonance (NMR). Both these methods are time consuming and expensive in order to be used for each of the protein sequence available. Furthermore, problems associated with the protein purification, low yields, and impurity of samples, makes the task of structure determination especially challenging for the membrane proteins. In addition, the static nature of X-ray structures and technical difficulties associated in resolving the structure of high-energy intermediates necessitates the use of protein structure prediction methods, which strive to model the protein structures as close to their native states as possible.

According to the classical definition, protein structure prediction methods can be broadly divided into three different categories; *a)* Template-based modeling (Homology modeling), *b)* Fold-recognition and *c) ab-initio* or *de-novo* or free modeling [183]. However, all three methods are sometimes used in tandem in specific cases of protein modeling [183,184]. To serve large-scale protein structure prediction endeavors, such as modeling of all the members of G-protein coupled receptors (GPCRs), requires web-based automated servers [185]. These servers are often based on the above listed methods or some combination of them. They are also capable of generating protein models with minimal manual intervention, and have shown great success in the recent Critical Assessment of protein Structure Prediction (CASP) experiments [186]. Based on the CASP results over the years, the most successful method for protein structure prediction is the template- or homology-based modeling [187,188]. Both web-based and standalone tools are available to perform the homology-based modeling.

3.3.1 Homology modeling

Homology modeling is based on the principle that two proteins that share a common evolutionary origin are homologous to each other, and may also have a high structural similarity. The sequence of the protein for which a three dimensional structure is sought (target sequence), is modeled by using the structural information obtained from the homologous protein (template), aided with the sequence alignment between the two. The sequence and structure of the template protein can be optimally determined with the help of many web-based tools [179,189]. The most commonly used method is BLAST, or its variants, in which a target sequence is submitted as a 'query', and the complete sequence database is searched. Based on the search results, possible templates are identified along with a pairwise or multiple sequence alignment. Fold-recognition methods are also based on the similar principles, in which a model of the target sequence is

constructed, in addition to the identification of template(s). The Fold-recognition methods make use of pre-defined sequence-structure relationships [190,191]. Many web-based servers that perform either task of template search or fold-recognition are available [192,193], and a number of programs are listed here: http://www.russelllab.org/gtsp/foldrec.html.

In homology modeling sequence alignment between the target and template sequences play a critical role. It has been suggested that at least 40% sequence similarity is required, in order to generate quality homology models [194]. In the case of lower sequence similarities (known as 'Twilight zone'), an optimal alignment is often difficult to obtain, and main-chain and side-chain positions are likely to be modeled incorrectly [195]. It is often necessary to generate larger multiple sequence alignments in order to obtain optimal alignments between the template and target sequences. This allows clear identification of regions of high and low sequence conservation [192]. Additional input of structural information in the form of constraints is also known to increase the success rate of homology modeling [175].

A general protocol of the homology modeling process can be described as follows, and may require repetition and manual intervention at different stages;

a) generation of structural alignment of the available template structures,

b) alignment of template sequence(s) to the structural information from *a)*,

c) generation of a larger multiple sequence alignment that contains target sequence and other homologous sequences,

d) addition of the alignment *c)* to *b)* through different methods of *sequence-to-profile* or *profile-to-profile* alignments.

Finally, from the larger alignment (step *d* above), the target-template sequence alignment is fed into the homology modeling program for structure generation. *Profile-to-profile* alignments have successfully been used in the past for homology-based modeling. They tend to improve the identification of deviating patterns in the multiple sequence alignment [196,197].

Many different algorithms exist that construct a model for a given target sequence. Some popular algorithms implemented into programs are; satisfaction of spatial restraints (MODELER) [194], segment matching (LOOK) [198], rigid-body superposition (SWISS-MODEL) [199] and artificial evolution (NEST/JACKAL) [200,201].

3.3.1.1 MODELER: Model building and validation

MODELER is one of the most popular programs for homology-based modeling. Many recent surveys have found its enhanced performance over a diverse set of proteins [202,203]. Apart from having a fully automated model generation protocol, MODELER also allows manual intervention, which could be used to improve the quality of the models. In recent CASP

experiments, manual interventions have been shown to improve the success rate of homology models [204,205,206,207,208].

MODELER requires a target sequence, template-target sequence alignment and template structure as inputs for the automated model generation procedure. In difficult cases of modeling, such as when the target-template sequence identity is low (<40%), experiment based input in the form of distance or angle constraints, along with modified MODELER parameters, may improve the quality of models. MODELER works on a principle that is similar to the concept of distance geometry, and combines it with an energy function optimization, which can be compared with the common forcefield-type potential energy function optimization [194]. A brief description of MODELER's functionality is described next.

During a single run of MODELER, 'features' such as, bond lengths, bond angles, main-chain and side-chain dihedrals, Cα-Cα or N-O distances, *etc.* are built depending on the template structure and template-target sequence alignment. These features are then transformed into conditional probability density functions (*pdfs* or p_i), which depend upon the residue type, main-chain conformations, equilibrium values, gaps between residues, *etc*. The conditional *pdfs* are further combined into a single function *P* (of the form, Πp_i) and different types of optimizations are performed. Optimization of the function *F* (= ln*P*), with respect to Cartesian co-ordinates, yields a final model. The lower value of MODELER's objective function (*F*) means lower number of violations in geometry, and this model is more precise among an ensemble of models that are generated. This process can be repeated, and also be combined with the changes in the alignment, especially in the regions of low sequence conservation. Such manual interventions are often required to increase the quality of the models.

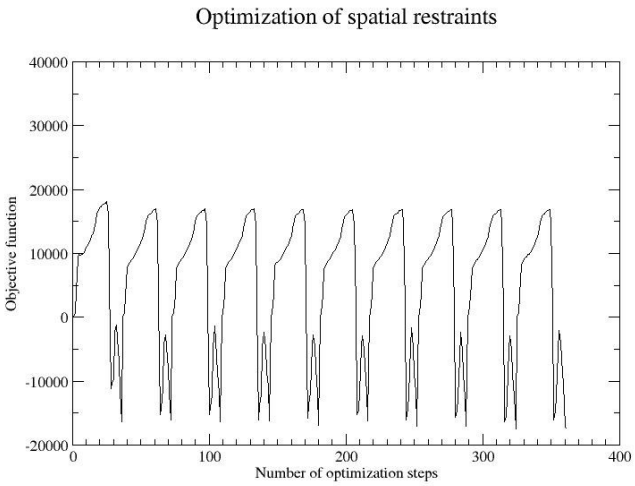

Figure 3.2 MODELER's objective function value *vs* number of optimization steps. 10 optimization cycles are performed to obtain a final model. Each cycle consists of a Conjugate Gradient optimization of the variable target function, followed by Molecular Dynamics Simulated Annealing protocol.

After the homology models have been constructed, different quality checks and structure validations are performed in order to identify potential stereo-chemical errors, errors in

backbone dihedrals, planarity of rings, side-chain packing, atomic contacts, *etc*. Many different programs exist for this purpose; Procheck, Molprobity and Anolea, to name a few [209,210,211,212]. Depending upon the results of the quality check, different types of refinements can be performed including more accurate forcefield based energy minimization and molecular dynamics [213,214,215]. With these refinements, the quality of the model can be improved, though its accuracy is not guaranteed. Currently there are no methods that would identify 'near-nativeness' of a constructed model. It has often been observed that the structure of a modeled protein is more similar to the template, than to its own native structure. The likely reasons for this are the problems associated with the modeling protocols. Despite this, recent CASP experiments have shown progress in this field, where near-to-native models were indeed obtained especially in those cases in which sequence similarity between target and template was high [188]. Along with the improvements in modeling methodology, validation tests have also evolved over time. In recent CASP exercises, modeled proteins were subjected to an ensemble of validation tests, which led to the identification of many false positives [205].

3.3.1.2 MODELER: Problems and Loop modeling

Homology models are sometimes rendered inaccurate due to errors in the template-target sequence alignment. More specifically, the most variable parts of the alignment *i.e*. the loops, are often modeled incorrectly in which the side-chain conformations are predicted erroneously. Furthermore, most of the spatial-restraint violations also occur in such regions in MODELER calculations (publication I). In these cases, it is often advisable to combine data from other sources to the homology modeling program *e.g.* in the form of distance or angle constraints.

When no other data are available, the modeling of loops can also be performed *de-novo* [216], in which loops are modeled by using forcefield based methods, or by searching possible templates from available libraries. Knowledge-based methods have also been highly successful in predicting correct side-chain dihedrals. These methods are often used in conjunction with homology modeling, in which side-chains are modeled by using side-chain rotamer libraries [217]. Both backbone-dependent and backbone-independent rotamer libraries are available [217,218].

Chapter 4

Density functional theory

The last two decades have seen a substantial growth in the use of Density Functional Theory (DFT) in modeling biochemical problems. This has been possible due to the higher accuracy and lower computational cost of DFT, relative to many *ab-initio* methods. The advent of novel algorithms, highly parallelized programs and increase in computational power have enabled the routine treatment of metal-containing biological systems up to hundreds of atoms. One hybrid density functional (B3-LYP) has dominated the field of biomolecular computations due to its superior performance in calculating accurate energetics of the reactions. In this chapter, basic principles of DFT, its functionality and performance are briefly discussed.

4.1 Early ideas in DFT

Early work by Fermi and Thomas led to the foundations of DFT [219,220]. In their theoretical model kinetic energy (T) of the system was expressed as a functional F of electron density or particle density $\rho(r)$.

$$T = F[\rho(r)] \qquad 4.1$$

where $\rho(r) = \int \Psi^* \Psi \, dv$ and Ψ is the wavefunction. 4.2

However, their model failed to provide a greater accuracy due to the incomplete description of kinetic energy and exchange energy, and absolute omission of electron correlation. Later work by Hohenberg and Kohn laid the foundations of contemporary DFT calculations [221], by which it was suggested that if ground state electron density of a system is known, the minimal energy and other properties can be readily evaluated.

4.2 Formulation of DFT

DFT, though inherently different from Hartree-Fock (HF) theory, still share substantial similarity with the latter. In the sections below, basic formulation of DFT is briefly discussed along with the critical differences to HF theory.

4.2.1 The Schrödinger equation and the Self-consistent Field method

The Schrödinger equation after Born-Oppenheimer approximation [222] can be written as,

$$\left(\sum_i \frac{-\hbar^2}{2m_e} \nabla_i^2 - \sum_{i,A} \frac{Z_A}{r_{iA}} + \sum_{i,j} \frac{1}{r_{ij}} \right) \boldsymbol{\Psi} = \boldsymbol{E} \boldsymbol{\Psi} \qquad 4.3$$

Here, the first term from left corresponds to the kinetic energy of electrons, the second term

corresponds to the Coulombic interaction between the electrons (i) and nuclei (A), and third term is the electron-electron interaction. Physical constants and Coulombic interaction between nuclei are omitted for simplicity. Solutions of eq. 4.3 are readily available if the third term is ignored, which would correspond to no electron interaction. Energy ($\boldsymbol{E}$) in such a case is high and incorrect, and is equal to the sum of energies of individual electrons. Moreover, the solutions ($\boldsymbol{\Psi}$) are eigenfunctions of individual uncorrelated electrons, known as the Hartree Product. Therefore, explicit treatment of the last term is thus necessary in order to obtain the correct $\boldsymbol{E}$ and $\boldsymbol{\Psi}$, which complicates the solutions of eq. 4.3 in a multi-electron system (i>1). All methods such as DFT, HF or post-HF, which aim to provide solutions to the multi-electron Schrödinger equation (eq. 4.3), are based upon the self-consistent field (SCF) method. In this method the wavefunction ($\boldsymbol{\Psi}$), which is expressed in a Slater-determinant formalism [223], is varied until the energy of the system is minimized below a given threshold value.

Equation 4.3 results in two 'two-electron energy' terms. The first two-electron integral is a Coulombic integral J_{ij}, and represents the repulsion between two electrons,

$$J_{ij}=\sum_i \sum_j \int\int (\Psi_i(r_i))^2 \frac{1}{r_{ij}} (\Psi_j(r_j))^2 dv_i d\omega_i dv_j d\omega_j \qquad 4.4$$

In HF theory $\Psi_i(r_i)$ are the spin-orbitals, and in DFT $(\Psi_i(r_i))^2$ is the electron density (see eq. 4.2). The second two-electron integral is the Exchange integral (K_{ij}), which appears in HF theory only.

$$K_{ij}=\sum_i \sum_j \int\int \Psi_i(r_i)\Psi_i(r_j)\frac{1}{r_{ij}}\Psi_j(r_j)\Psi_j(r_i) dv_i d\omega_i dv_j d\omega_j \qquad 4.5$$

It has been observed that this term provides a correction to the over-estimated Coulombic term (eq. 4.4), but does not capture the electron correlation effects compared to the DFT or post-HF methods [222].

The total energy in HF method is thus given by,

$$E=\sum_i h_i+\sum_i \sum_j (J_{ij}-K_{ij}) \qquad 4.6$$

and, in DFT,

$$E=\sum_i h_i+\sum_i \sum_j J_{ij}+E_{XC}[\rho(r)] \qquad 4.7$$

All terms in eq. 4.7 are the functions of electron density. The first term on the right (h_i) collectively represents the kinetic energy of electrons in a non-interacting system (T), along with the nuclei-electron interaction (V). The second term is the classical Coulombic electron-electron interaction (U) and the third term contains all the remaining contributions for which the exact formulation remains unknown. If the last term exchange correlation energy (E_{XC}) is known, the exact energy of the system can be obtained. It is understood that E_{XC} comprises of contributions

from kinetic energy (in an interacting system), in addition to exchange and correlation contributions [224].

4.2.2 Kohn-Sham (KS) orbitals

The equation 4.7 is solved with the use of Lagrange multipliers under the constraint that integration of electron density over space yields a total number of electrons in the system. This yields KS equations,

$\hat{f_i^{KS}}\Psi_i=\epsilon_i\Psi_i$, where *f* is the one electron operator $\hat{f_i^{KS}}=\hat{h_i}+\sum_j \hat{J}_{ij}+\hat{V_{XC}}$

The KS equations are solved iteratively until the predefined threshold is reached. The required initial density is usually constructed from the simpler methods such as Extended-Huckel method [225] or from the lower level of theory. It has been recognized that final $\boldsymbol{\Psi}$ obtained after solving analogous HF equations does not represent the correct wavefunction due to the missing electron correlation. In such a case it is more like a non-interacting system, and energies reported are also not exact [224]. In the case of DFT, approximate functions are used to describe the exchange-correlation potential, V_{XC}. These approximations (section 4.4) have given impressive results and current research in the field of DFT is primarily aimed at improving the formulations of V_{XC}. The solutions of KS equations are known as KS orbitals (Ψ_i) [226], which are inherently different from the HF orbitals, though their use in calculating different properties of molecules to a high accuracy has been reported [227,228,229].

4.3 Basis sets

The spatial part of spin-orbital $\Psi(r,w)=\chi(r)\alpha(w)$ can be expanded as a Linear Combination of Atomic Orbitals (LCAO),

$$\chi_l(r)=\sum_m c_{lm}\phi_m$$

where c_{lm} are the coefficients and Φ_m are the basis functions [222,230]. Φ_m are not necessarily orthogonal, but may be normalized. During a SCF iteration, energy of the system is minimized by varying the coefficients until thresholds for change in energy and coefficients are reached.

Earlier basis functions were analogous in formulation to the hydrogenic atomic orbitals, and were called as Slater-type orbitals (STOs) [231]. However, Gaussian-type of orbitals (GTOs) were later preferred [232], due to the complications in solving two-electron integrals (eqs. 4.4 and 4.5). The GTOs can be expressed as,

$$\phi(r_1-r_c)=N(x_1-x_c)^i(y_1-y_c)^j(z_1-z_c)^k e^{-\alpha(r_1-r_c)^2}$$

Here *N* is the normalization constant, x_1, y_1 and z_1 are the co-ordinates of electron, x_c, y_c and z_c are

co-ordinates of the center, *i*, *j* and *k* are the non-negative integers and α is a positive exponent. Multiple GTOs are often added together to approximate a single STO, a process known as 'contraction' [233,222]. For example, a STO approximated by *n* GTOs can be written as,

$$\phi(r)_i^{STO-nG} = \sum_n d_{in}\psi_n(r)$$

The contraction coefficients d_{in} and the exponential constant α are usually obtained after iterative refinement procedure [222]. When compared to the simpler and less accurate STO-nG type of basis sets, split-valence (SV) type of basis sets are currently used more often. In SV basis sets, core and valence electrons are treated separately. One such commonly used basis set is 6-31G* (Pople notation [234,235,236]). In this notation the first number corresponds to the number of basis functions (or gaussian primitives) used to define the core electrons *e.g.* the *1s* orbital. The last two digits denote the number of two separate contraction schemes that are used to treat the valence electrons. For example, 3 Gaussian primitives, and separately 1 Gaussian primitive for the valence electrons (*2s* orbital onwards). These basis functions are often added with polarization (*) and diffuse functions (+) in order to provide a more realistic description of chemical bonding [236].

4.4 DFT: Exchange Correlation (XC) functionals

In HF or post-HF methods the results can be improved systematically by increasing the size of the basis set or by increasing the level of theory. However, there exist no systematic way of improvement in DFT, and quality of results mainly depends upon the formulation chosen for E_{XC}.

4.4.1 Local (Spin) density approximation (LSDA)

This is the most basic type of approximation for XC functional. If the electron density varies very slowly (uniform electron gas), the E_{XC}^{LDA} can be written as,

$E_{XC}^{LDA} = \int \rho(r) e_{XC}(\rho)\, dr$ or in the case when system has unpaired electrons (*open-shell* system),

$$E_{XC}^{LSDA} = \int \rho(r) e_{XC}(\rho_\alpha + \rho_\beta)\, dr$$

where $e_{XC}(\rho)$ is the exchange-correlation energy per particle weighted with the density ρ(r) [226]. In metal systems or in proteins, in which the variations in electron density are far more drastic, the LDA fails to give satisfactory results [237,238]. For example, when this approximation is used in calculating the thermochemical data on the G2 set of molecules, average error as much as by 36 kcal/mol is observed in the energies [239]. However, it performs

satisfactorily in evaluating the equilibrium structures and frequencies of molecules [224]. The commonly used DFT functionals associated with L(S)DA approximation are; S-VWN [240] and PW-LDA [241].

4.4.2 Generalized gradient approximation (GGA) or meta-GGA functionals

Inclusion of the gradient of density (∇) into the functional led to a great improvement in the results [224]. Energy in this case is given by,

$$E_{XC}^{GGA} = f(\rho_\alpha, \rho_\beta, \nabla(\rho_\alpha), \nabla(\rho_\beta))dr$$

while in the case of meta-GGA,

$$E_{XC}^{meta-GGA} = f(\rho_\alpha, \rho_\beta, \nabla(\rho_\alpha), \nabla(\rho_\beta), \nabla^2(\rho_\alpha), \nabla^2(\rho_\beta))\,dr$$

Here second derivatives (∇^2) are also included in the calculations. The well-known functionals associated with GGA are B-LYP, B-P86 [242,243,244], whereas with meta-GGA; TPSS [245,246,247].

4.4.3 Hybrid functionals

The E_{XC} in hybrid density functional also consists of the exact exchange interaction calculated from the HF theory (eq. 4.5). One of the most popular hybrid density functional is B3-LYP [243,248,249],

$$E_{XC}^{B3LYP} = (1-a_0)E_X^{LSDA} + a_0 E_X^{HF} + a_X \Delta E_X^{B88} + a_c E_C^{LYP} + (1-a_C)E_C^{VWN}$$

where constants a_0, a_x and a_c are 0.2, 0.72 and 0.81, respectively. Each of them were obtained by fitting the calculated heat of formation of small molecules to the experimental data [249]. The E_X^{LSDA} describes 80% exchange interaction in a fully interacting system, whereas 20% comes from the HF theory. ΔE_X^{B88} is the gradient correction to the exchange term, and E_C^{LYP} and E_C^{VWN} consists of the correlation and its gradient correction, respectively.

4.5 Functionality in DFT

4.5.1 Coulombic term

The one-electron integrals are calculated in a similar way in both HF theory and DFT, and computational demand for these integrals is also relatively low. However, the most demanding calculation is the numerical computation of two-electron integrals (eqs. 4.4 and 4.5). In the HF method, these integrals are explicitly evaluated, whereas in the DFT, the inherent separation of the Coulombic and Exchange terms in the form of J and E_{XC} simplifies and speeds-up the calculations. Furthermore, the Coulombic term is handled by the methods of density fitting in DFT, where M^4 dependency is reduced to M^2N by the use of auxiliary basis sets [250,224]. Here,

M is the number of basis functions, and *N* is the dependency of density that is obtained when atom-centered basis functions are replaced with an approximate density (eq. 4.2). This has been shown to provide a speed-up of one order of magnitude with little loss of accuracy [251]. In a popular DFT program, TURBOMOLE [252], this approach is used as 'resolution-of-identity (RI)-J, where *J* stands for Coulombic term [224]. Other DFT codes also utilize similar methods. Moreover, many other advanced grid-based or grid-free numerical integration methods have been developed and are used in many DFT codes [224].

4.5.2 Solvation effects in DFT

Solvation effects are usually modelled by mimicking the dielectric continuum of proteins. A commonly used methodology is the Conductor-like screening model (COSMO) [253]. In this method the solvent is treated in a scaled conductor manner such that the polarization of the solvent due to the solute is modelled by scaling the screening charges on the cavity surface by a factor of $f=(\frac{\epsilon-1}{\epsilon+0.5})$, where ϵ is the dielectric constant. The interaction energy between the solute and the scaled surface charges is then evaluated and optimized during the SCF iterations. The cavity surface can be represented by the Van der Waals surface of the molecular system. The COSMO solvation model has been implemented in many quantum chemistry programs [253,254,255] and adequate results have been obtained for both high and low dielectric systems [253,254,255].

4.6 Performance of DFT

The power of DFT lies in eq. 4.2, which allows the reduction of dimensionality of the problem by expressing *3N* Cartesian co-ordinates of electrons into just three spatial variables of density.

4.6.1 Geometry optimization

Highly accurate molecular structures and ground-state energies have been obtained based on the geometry optimizations using DFT [256,257,258,259]. It has been observed that the basis set requirements are not stringent in the DFT-based geometry optimizations. However, at least split-valence (SV) or triple-zeta valence (TZV) quality basis sets are recommended to achieve high-quality results such as in transition metal containing systems [256]. On the other hand, the choice of the functional has been found to be critical. The performances of GGA (BP86) and meta-GGA functionals (TPSS) are superior to the commonly used hybrid functional B3-LYP, especially in the case of first-row (*3d*) transition metal containing systems [260,261,262,263]. The GGA functional BP86 is a popular choice of method for geometry optimizations due to its

high computational performance (section 4.5.1). Recent research has also shown the improvement in quality of optimized structures with the newer meta-GGA type TPSSh functional, which incorporates 10% exact exchange [260].

4.6.2 Enzyme mechanism modeling

Due to its high accuracy and low computational cost, DFT has been the workhorse of enzyme mechanism modelling. Large cluster model systems of the enzyme active sites that contain *ca.* 200 atoms have been used in the DFT calculations in order to assess the energetics of reaction [264]. It has been shown that with a sufficiently large cluster model a convergence in energies can be obtained [264,265,266]. However, universal application of this approach has been debated, and it has been found that the convergence with the system size is slow to achieve specially in comparison to the hybrid quantum mechanics/molecular mechanics (QM/MM) approaches, which are also often more accurate [267]. In the field of enzyme mechanism modelling, the performance of hybrid density functional B3-LYP has been found to be exceptionally superior compared to the GGA functionals such as BP86 or PBE [268]. However, errors in relative energies of up to 5 kcal/mol have also been reported for the transition metal containing systems [268]. For the G2 reference set of molecules, the errors in energies are in the range of 2-3 kcal/mol of the experimental values [268,269]. A higher accuracy is sought in the energetics of an enzyme mechanism modelling, which puts a lower limit on the quality of basis sets to be used. At least TZV quality basis sets are recommended, which could be added with the polarization or diffuse functions [270].

4.6.3 Problems associated with DFT

Despite the enormous success, DFT calculations also suffer with the errors. The three most commonly occurring errors in the DFT calculations are: *a)* Self-interaction error, *b)* errors due to the missing static correlation, and *c)* errors due to the missing van der Waals interactions [268]. Self-interaction errors beyond the one-electron case still remains enigmatic. However, recent developments have highlighted their sources, and improvement in density functionals is an active research field [224]. The errors in static (non-dynamical) correlation stems from the fact that DFT is a single determinant based method, whereas in cases for which near-degenerate determinants are required to define the ground state, multi-configurational SCF (MC-SCF) methods are the popular choice [256]. It has also been suggested that the B3-LYP functional is designed to take care of these two errors, such that they cancel each other out in some cases [268].

The errors sourcing from the incomplete treatment of van der Waals interaction

(dispersion), results in the inconsistencies that are observed during geometry optimizations or single-point calculations. Since these interactions are quantum mechanical in nature, and can be considered similar to the exchange interaction of HF theory (attractive dispersion) [271], empirical corrections to DFT energy have been applied. These corrections have shown improvements in the structure quality and energetics [271].

Chapter 5

Continuum electrostatics

Continuum electrostatic calculations provide a solid and efficient means to calculate the pK_a of a titratable residue in a protein. As the name implies, the solute (protein), and the solvent surrounding the solute *i.e.* either the membrane or the aqueous phase, are modeled as continua with fixed dielectric constants. The Poisson-Boltzmann (PB) equation is then solved in order to calculate the electrostatic potential at a number of grid-points. The electrostatic free energy of placing a point charge on a grid point is thus calculated, from which the pK_a is calculated. This method assumes that pK_a of a titratable residue depends upon the electrostatic interactions only, and all other effects of bond breaking and formation cancels out.

The protonation behavior of a residue in protein may change under the influence of change in redox state, conformational state or chemical state (*e.g.* mutation). Such effects are often successfully modeled with the help of continuum electrostatic methods. In this chapter, applications and limitations of these methods are briefly discussed, along with a theoretical description.

5.1 pK_a calculations : general theory

For any titratable residue at equilibrium,

$HA \rightleftarrows H^+ + A^-$

In this case the acid dissociation constant can be written as,

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

which leads to the well-known Hendersson-Hasselbalch equation,

$$pH = pK_a + \log\left(\frac{[salt]}{[acid]}\right) = pK_a + \log\left(\frac{[A^-]}{[HA]}\right) \qquad 5.1$$

Probability (P), that the titratable residue is protonated can be written as,

$$P = \frac{[HA]}{[HA]+[A^-]} = \frac{1}{1+10^{(pH-pK_a)}}$$

or, $$pK_a = pH + \log\left(\frac{P}{1-P}\right) \qquad 5.2$$

Therefore, in order to calculate pK_a of a residue, its protonation probability P needs to be evaluated, and the pH at which P=0.5, that would correspond to its pK_a [272,273]. This requires a titration to be performed, and it is normal in protein titrations to obtain non-sigmoidal curves,

which most likely reflect the interactions between multiple titration sites [273].

Probability P in eq. 5.2 can be expressed in the form of a canonical ensemble formulation [274,275],

$$P = \frac{\sum_{n=1}^{n=2^N} x_\mu^n e^{-\beta G^n}}{\sum_{n=1}^{n=2^N} e^{-\beta G^n}} \qquad 5.3$$

where

$$G^n = \sum_{\mu=1}^{N} \left((x_\mu^n - x_\mu^o) RTln10 (pH - pK_{intr,\mu}) + \sum_{\nu>\mu}^{N} W_{\mu\nu} (x_\mu^n - x_\mu^o)(x_\nu^n - x_\nu^o) \right) \qquad 5.4$$

Here, G^n is the energy of a protonation state n of the protein and there are 2^N such protonation states. N is the number of titratable residues in the protein, μ (or υ) is the titratable group, and x_μ^n takes the values 1 (protonated) or 0 (deprotonated). The variable x_μ^o is the reference uncharged state of the group μ, with its values 1 or 0, for acids and bases, respectively. Additionally, $pK_{intr,\mu}$ is the pK_a of a group μ when all other titratable sites are in their reference protonation states (neutral). The $W_{\mu\upsilon}$ term is the interaction energy between two charged titratable sites. In order to evaluate G^n, the pK_{intr} and site-site interactions ($W_{\mu\upsilon}$) are first calculated. Since the calculation of P increases enormously with the increase in the number of titratable groups (N), the Metropolis Monte Carlo methods are generally used [275].

The pH independent part, $pK_{intr,\mu}$ in eq. 5.4, is first evaluated from the equation,

$$pK_{intr,\mu} = \frac{\Delta\Delta G_{intr,\mu}}{2.303RT} \quad \text{where} \quad \Delta\Delta G_{intr,\mu} = \Delta\Delta G_{Born} + \Delta\Delta G_{Back} \qquad 5.5$$

$\Delta\Delta G_{Born}$ is the Born solvation energy associated with the transfer of residue (μ) from the aqueous phase (ϵ=80) to the protein phase (lower dielectric), whereas $\Delta\Delta G_{Back}$ corresponds to the interaction of residue (μ) with all the other background charges present, *i.e.* solvent molecules in the aqueous phase, and other charges in the protein [276]. See Figure 5.1

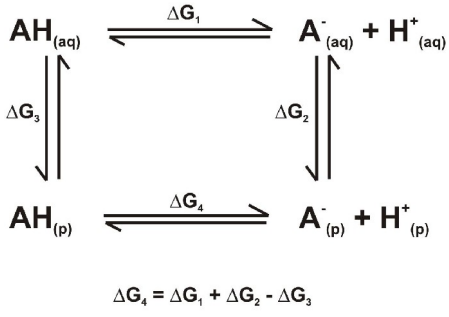

Figure 5.1 The thermodynamic cycle showing deprotonation of a residue in the aqueous phase and in protein. Free energy of desolvation (ΔG_2 or ΔG_3), each consists of 'Born' and 'Background' contributions. ΔG_4 is unknown and ΔG_1 is known.

Each of the contributions in eq. 5.5 are further disaggregated into individual terms,

$$\Delta\Delta G_{Born} = (G_{Born}^{A^-_{prot}} - G_{Born}^{AH_{prot}}) - (G_{Born}^{A^-_{aq}} - G_{Born}^{AH_{aq}}) \quad \text{and}$$

$$\Delta\Delta G_{Back} = (G_{Back}^{A^-_{prot}} - G_{Back}^{AH_{prot}}) - (G_{Back}^{A^-_{aq}} - G_{Back}^{AH_{aq}}) \quad .$$

Next, the pH dependent part $W_{\mu\upsilon}$ (eq. 5.4) is calculated. It corresponds to the site-site interaction energy for when both the sites are charged, and is usually presented in the form of a symmetric matrix where diagonal elements are zero [276].

Both p$K_{\text{intr},\mu}$ (eq. 5.5) and $W_{\mu\upsilon}$ are obtained from the following expression,

$$\Delta G_{es} = \int_V \rho(\boldsymbol{r})\phi(\boldsymbol{r})dV \qquad 5.6$$

where $\rho(\boldsymbol{r})$ is the charge density and $\phi(\boldsymbol{r})$ is the electrostatic potential, which is calculated by solving Poisson-Boltzmann (PB) equation.

5.2 Poisson-Boltzmann equation

The non linearized Poisson-Boltzmann equation can be written as,

$$\nabla(\epsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r})) = -4\pi\rho_{prot}(\boldsymbol{r}) - 4\pi\sum_i z_i q n_o \mathrm{e}^{-\beta z_i q\phi(\boldsymbol{r})} \qquad 5.7$$

where ϵ is the space-dependent dielectric constant, $\rho_{prot}(\boldsymbol{r})$ is the charge density, which in the case of a protein as solute would correspond to the point charges located on atomic nuclei, z_i is the charge of ion i, n_o is the concentration of ion i in an non-interacting system and q is the protonic charge [277,278]. This equation can be linearized by expanding the exponential, and by assuming low concentration of ions and electro-neutrality. It can thus be written as,

$$\nabla(\epsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r})) = -4\pi\rho_{prot}(\boldsymbol{r}) + 4\pi\sum_i (\beta z_i^2 q^2 n_o \phi(\boldsymbol{r})) \qquad 5.8$$

The analytical solutions of eq. 5.8 for complex systems such as proteins do not exist, hence numerical methods are applied. The most commonly used methods are the Finite Difference methods, by which electrostatic potential at a number of grid-points is obtained [279,280,281]. Mapping techniques are then used to calculate the approximate potential at the location of charge in order to obtain the final electrostatic free energy. Moreover, in order to avoid numerical errors that arise from the lower resolution of grid, 'focusing' techniques are applied in which the increased density of grid points at the site of interest is used such as at the location of titratable residue. As discussed in section 5.1, when p$K_{\text{intr},\mu}$ and $W_{\mu\upsilon}$ are obtained, MC sampling is performed to calculate the final pK_a of a residue (eq. 5.2 and 5.3).

5.3 Strength and weakness of continuum electrostatics method

Continuum electrostatic methods have been successfully used on a variety of enzyme systems

[282,283,284,285,286,287,288,289]. Enhancements of these methods by including conformational flexibility also led to improvements in the pK_a predictions [290,291,292,293,295]. Many different categories of enzymes have also been studied including the difficult cases of redox proteins, in which the events that correspond to redox coupled protonation have been successfully reproduced [296,297,298]. It has been suggested that even in the cases when variations in the electrostatic potential are large, the linearized PB equation method can be successfully used to calculate the electrostatic properties [277].

Notwithstanding the success, these calculations also exhibit limitations due to the approximations in the methodology involved [299]. In PB theory it is assumed that the pK_a of a residue is solely dependent upon electrostatic free-energy, whereas the effects of bond-breaking or formation or relaxation are neglected. Such effects may be critical and to disregard them may lead to errors in pK_a predictions. The treatment of dielectric constant inside the protein is an another complicated issue [300]. Although values from 2-5 are commonly used in electrostatic calculations, other values that range from 4-20 have also been used and led to impressive results [300].

The explicit treatment of water molecules in the continuum electrostatic calculations have been recommended specially in the cases of metal-ligation [301,302]. However, it has also been found to be problematic due to the sensitivity of water-bound hydrogen bonds to the redox or conformational changes in the protein [285,301]. Moreover, as there are no straightforward ways to treat the dynamic hydrogen bonds in the continuum methods, it is advisable to avoid the use of explicit water molecules [303].

Many successful formulations have been developed in order to obtain pK_a values averaged over an ensemble of structures. However, difficulties in sampling the conformational space corresponding to the pH range investigated experimentally or the inclusion of non-biased conformational flexibility are currently the bottlenecks [274].

Chapter 6

Molecular Dynamics Simulations

Molecular dynamics (MD) simulations allows the study of time evolution of a system of particles. As system traces its trajectory in the phase-space, with the help of statistical mechanics and mathematical tools, the microscopic properties such as positions and momenta can be associated with the observable macroscopic properties such as temperature and pressure. MD simulations are very popular in the field of computational biochemistry. Different categories of proteins are simulated in a variety of modeled surroundings in order to understand functionally important conformational changes. Quantitative estimates of root-mean square deviation/fluctuation, rate of diffusion, time-resolved fluorescence spectroscopy, side-chain motions, relative free-energies, *etc.* can also be obtained through the equilibrium MD simulations. The range of motions that can be studied using MD simulations is from fast atomic-level fluctuations, to slow domain movements or folding events. The increase in computational power has advanced the simulation times up to microseconds or milliseconds, from which reasonable and more accurate statistics can be obtained. This chapter provides a brief introduction to the MD simulations.

6.1 General Theory

N number of particles with known positions and momenta can be considered as a point on a multidimensional (*6N*) phase space. As time passes the point moves through this space and assembles a trajectory. If the time allowed for propagation is long enough, then it can be assumed that the average of any property over time is equal to its ensemble average. This is called the *Ergodic hypothesis*. Due to this longer MD simulations are often desired in order to obtain a reasonable statistics. However, the size of the system such as a biomolecular system, always acts as a bottleneck.

MD simulations are built on the platform of classical physics, in which the particles are described as spheres with van der Waals radii possessing mass and charge. The bonds, angles, dihedral angles between atoms are imitated as springs, added with non-bonded interactions; van der Waals and Coulombic interactions. From the initially known position (PDB) and momentum (Maxwell-Boltzmann distribution) of each particle, the potential energy of the system can be calculated, and the forces on all atoms are evaluated. Due to the coupled movement of particles and larger number of degrees of freedom, the equations of motion are solved numerically. Newton's equations of motion characteristically conserve energy and momentum and are fully

reversible. Consequently, MD simulations also conserve energy and are in principle reversible. The commonly used form of the potential energy function is,

$$V(r)=\sum K_b(b-b_o)^2+\sum K_\theta(\theta-\theta_o)^2+\sum K_\phi[1+\cos(n\phi-\delta)]+\sum \epsilon_{ij}[\left(\frac{A_{ij}}{r_{ij}}\right)^{12}-\left(\frac{B_{ij}}{r_{ij}}\right)^{6}]+\sum \frac{(q_i q_i)}{(\epsilon r_{ij})}$$

where *V(r)* is the total potential energy. The sums in above equation (eq. 6.1) cover all possible bonds, angles, dihedral angles and atom-pairs. The first three summation terms correspond to the bonded interactions, whereas the last two represent the non-bonded interactions. Both bond (first) and angle (second) terms are harmonic approximations, whereas the third dihedral angle term is usually represented in the form of a periodic function. K_b, K_θ, and K_ϕ are the force constants that are defined for each type of bond, angle and dihedral, respectively. They are usually determined by the high-level quantum chemical calculations on small model systems or by spectroscopy [304,305]. The terms b_o and θ_o are the equilibrium values of bonds and angles, respectively, and are also determined through experiments or computational procedures [304]. In the equation 6.1, n is the multiplicity and δ is the phase, which is usually 0˚ or 180˚, however, other values have also been suggested. Moreover, ϵ is the dielectric constant and r_{ij} is the distance between the atoms i and j. A_{ij} and B_{ij} are the coefficients to model the 12-6 (repulsive-attractive) Lennard-Jones (LJ) potential, and q_i, q_j represents the partial charges on atoms obtained through quantum chemical calculations or other procedures [304,305]. From the potential energy, the forces on atoms are calculated by the relationship,

$$\vec{F}_i=-\nabla V(\vec{r}_i)$$

The potential energy equation, its derivatives, and above listed parameters such as force constants, equilibrium values, *etc.*, are collectively referred to as 'forcefield'. Various standard forcefields are popular in the field of biomolecular computations such as CHARMM, AMBER or OPLS [306,307,308,309]. Each forcefield has its own methodology to determine the parameters, and it is usually not advisable to mix the parameters from two different forcefields. In order to describe and simulate the protein systems, forcefield and parameter sets are compiled as efficient programs or software, which can be run on multiple processors. NAMD is one such program that is capable of achieving high scalability on multiple processors [310].

Since partial charges used in potential energy equation (6.1) are non-variable, most of the current development in this field is concentrated on the inclusion of polarization effects [311]. Recent studies have indeed shown improvements with the inclusion of fluctuating charges [311,312,313].

6.1.1 Integration schemes

In order to achieve both speed and accuracy in a simulation, efficient integration schemes are suitably coded in the MD simulation packages. A brief overview of the most popular integration schemes is given in this section.

In general, new positions and velocities of all atoms at time ($t + \delta t$) can be obtained from the Euler's method, in which both position and velocity terms are expanded in Taylor series,

$$r(t+\delta t)=r(t)+r'(t)\frac{\delta t}{1!}+r''(t)\frac{(\delta t)^2}{2!}+O((\delta t)^3)\Rightarrow r(t+\delta t)=r(t)+v(t)\frac{\delta t}{1!}+\frac{\vec{F}}{m}\frac{(\delta t)^2}{2!}+O((\delta t)^3)$$

and $$v(t+\delta t)=v(t)+\frac{\vec{F}}{m}\frac{(\delta t)}{1!}+O((\delta t)^2) \qquad \text{6.2 and 6.3}$$

However, such basic integration schemes are known to violate the characteristic time-reversibility, and they suffer with larger errors in the order of ($O(\delta t)^3$). Therefore, more stable algorithms have been proposed, such as Verlet integration [314]. In this scheme, a second Taylor expansion is used in addition to the one already described in Euler's method (eq. 6.2 and 6.3),

$$r(t-\delta t)=r(t)-v(t)\frac{\delta t}{1!}+\frac{\vec{F}}{m}\frac{(\delta t)^2}{2!}-O((\delta t)^3) \qquad \text{6.4}$$

Addition of the two expansions 6.2 and 6.4 yields,

$$r(t+\delta t)=2r(t)-r(t-\delta t)+\frac{\vec{F}}{m}\frac{(\delta t)^2}{2!}+O((\delta t)^4) \qquad \text{6.5}$$

This new formulation clearly provides an improvement due to the smaller magnitude of error in position ($O(\delta t)^4$), whereas velocities still retain error, due to the ($O(\delta t)^2$) or ($O(\delta t)$) term,

$$v(t)=\frac{r(t+\delta t)-r(t-\delta t)}{2\delta t}+O((\delta t)^2)$$

or $$v(t+\delta t)=\frac{r(t+\delta t)-r(t)}{\delta t}+O((\delta t)) \qquad \text{6.6}$$

In order to limit errors in the velocity calculation of the Verlet integrator, another scheme has been suggested, where,

$$v(t+\delta t)=v(t)+\frac{a(t)+a(t+\delta t)}{2}(\delta t) \qquad \text{6.7}$$

This formulation is called Velocity Verlet [304], and it helps in achieving higher accuracy with only minor increase in the computational cost [304]. One of the popular MD package NAMD uses the Velocity Verlet integrator by default, whereas CHARMM utilizes the Leap-Frog Algorithm, which has a high similarity with the former integration algorithm.

Irrespective of the integration schemes described above, the limited accuracy of computer's floating point operations and the stochastic nature of commonly used non-Newtonian equations such as, Langevin dynamics or Brownian dynamics, always causes a simulation to

diverge from its original trajectory [315]. This phenomenon is explained in terms of the Lyuapunov instability, for which any two simulations that start from the same point in the phase space will diverge over a period of time. This may not be a limitation in cases when a sufficiently large phase-space needs to be explored. However, the existence of some degree of synchronization in trajectories due to the 'seed' problems of computers may lead to less divergence among trajectories than desired [316].

In order to increase the length of simulation times, numerous schemes have been developed. In one of the most commonly used methods, the fastest motions of the lightest atoms (hydrogen atoms) are ignored, and a time-step of 1 or 2 fs is easily used [317,318]. Furthermore, when simulations are performed using Coarse-grained models [319], much larger time-steps can be used, even though such simulations do not provide details on the atomic-level. Other kinds of simulation approaches are also utilized in order to explore the large conformational space. These include simulated annealing, replica-exchange MD, torsion-angle dynamics, *etc.* In addition, these approaches can also allow in overcoming energy-barriers which are inaccessible to equilibrium MD simulations.

6.2 Statistical ensembles and thermodynamics

MD simulations are performed by mimicking the experimental conditions. The macroscopic details of the environment are conveniently associated with the microscopic details that have been obtained from the MD simulations through the theory of statistical mechanics and thermodynamic ensembles [320].

The NVE or microcanonical ensemble corresponds to a constant number of particles (N), volume (V) and energy (E), and is the most basic kind of ensemble in statistical thermodynamics. MD simulations by solving Newton's equations of motion conserve energy (E) and momentum by default, thus reproducing the NVE ensemble. Moreover, absence of any external force in the simulation renders the total energy value to remain constant throughout the simulation, though an initial increase in kinetic energy and decrease in potential energy is unavoidable as the system relaxes. Each snapshot in the simulation corresponds to a microstate with fixed *N*, *V* and *E*. The number of such microstates, corresponding to the energy E at equilibrium, is given by,

$$\Omega(N,V,E)=\sum_i \delta(H_i(\vec{p},\vec{q})-E) \quad \text{where} \quad \delta=\begin{cases}0, H_i\neq E\\ 1, H_i=E\end{cases} \qquad 6.8$$

where Ω is the microcanonical partition function, H_i is the Hamiltonian and vectors p and q define the momenta and positions of all particles in the system. Summation in eq. 6.8 covers all microstates i. The partition function Ω is an important quantity in statistical thermodynamics

from which all other thermodynamic functions such as entropy can be calculated,

$$S = k_B \ln(\Omega) \quad . \qquad 6.9$$

However, in practice it is nearly impossible to calculate *Ω* due to the enormous size of the phase space. The most important consequence of this ensemble formulation is that all microstates are equally likely, therefore probability (P_i) of finding the system in any one of the microstate is *1/Ω*. Moreover, from the probability and the assumption that the time average is equal to the ensemble average, the average of any property A can be calculated as,

$$\langle A \rangle = \sum_i A_i P_i \qquad 6.10$$

NVT (canonical) ensemble on the other hand corresponds to the constant temperature (T) instead of energy (E), which is achieved by connecting the system to an external thermal reservoir. The mathematical description of this ensemble forms the basis of statistical mechanics where the formulation of partition function (*Q*) turns out to be different from above,

$$Q = \sum_i e^{-\beta E_i} \qquad 6.11$$

Here *i* covers all the microstates, and E_i is the energy of the i^{th} microstate. Similar as above, the thermodynamic average of property *A* can be written as,

$$\langle A \rangle = \frac{\sum_i A_i e^{-\beta E_i}}{\sum_i e^{-\beta E_i}} \qquad 6.12$$

There are other ensembles such as the isothermal-isobaric (NPT) or grand canonical ensemble that are similar in formulation to these aforementioned ensembles [320]. Depending upon the type of simulation, the appropriate ensemble theory can be used to calculate the average properties, energetics and other thermodynamic functions.

6.3 Molecular Dynamics: Methods for constant temperature and pressure

6.3.1 Constant temperature

Alternative methods have been developed in order to keep the temperature constant during the MD simulations. The earliest methods relied upon the simpler velocity re-scaling techniques [304]. Since, average kinetic energy is directly related to the temperature through *‹KE›* = *3/2Nk$_B$T,* scaling factors such as $\sqrt{T_o/T}$ were proposed, in which T_o is the desired temperature and *T* is the temperature at time *t*. However, these methods and successive refinements [321] did not reproduce the desired canonical distributions. Therefore, more advanced temperature control methods, such as the Nosé-Hoover thermostat were developed. In the Nosé-Hoover thermostat method additional terms are integrated into the canonical Hamiltonian and equations of motion

are solved [322,323]. This method successfully led to the desired canonical distributions, and since then it has been used in many MD simulation studies.

Currently another method of choice is Langevin dynamics, which introduces frictional and random forces into the Newton's equation of motion, thereby mimicking solvent effects [324]. It has successfully been used to maintain constant temperatures in MD simulations, and has produced the required ensemble distributions [324]. Langevin equations are given by,

$$\vec{F}_i = -\nabla V(\vec{r}_i) - \gamma m_i \dot{r} + W_i(t) \qquad 6.13$$

where γ is the friction coefficient and W is the random force. Damping or friction coefficient γ describes the amount of friction in the system. Larger values would cause a high friction and that would result in a slower diffusion, such as in Brownian dynamics, whereas too small a value may result in larger temperature fluctuations.

6.3.2 Constant Pressure

Constant pressure and volume effects are simulated by placing the macromolecule inside a simulation cell of variable size and geometry. Depending upon the requirements either the pressure fluctuates (NVT ensemble) or the volume (NPT ensemble). Methods have been developed by which the volume of the simulation cell is scaled by a factor at each time step in order to keep the pressure constant [321]. Methods such as the Nosé-Hoover Langevin Piston pressure control [327], along with other methods [325,326], allows the generation of required NPT ensemble, and are used commonly in the MD simulations.

6.4 MD simulations on biomolecules

Current research in the field of MD simulations is aimed at modeling the biological systems as realistic as possible. In the case of simulations on membrane proteins, an atomic-level membrane model surrounding the protein can be constructed [328,329], which helps to avoid the surface effects at the protein surface. The hydrophilic region is also modeled by the inclusion of explicit water molecules and ions [328,329]. However, micro- or nano-second length simulations on such massive systems are prohibitive, especially when many redox or chemical states are to be simulated. Therefore, such simulations on membrane proteins can be performed in vacuum, specially in those cases in which the region of interest lies deep inside the protein surface, and where surface effects are small.

MD simulations on proteins are also performed by employing Periodic Boundary Conditions (PBC). In this case a replica of the central simulation cell is extended in all three spatial dimensions, such that atoms at the surface of the cell do not encounter a vacuum. This

reduces the surface effects and provides a more realistic description of the bulk properties [304]. The dynamics of the central cell is replicated in all the other cells, whereas the number of particles in the central cell can be kept constant by allowing the passage of particles through the boundaries [304].

Due to the large computational cost associated with the calculation of non-bonded interactions, MD simulations are often performed using the smoothing cut-off procedures. However, non-bonded electrostatic interactions are long-ranged and despite the shielding effects from the solvent and parts of the protein, truncation by smoothing procedures sometimes yield incorrect results, especially in cases of charged complexes such as DNA [330]. Therefore, Ewald summation techniques (or Particle-Mesh Ewald for periodic systems) are used to treat the long-range electrostatic effects, and are popularly used in the MD simulations on proteins or nucleic-acids [331,304].

Simulations of longer time scales are desired in order to obtain reasonable quantitative results. This could be achieved by increasing the time-step of integration [332], though that occurs at the cost of accuracy. Therefore, simulation techniques such as constrained simulations have been developed, in which certain degrees of freedom are constrained such as bond length and angles, which helps in increasing the time-step of integration. One such prominently used algorithm is SHAKE [317]. In this algorithm, bond lengths associated with the hydrogen atoms are often constrained, which leads to an increase in the time-step from 1 fs to 2 fs and without the loss of accuracy.

The MD simulations on proteins such as those which perform redox chemistry require partial-charges and parameters for the metallic co-factors, and those are often not available in the standard forcefields, due to the massive task of parameter development and refinement. Multiple methods exist to generate the partial charges for such systems, and enable their successful use in the simulations [333,334,335,336,337]. Optimization of non-bonded parameters is also a mammoth task similar to the partial charge development [305]. However, non-bonded LJ-parameters (ϵ_{ij} and R_{ij}) of hetero-diatomics are often obtained from the available LJ parameters of individual atoms [304,305,306] by using combination rules like,

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \quad \text{and} \quad R_{ij} = \frac{R_i + R_j}{2}$$

This helps in avoiding technical and methodological challenges that occur in the parameter development with little loss in accuracy. Nevertheless, the forcefield parameterization of metal centers present in an enzyme is not a straightforward automated task and requires optimization of several parameters at the same time [338].

Chapter 7

Aims and methods

7.1 Aims

The following were the specific objectives of the work presented in the thesis:

A) To identify the conserved amino acid residues and proton transfer pathways in the *cbb*$_3$-type oxidases.

B) To model the differential local effects and their importance in oxygen reduction and proton pumping mechanism of A- and C-type oxidases.

C) pK_a calculations to identify the proton-loading sites in A- and C-type oxidases.

D) To identify the functional groups involved in the redox coupled proton transfer in C-type oxidases.

7.2 Methods

7.2.1 Multiple sequence alignments and homology modeling

In studies (publications) I and II, multiple sequence alignments that comprised the catalytic subunits from A-, B- and C- type oxidases were generated (Fig. 3.1), and used to construct the homology model of C-type oxidase from *R. sphaeroides*. Due to the low pairwise sequence identity between the sequences (~20%), there was a high probability of generating incorrect alignments and incorrect homology models. Therefore, in order to achieve as high an accuracy as possible, a structural alignment was first generated in which the catalytic subunits of A- and B-type oxidases were aligned by superimposing the helical segments by using VMD/STRIDE [339,340]. Subsequently, enzyme sequences were aligned to the structural alignment, which was marked with the secondary structure definitions based on the structure of A-type oxidase from *B. taurus* [69]. The secondary structure definitions corresponding to the latter structure were used, since it was the highest resolution crystal structure available at that point of time. The sequences of A- and B-type oxidases were aligned to the sequence-structure alignment in small subsets using the *sequence-to-profile* methodology. The presence of conserved histidine residues over the span of sequence helped in achieving reasonable global alignment. Furthermore, the presence of a secondary structure profile did not allow insertions in the conserved helical segments, and also restrained the length of the TM regions.

C-type oxidases, though deviant from other members of HCOs, form a conserved subclass among themselves. Sequences of C-type oxidases from many different organisms were first aligned, and then added to the above alignment by *profile-to-profile* methodology. Multiple

trials were performed in order to assess the variability of the non-conserved regions. Available structural data on the A- and B-type oxidases were also used in order to manually adjust the alignments in the regions of low sequence conservancy. Additionally, sequence alignments of *b*-subunit of NORs were also found instrumental due to the local structural similarity between the catalytic subunits of NORs and C-type oxidases. All alignments were performed using the CLUSTAL program [180].

Homology modeling was performed with the program MODELER (version 7v7) [194]. From the alignment obtained above, the two template sequences (catalytic subunits of A-type oxidase from *B. taurus* and B-type oxidase from *T. thermophilus*) were extracted, along with the target sequence of C-type oxidase from *R. sphaeroides*. The use of two template structures allowed MODELER to select the better of the two along the sequence length, but at a cost of larger violations in the optimization schedule. However, it was later realized that most violations occurred in the loop regions where template and target sequences were the most divergent. The topologies and parameters of standard amino acid residues and metal centers were obtained from the standard forcefield or from the 'in-house' developed topologies/parameters [341]. No changes were observed in the final models upon change in the redox states of the metal centers. Based on the analysis of template structures, special restraints available in MODELER were used to model the interactions between protein and heme propionates. Out of total 50 models generated in each trial, five with the lowest value of MODELER objective function (section 3.3.1.1) were selected for further analysis. They were subjected to stereochemical checks using PROCHECK [209] and Molprobity [210]. Though MODELER's inherent optimization method involves an energy minimization step, separate minimizations were also performed using NAMD [310]. The energy minimization causes a minimal drift in the structure, and removes steric clashes and strains from the system, whereas molecular dynamics may degrade the models far from their native state [342,343,344].

7.2.2 Continuum electrostatic calculations

In studies IV and VI, continuum electrostatic calculations were performed on the crystal structures and also on the snapshot structures that were obtained from the MD simulations. In the latter case, in which conformational sampling was achieved through MD simulations, it was found critically important to select a representative set of structures. This was due to the larger fluctuations in pK_a values, even when the structural variations were non-local. Previously methods have been described by which individual conformers can be weighted using the forcefield based non-electrostatic energies. However, such approaches were found to be unsuccessful in the current case due to the biasing of one single conformer over all the others,

even when the difference between non-electrostatic energies of different conformers was small (unpublished information).

Continuum electrostatic calculations presented in publications IV and VI correspond to the enzymes from A- and C-type subfamilies, respectively. In publication IV, a representative set of structures were selected from the simulation trajectories that correspond to the two different conformations of A-propionate of heme a_3, *i.e.* 'open' and 'closed'. The 'open' conformation corresponds to the case when the hydrogen bond between the A-propionate and AI:Asp364 is broken, whereas in 'closed' it remains intact. Continuum electrostatic calculations were then performed for the different redox states. The pK_a values were averaged over the number of selected snapshots. In addition, auxiliary calculations were performed by mutating AI:Asp364 to an asparagine, thereby mimicking the propionate-protein interaction in the *cbb*$_3$-oxidases. The titration behavior of the whole protein was also analyzed in different redox states.

In study VI, pK_a calculations were performed on the crystal structure of *cbb*$_3$-oxidase and on the snapshots obtained from a short 250 ps MD simulation. The titration profile of the whole two subunit enzyme, and also the pK_a values of a larger set of active-site residues were obtained in different redox states.

Due to the dynamic nature of the hydrogen bonding patterns, explicit water molecules were not modeled in either of the above calculations, whereas solvent cavities and protein/membrane dielectric continua were modeled with ϵ=80 and 4, respectively. Water/OH^- ligands of the metal centers were treated explicitly when required. Software MEAD [276] was used to calculate the $\Delta\Delta G_{Born}$, $\Delta\Delta G_{Back}$ and site-site interaction energies (W_{ij}), whereas KARLSBERG [275] was used to perform the Monte-Carlo sampling in order to obtain the final pK_as. It has been observed before [284], and also confirmed in this work, that the variation in the pK_{int} values due to the structural variations is less, relative to the final pK_a values. In a number of cases it was also observed that $\Delta\Delta G_{Born}$ and $\Delta\Delta G_{Back}$ nearly canceled each other, and final pK_a were perturbed due to the W_{ij} terms only. Similar observations have also been reported previously [276].

In the calculations, charges of amino acid residues were obtained from the forcefield CHARMM [307], whereas charges of metal centers in different redox states were calculated using high-level DFT calculations, as described in the next section.

7.2.3 Density functional theory calculations

High structural similarity in the active-site region of all HCOs allowed DFT calculations to be performed on the cluster models constructed from the homology model of *cbb*$_3$-oxidase (publications III and V). In contrast, in the case of A-type oxidase, cluster models were built

from the available crystal structures (publications III and V). Additionally, DFT calculations were performed on the models built from the recently solved crystal structure of *cbb*$_3$-oxidase, as discussed in the publication VI.

DFT calculations were performed in different redox states, and in the absence or presence of non-proteinaceous metal-ligands such as OH^- or H_2O. Optimized geometries, electronic energies, spin-densities, redox potentials and partial charges were obtained from the calculations. In order to prevent any larger deviation from the initial structure, the termini atoms such as Cα or Cβ were kept fixed during the geometry optimizations. The GGA density functional BP86 [242,244] was used in all the geometry optimizations, with SVP basis set on all C, N, H and O atoms, whereas metals such as Fe and Cu were treated with the TZVP basis set [345,346]. The default convergence criteria of the program TURBMOLE [252] were used during the geometry optimizations, which were performed in the gas-phase. The Multipole Accelerated Resolution of Identity-*J* (MARI-*J*) algorithm available in TURBOMOLE was used to speed up the calculations [347]. Figure 7.1 shows the optimized geometries of metal centers present in the *cbb*$_3$-oxidase.

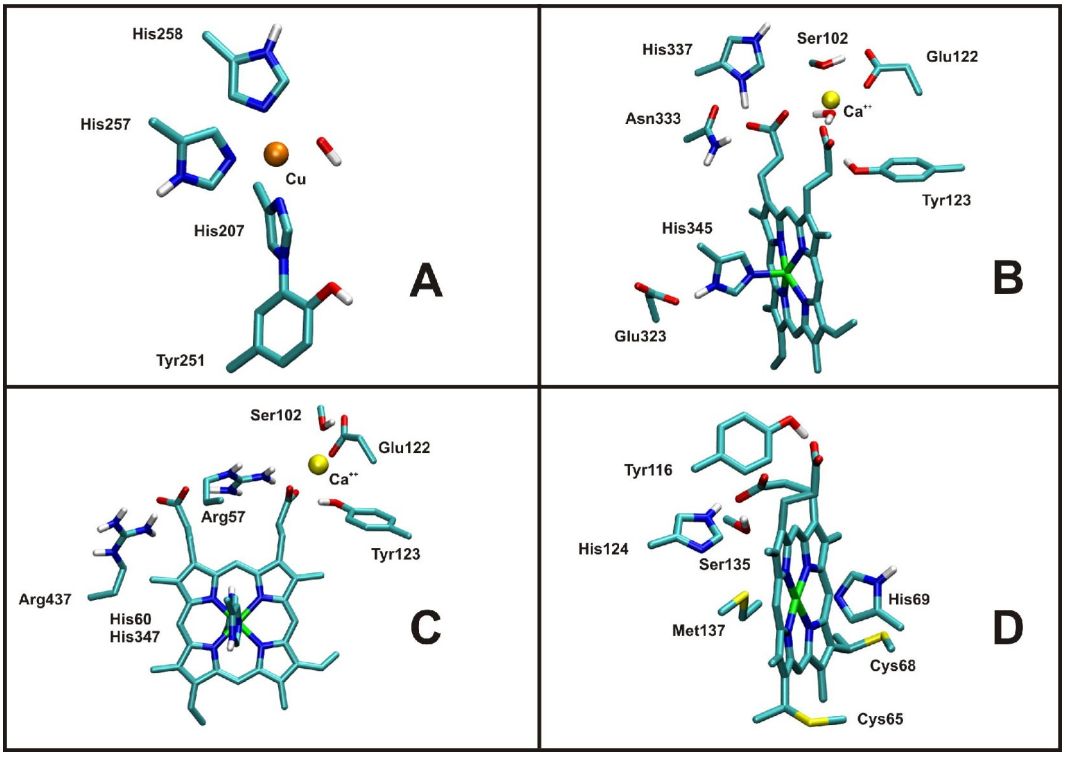

Figure 7.1 DFT optimized geometries of the A) Cu_B, B) heme b_3, C) heme *b*, and D) heme *c* metal centres present in the *cbb*$_3$-oxidase from *P. stutzeri*. The geometry optimizations were performed with the GGA density functional BP86, with SVP basis set on C, N, H, O and Ca atoms, and TZVP basis set on Fe and Cu atoms. Figure is published after permission from the publisher.

As higher accuracy was sought for the evaluation of properties such as relative energies,

spin densities, redox potentials and partial charges, hybrid density functional B3LYP [249] was used along with the TZVP basis set on all atoms (publications III, V, VI). The protein dielectric was modelled as a continuum with ϵ=4 using the COSMO methodology [253] available in TURBOMOLE. Unrestricted (*open-shell*) DFT calculations sometimes suffer from spin-contamination, in which the expectation value of the total spin-squared operator $<S^2>$ differs from the actual $S(S+1)$ [348,349]. In the calculations performed here, the extent of this contamination was found to be minimal. Moreover, the LUMO-HOMO gap was also found to be positive in all the calculations reported here. These analyses suggest that obtained results can be considered to be reliable.

The partial charges of all the atoms of the active-site models were obtained using the Merz-Kollman scheme [333] available in TURBMOLE. According to this methodology, the quantum chemical electrostatic potential is made to fit the potential due to the point charges centered on atomic nuclei. In study VI, these partial charges were used for the pK_a calculations and MD simulations.

7.2.4 Molecular dynamics simulations

In study II molecular dynamics simulations were performed on the water-loaded homology model, whereas in study VI simulations were performed on the recently solved crystal structure of *cbb*$_3$-oxidase. All simulations were performed by using the program NAMD [310]. The water molecules were predicted in protein cavities based on the interaction energy approaches [350]. Different energy cut-offs of water prediction were applied and tested in study VI. All simulations were performed in vacuum, and based on the root mean square deviation (rmsd) *vs* time plot, it was observed that the system remained stable over nanosecond timescales (Figure 7.2).

In the simulations presented in study VI, the temperature was kept constant at 310 K by using the Langevin dynamics method available in NAMD with damping coefficient = 0.1 ps^{-1}. Time step of 1 fs was used in all the simulations, and data were saved every 100 steps. The non-bonded cut-off of 20 Å and smoothing criteria of 18 Å were applied. The standard forcefield parameters along with the 'in-house' developed parameters were used for the protein-metal systems, whereas charges were obtained from the DFT calculations as described earlier (section 7.2.3). Standard protonation states of all amino acid residues were considered, except the ones which showed deviated pK_a values. Multiple simulations of 5 ns each, after 1 ns equilibration, were performed in different redox states. This most likely helped in exploring larger conformational space, and also allowed the water molecules to attain a relaxed state inside the protein interior.

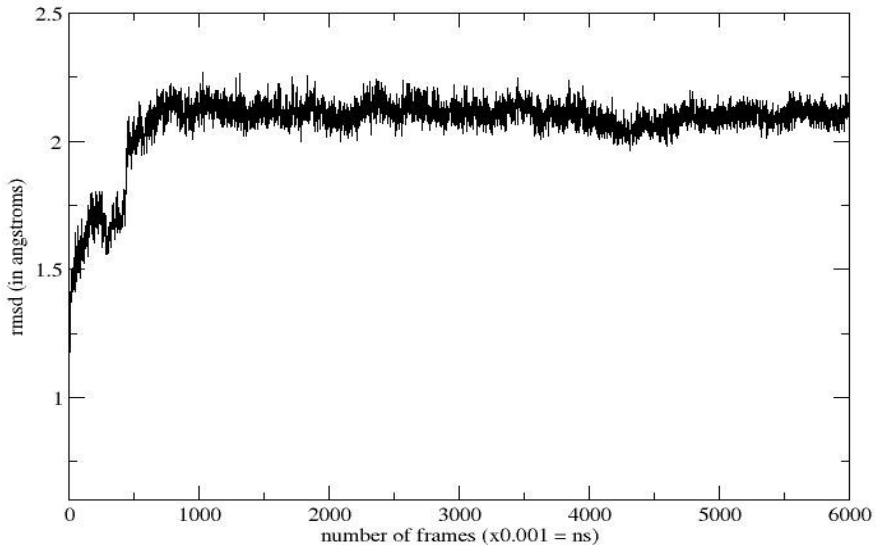

Figure 7.2 rmsd *vs* time plot, based on a 1 ns equilibration and 5 ns production run in OROO redox state (publication VI). C*α* atoms were used to align and superimpose the structures on the initial frame.

In study II, short MD simulations of 450 ps were performed in order to identify the stable arrangement of side-chain of cross-linked tyrosine in the water-loaded homology models. Due to the likelihood of degradation in quality of homology models during the MD simulation, harmonic constraints were put on all the C*α* atoms.

Chapter 8

Results and discussion

8.1 Homology model of *cbb*$_3$-oxidase and its comparison with the crystal structure

The superposition of the helical segments of the homology model of *cbb*$_3$-oxidase on the crystal structure yields 2.35 Å *rmsd* for the C*α* atoms (Figure 8.1A and 8.1B). It is observed that the region from helix VI to VIII, and some of the interactions between protein and heme propionates are correctly predicted by the homology models. The crucial active-site tyrosine (section 2.5.3), which is fully conserved in the A- and B-type oxidases in helix VI, is located in the helix VII of *cbb*$_3$-oxidases. The homology models provided the first insight as to where the tyrosine CI:Tyr251 in helix VII can possibly make a covalent bond with the histidine ligand (CI:His207) of Cu_B (publications I and II). Figure 8.1C shows the side-chain of tyrosine in proximity of the histidine ligand and also nearly matching the crystal structure conformation. This prediction is confirmed by the site-directed mutagenesis, mass-spectrometry and recent crystal structure data [43,156,157]. Moreover, the side-chain dihedral $\chi 2$, inter-planar angle between histidine and tyrosine, and the improper torsion ($C\varepsilon 1_{His207}$-$C\varepsilon 1_{Tyr251}$-$N\varepsilon 2_{His207}$-$C\delta 2_{His207}$) in the energy minimized crystal structure are 137°, -14° and 176°, respectively. These values closely match the predicted values in the refined homology model (publication II).

In the region from helix VI to VIII the local similarity between the template sequence (1EHK.pdb) and target sequence is relatively high. Hence, a putative proton channel is predicted in the C-type oxidases in the region where a functional proton channel exists in both the A- and B-type oxidases (publication I). The presence of this channel in C-type oxidases is confirmed with the site-directed mutagenesis studies and recent X-ray structure [43,156]. Figure 8.1C shows the correctly predicted backbone and sidechains of amino acid residues involved in the proton channel.

However, due to errors in the alignment, particularly in the regions from helix III to V, and loop between helices XI and XII, some backbone and side-chain atoms are modelled incorrectly (Figure 8.1A and 8.1D). This is either due to the low sequence conservancy between the template and target sequences or due to the incorrect modelling procedure applied.

Multiple sequence alignment in the regions of low sequence similarity tend to depend upon the order in which sequences are added and aligned. Structural alignments, multiple trials and manual intervention play important roles in assigning the variable parts more accurately. In the multiple sequence alignment obtained in study I after several trials, it is observed that besides

the fully conserved metal-ligating histidines there are only few other highly conserved residues in the whole HCO superfamily. Two of these are; CI:Pro282 in helix VIII, and CI:Gly324 in helix IX. When compared with the crystal structure and structural alignment [43], both these residues are correctly predicted in the alignments and homology models. Proline is located in the region of the K-channel of A-type oxidases, where it most likely provides the helix with a functional flexibility [74,351]. On the other hand, glycine is located near the high-spin heme, and due to the absence of side-chain it packs closely to the heme and its histidine ligand. Glycines are known to be located in the regions where tight packing is required [352], and indeed a bulky residue at such a position affects the activity of the oxidase [353]. Additionally, many of the glycines (CI:Gly391 and CI:Gly403), which were used as a yardstick during the alignment generation are also modelled correctly.

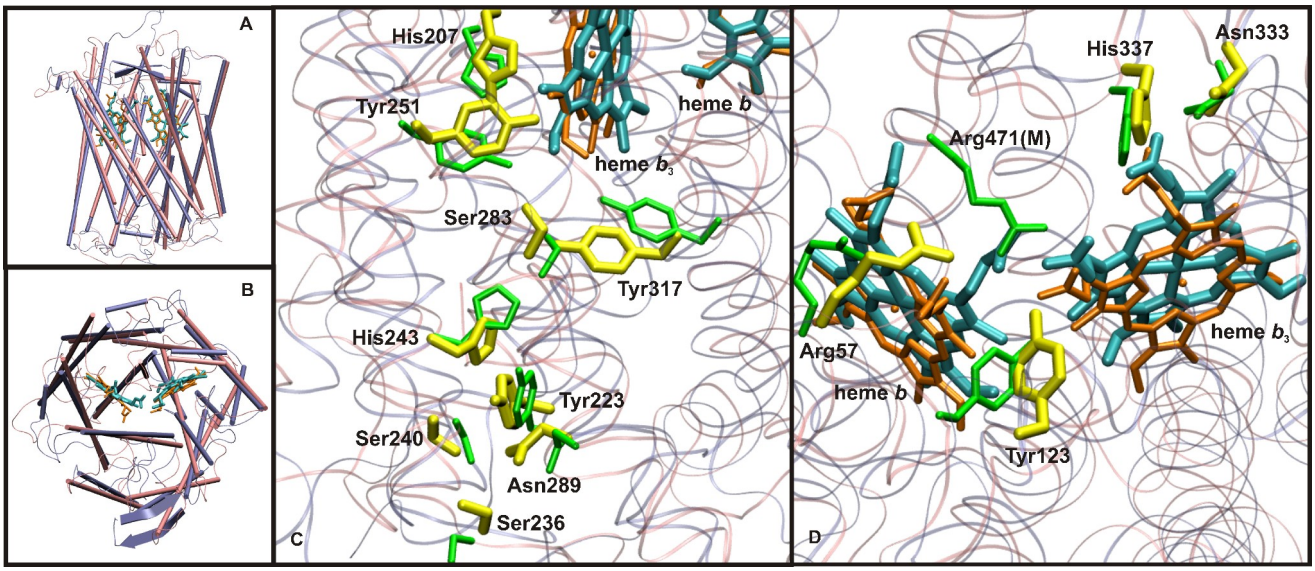

Figure 8.1 Local and global comparisons between the homology model of the catalytic subunit of *cbb*$_3$-oxidase from *R. sphaeroides*, and the same subunit from the crystal structure of *P. stutzeri* enzyme. Side (A) and top (B) view of the superposed structures; pink, model; ice blue, crystal structure. Closely matching side-chain conformations between the model (green) and the crystal structure (yellow), in the region of the putative proton-channel (C), and above heme propionates (D). All amino acid numbering corresponds to the subunit N of the *cbb*$_3$-oxidase from *P. stutzeri*, except the incorrectly modelled Arg471(M). Hemes are shown in licorice representations (cyan for x-ray structure, and orange for the model).

The section from helix III to V is variant in all the three subfamilies of HCOs (publication I). In the A- and B-type oxidases, amino acid residues from loop III-IV participate in charged interactions with the heme propionates, and are important in proton pumping. Based on local similarities in the template and target sequences, such interactions were modelled

explicitly in studies I and II (see also section 7.2.1), and interestingly, part of the predictions turned-out to be correct, namely that: *a)* a highly conserved arginine (CI:Arg57) from helix II interacts with the low-spin heme propionates, and *b)* a conserved tyrosine from the loop III-IV (CI:Tyr123) hydrogen bonds to the D-propionate of the high-spin heme [43] (Figure 8.1D). In contrast, the explicitly modelled interactions between the D-propionate of high-spin heme and the residues from loop XI-XII are found to be incorrect. Instead of a modelled arginine (R471(M) in Fig. 8.1D), a Ca^{+2} ion neutralizes the negative charge of the propionate [43]. Nevertheless, part success in the correct identification of heme propionate partners highlights the role of manual intervention in homology modelling.

In the first models the role of a highly conserved glutamate in helix IX (CI:Glu323) was unclear. However, a putative proton channel was suggested based on its location and mutational data in study I. Recent biochemical studies in conjunction with redox titrations identified a hydrogen bond between the glutamate and the histidine ligand of high-spin heme [159], which is also confirmed by the recent crystal structure [43]. This experimental input led to the improvement in models, and another highly conserved interaction between the glutamate and the tryptophan (CI:Trp341) is predicted in study III. This prediction is also supported by the confirmation from the recent structure [43]. As discussed in a later section, pK_a calculations and MD simulations indeed suggest a role of this glutamate in proton transfer (publication VI).

Due to the difficulty in aligning one of the two conserved tryptophans in helix VI of *cbb*$_3$-oxidases to the only conserved tryptophan in the A-type oxidases, two different models 'continuous' and 'gapped' were constructed in studies I and II. Comparison with the recent crystal structure and the structural alignment presented in the reference [43] reveals that one of the two predicted models is correct, and manual insertion of a gap in the region led to incorrect results. The tryptophan (AI:Trp236) forms a stacking interaction with the histidine ligand of Cu_B, and its mutation to other residues sharply decreases the catalytic activity of the oxidase [354]. A similar result can also be expected in the *cbb*$_3$-oxidases for analogous tryptophan CI:Trp203.

8.2 Quantum chemical modelling of the active-site

The overall arrangement of the active-site is similar in both A- and C-type oxidases. Nonetheless, there are local structural differences, which might be important in the oxygen reduction and proton pumping mechanisms. The DFT calculations were performed on the cluster models of both the oxidases in order to identify the differences in energies and structure of the intermediate states. In the case of *cbb*$_3$-oxidase, a model system that comprised of BNC region was constructed from the homology model, whereas for the *aa*$_3$-oxidase, available crystal structure was used. The models are relatively large and comprise *ca.* 200 atoms.

DFT calculations showed large differences in the spin-density distribution on heme and its histidine ligand in the two model systems (A and C) in the oxidized state (Figure 8.2).

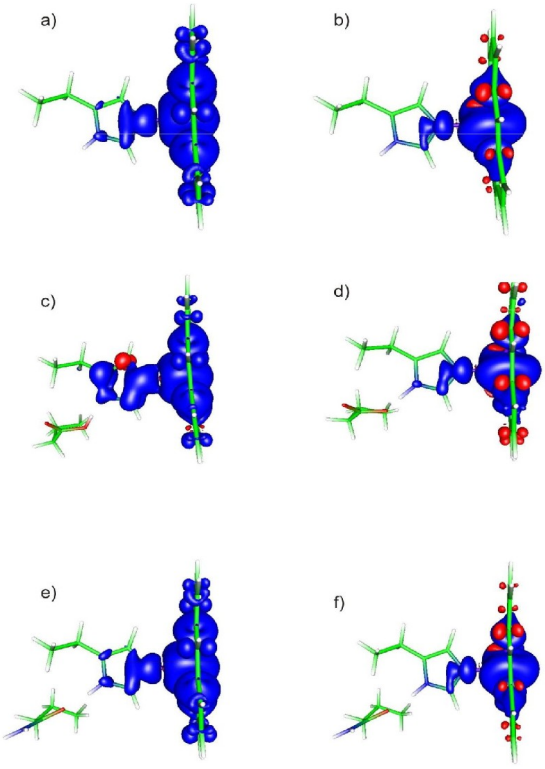

Figure 8.2 Spin densities in small model systems. Left (oxidized), right (reduced). *a), b)* corresponds to the active site of A-type oxidases, whereas *c), d)* to the C-type oxidases. *e), f)* systems are variants of *c), d)* in which glutamate is *in-silico* mutated to a glutamine. Isocontour plots at the value of 0.001 $e/Å^3$ are shown with colours blue and red representing the excess α and *β* densities, respectively.

The differences in spin density on histidine in the oxidized case are clearly visible in *a), c)* systems, whereas *e)* shows spin distribution more similar to *a)*. Figure is published after permission from the publisher, and is from publication III.

These differences in spin-density can be associated with the rhombic EPR spectrum that is obtained in the C-type oxidases as compared to the axial spectrum of A-type oxidases [159]. Preliminary calculations also suggest differences in the EPR *g*-tensors and zero-field splitting parameters (*E* and *D*) for the two model systems (unpublished data). Furthermore, upon reduction of heme, spin distributions in both the model systems are found to be similar as shown in Figure 8.2b and 8.2d. The most likely reason for the observed differences is the presence of the glutamate (CI:Glu323) in the proximal cavity of heme b_3, which hydrogen bonds with the histidine ligand of the latter (Fig. 8.2c and 8.2d). This unique interaction in the *cbb*$_3$-oxidase modulates the geometrical and also the electronic structure of the BNC. The proximal site architecture also resembles that is found in the peroxidases, where an analogous aspartate modulates the electronic and geometrical characteristics of the system [160]. Additionally, in study III, a redox dependent proton movement is observed in both small and large cluster model systems of *cbb*$_3$-oxidase. The proton shared between the histidine and glutamate is found closer to the glutamate in the oxidized state, whereas in the reduced state it is more closer to the histidine (Figure 8.2c and 8.2d). The preference of proton on one of the residues is also

supported by the energetic and frontier orbital analysis (publication III). However, in peroxidase enzymes in which an analogous system (Fe-His-Asp) exists, the position of the shared proton has been found to be closer to the histidine ligand based on the hybrid QM/MM calculations [355].

The possible effect of hydrogen bonding between histidine and glutamate on the energetics of oxygen reduction reaction was also studied. Models of A- and C-type oxidases were constructed in a similar manner as described above, and the reaction sequence $\mathbf{A \rightarrow I_P \rightarrow P_M}$ was studied. It is observed based on the analysis of electronic energies that the peroxy intermediate ($\mathbf{I_P}$) is more stable in the cbb_3-oxidases compared to the aa_3-oxidases (publication V). This finding would lead to a higher occupancy of intermediate $\mathbf{I_P}$ ([$\mathbf{I_P}$]) and a higher rate of $\mathbf{P_M}$ formation (d[$\mathbf{P_M}$]/dt). Consequently, this would also lower the Michaelis-Menten constant $\mathbf{K_M}$ in the case of cbb_3-oxidases [356]. Interestingly, it is well-known that Michaelis-Menten constant $\mathbf{K_M}$ in cbb_3-oxidases is low in the nM range, and they possess a relatively higher affinity for oxygen as compared to the other HCO superfamily members. The two most likely reasons for the stabilization of $\mathbf{I_P}$ intermediate in cbb_3-oxidase are, *a)* the presence of negatively charged glutamate in the proximal site of heme b_3, and *b)* different sidechain orientation of cross-linked tyrosine (CI:Tyr251). Both these characteristic features of C-type oxidases also classify them as being different from the A- or B-type oxidases [43]. Moreover, as observed in study V, *ca.* 10 kcal/mol electrostatic stabilization of $\mathbf{I_P}$ intermediate in cbb_3-oxidases is most likely caused by these distinct structural features.

The electronic energies of other intermediate states such as **R**, **A** and $\mathbf{P_M}$ are found to be similar in both the oxidases. The approximate 20 kcal/mol exergonicity from $\mathbf{R \rightarrow P_M}$ if combined with the 10 kcal/mol entropic contribution [357] would yield 10 kcal/mol overall exergonicity, which is in harmony with the previous similar DFT studies made on the A-type oxidases [115,128,129].

In general it is difficult to judge the formal oxidation states of metal centers based on the Merz-Kollman (MK) charges, as charges tend to migrate to metal-ligands or other locations. The same holds true for other functional groups, however, in the current calculations MK charge analysis of the cross-linked tyrosine shows that it has a neutral charge in the $\mathbf{P_M}$ state (0.17 and 0.02 in A and C models, respectively), which implies that it most probably forms a neutral radical and provides the fourth electron during the oxygen splitting reaction (publication V). Figure 8.3 also confirms this where the unpaired electron distribution on the cross-linked tyrosine in the $\mathbf{P_M}$ intermediate is shown.

Figure 8.3 (inset) also shows that LUMO lies on the cross-linked tyrosine, such that upon reduction from $\mathbf{P_M \rightarrow P_R}$, cross-linked tyrosine would act as an electron acceptor, in conjunction

with the similar previous observations on the aa_3-type oxidases [21].

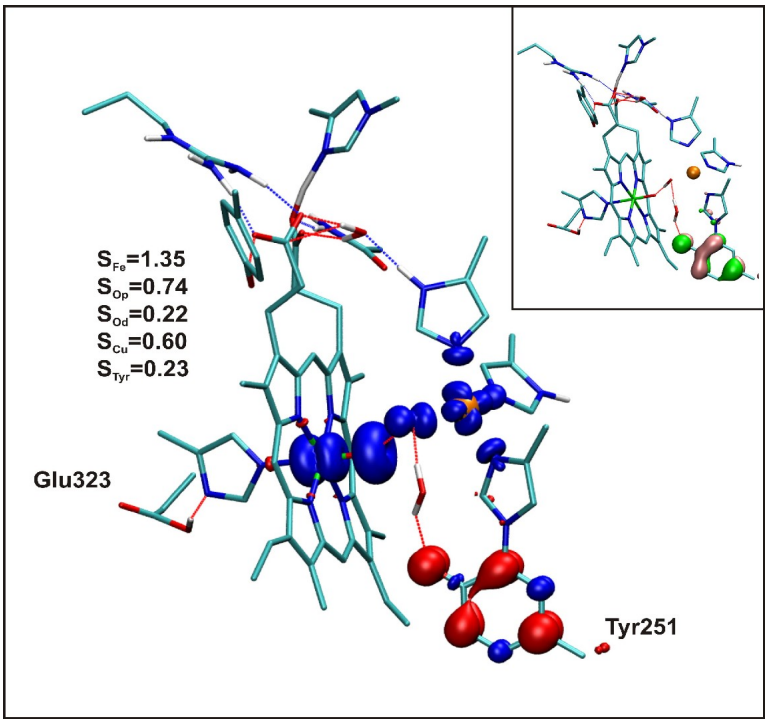

Figure 8.3 Spin density distribution in the proposed $\mathbf{P_M}$ intermediate in the catalytic cycle of C-type oxidases. Iso-contour plots are drawn at the value 0.005 $e/Å^3$. Blue and red are the excess α and β spin densities, respectively. The α density values are also listed for some specific moieties. Inset shows the presence of LUMO on the cross-linked tyrosine (Isocontour value: 0.05 $e/Å^3$). Atoms are shown in the following colours: C (cyan), H (white), N (blue), O (red), Fe (green) and Cu (orange). Only hydrogens attached to N and O are shown.

A highly conserved tryptophan forms a neutral or cationic radical in the catalytic reaction of peroxidases [358]. It makes stacking interactions to the heme and its axial histidine ligand in addition to a strong hydrogen bond with an aspartate. All these interactions in combination, modulates the electronic and geometrical structure of the whole system [160]. Although a similar structural arrangement does not exist in the cbb_3-oxidases, preliminary calculations on the extended cluster model does show the spin-density distribution on a distant tryptophan (CI:Trp400) (Figure 8.4). This tryptophan along with another (CI:Trp341) hydrogen-bonds with the proximal glutamate, but do not form any other interactions with the heme or histidine (Figure 8.4). The conjecture that tryptophan may play a role in electron transfer would require further experimental and computational studies. However, the high sensitivity of spin density on the dielectric constant support that the cross-linked tyrosine is the potential electron donor in cbb_3-oxidases too.

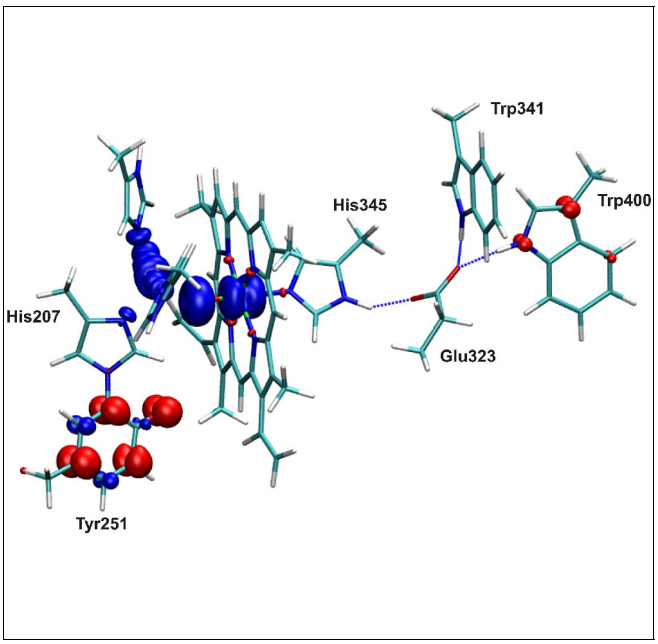

Figure 8.4 Spin density distribution in the $\mathbf{P_M}$ intermediate. The extended active site model is based upon the recently solved crystal structure of *cbb*$_3$-oxidase from *P. stutzeri* [43]. Isocontour plots are drawn at the value 0.005 *e*/Å^3. Blue and red colours represent the excess α and *β* spin densities, respectively. C (cyan), H (white), N (blue), O (red), Fe (green) and Cu (orange) atoms are shown along with hydrogen bonds as blue dotted lines.

8.3 Proton-loading site: pK_a calculations

The proposed proton pumping mechanism in the A-type C*c*Os require a transient proton-loading site (PLS) from which the pumped proton is eventually released into the P-side of the membrane. The additional requirement for this site is that its pK_a should modulate with the changes in redox state of the metal-centers and its interaction with the proton consumed at the BNC [21,98,136]. Previous thermodynamic assessment of the catalytic reaction of A-type C*c*O, in addition to independent computational studies, suggested that the A-propionate of high-spin heme is the PLS [21,137]. However, explicit variations in its pK_a due to the changes in redox states have not been observed.

MD simulations combined with pK_a calculations identified that the A-propionate shows high proton affinity (pK_a *ca.* 10) upon dissociation from the hydrogen-bonding partner AI:Asp364 ('open' conformation). In contrast, when the two are strongly hydrogen bonded ('closed' conformation), pK_a of propionate is much lower. Additionally, during the relatively short MD simulations of 5 ns, the open/closed conformations of A-propionate are found to be redox state dependent (publication IV). In the $\mathbf{P_M}$ state with heme *a* reduced, multiple closed ↔ open transitions are observed, whereas no such transitions are observed in other states such as $\mathbf{P_R}$. This study suggests that A-propionate is the transient PLS in the proton pumping mechanism of C*c*O.

The proton affinities of a larger set of residues have also been calculated from different redox states by averaging over the snapshot structures obtained from the equilibrium MD simulations. Although longer MD simulations or rigorous free energy calculations would be required to sample the large phase space and to obtain accurate estimates of energetics of redox dependent transitions, the increase in proton affinity of A-propionate upon 'opening' support the earlier proposals that reduction of heme *a* is linked with the proton-uptake [67,68,69,86,135]. Thus, A-propionate of high-spin heme could be the proposed PLS, in that it accepts a proton from the D-channel end point AI:Glu242. Furthermore, since region around the A-propionate is a complex network of hydrogen-bonds, it can be assumed that the charge of proton is distributed throughout the region rather than residing fully on the propionate itself.

Instead of an aspartic acid an asparagine (CI:Asn333) makes a hydrogen bond to the A-propionate of high-spin heme in the *cbb*$_3$-type oxidases [43]. The A-propionate ↔ Asn system could thus be assumed to fluctuate in the same way as the A-propionate ↔ Asp system does in the A-type oxidases. If it does occur, then a similar increase in the proton affinity of A-propionate could be expected (publication IV). pK_a calculations performed by mutating the aspartic acid into asparagine *in-silico,* in both 'closed' and 'open' conformations, led to an equivalent increase in the pK_a of propionate, which suggests that the same region could form the PLS in the C-type oxidases as well. A more intriguing picture emerged when MD simulations and pK_a calculations were performed on the *cbb*$_3$-structure as described in the next section.

8.4 Water networks in *cbb*$_3$-oxidase: MD simulations and pK_a calculations

It is well-known that both A- and C-type oxidases catalyse the reduction of molecular oxygen to water and convert the free-energy into proton pumping. Even so, proton pumping stoichiometry in the latter enzymes is still a subject of debate. It has recently been suggested that the C-type oxidases pump protons at nearly half the stoichiometry reported for the A-type enzymes [144]. However, this finding contradicts the view that C-type oxidases are highly efficient enzymes and are capable of 1 H^+/e^- pumping stoichiometry even under low-oxygen tensions. Notwithstanding the stoichiometry, the proton transfer pathways must be present in the catalytic subunit of C-type oxidases in order to supply protons either to the BNC or to the PLS. Earlier homology models and site-directed mutagenesis studies identified a proton channel in the region analogous to the K-channel of the A-type oxidases (publication I) [156]. Recent crystal structure also confirmed the presence of this proton channel [43]. However, this path does not complete the connection to the P-side of the membrane, which would be essential for proton pumping. In publication VI, a proton channel is identified in the catalytic subunit of *cbb*$_3$-oxidase from *P. stutzeri*, which may be utilized for redox-coupled proton pumping.

The work is based on three different computational approaches; DFT calculations, MD simulations and continuum electrostatic calculations. Since MD simulations on oxidases require partial charges of metal centers, ESP based charges were obtained after performing DFT calculations on the cluster models (publication VI). This type of approach has successfully been used before, on variety of enzyme systems [359,360,361]. Test calculations were also performed in order to gauge the variation in charges with different methods available in TURBOMOLE. For instance, a change in the dielectric constant from vacuum to 4 did not alter the charges to a greater extent. In addition, increasing the density of points in the ESP calculation had a smaller effect on the charges. Moreover, qualitative comparison of the obtained charges show similarities with the standard CHARMM charges especially for the *b*-type heme and glutamate residue, which corroborates the process used here. It has previously been argued that ESP based charges of buried atoms are not accurate [336], nevertheless, in the current study most of the results are compared between different redox states, hence, such errors would most likely have a small effect.

The continuum electrostatic calculations were performed on the crystal structure in addition to the snapshots structures obtained from MD simulations. Figure 8.5 shows the surface representation of electrostatic potential in a fully oxidized state of the enzyme. Although proton-affinities of a larger set of residues were obtained, two residues distinctively showed deviant profiles. First, histidine (CI:His337) displayed high pK_a in all the studied redox states. CI:His337 is located near the P-side of the membrane and hydrogen-bonds to the A-propionate of high-spin heme (Figure 8.1D and 8.6B) (publication I) [43]. The high proton-affinity of histidine in the C-type oxidases contrasts with the A-type oxidases in which the analogous histidine cannot be protonated due to its ligation to the redox inactive metal center (Mg^{+2}) [69]. On the other hand, the corresponding histidine in B-type oxidases is likely to be protonated due to its interaction with a negatively charged residue in the region [29], whereas no such interaction exists in the C-type oxidases [43]. The protonated histidine would neutralize the negative charge of the propionate in the crystal structure conformation. However, considering the fluctuations observed in the hydrogen-bonding between the histidine and the propionate during the MD simulations, it can be inferred that this region is likely to be a PLS region (publication IV), and dissociation of the hydrogen bond between the two may increase the proton affinity of the latter.

Second, the proximal glutamate (CI:Glu323, Figure 8.6A and 8.6B) shows protonation upon the reduction of heme b_3, whereas it may or may not be protonated when heme is oxidized. The hydrogen-bonding of glutamate to the histidine ligand of heme is a unique interaction known only in the C-type oxidases, and the observed protonation profile implicates a redox-coupled protonation mechanism. Moreover, since redox potential of heme b_3 is uniquely low (-

60 mV), it is likely that protonation of proximal glutamate tunes the redox potential of the former to allow its reduction, which would be a proton-coupled electron transfer kind of mechanism.

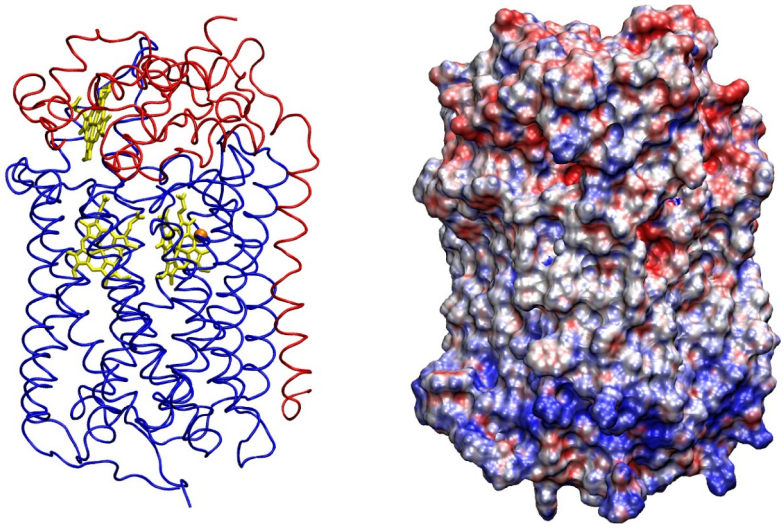

Figure 8.5 The two subunit C-type oxidase from *P. stutzeri* is shown in the ribbon representation (left). Subunit I in blue and subunit II in the red colour. Hemes are shown in yellow and copper in orange. The surface electrostatic potential representation in the fully oxidized state is also shown (Right). Blue corresponds to +5 *k*T/e and red, -5 *k*T/e.

The MD simulations were performed on the water-loaded two subunit enzyme in different redox states. Water molecules were predicted inside the protein interior by using the interaction energy approach as available in the DOWSER program [350]. Since this approach is based on crudely parametrised van der Waals and electrostatic energy interactions, different energy cutoffs were applied in order to analyse the water density inside the protein. More sophisticated thermodynamic based approaches of water prediction, which have previously been used in the A-type oxidases [83], have not been used here.

Multiple simulations of 5 ns each in different redox states showed the formation of water-chains inside the protein interior. The water-chain at the location analogous to the K-channel of A-type oxidases formed irrespective of the redox state or different energy cutoffs used for water prediction (Figure 8.6C). This stable water chain thus enables continuous supply of protons to the BNC for oxygen reduction chemistry to occur. Another water-chain is concurrently observed, which connects the N-side of the membrane to the proximal glutamate (CI:Glu323), and glutamate to the histidine-propionate pair (Figure 8.6A and 8.6B). This pathway could be used for redox-coupled proton pumping in the enzyme. The stable water-chains lead to the identification of certain amino acid residues which may be important for proton transfer reactions. This result requires site-directed mutagenesis studies in order to confirm the

predictions made here.

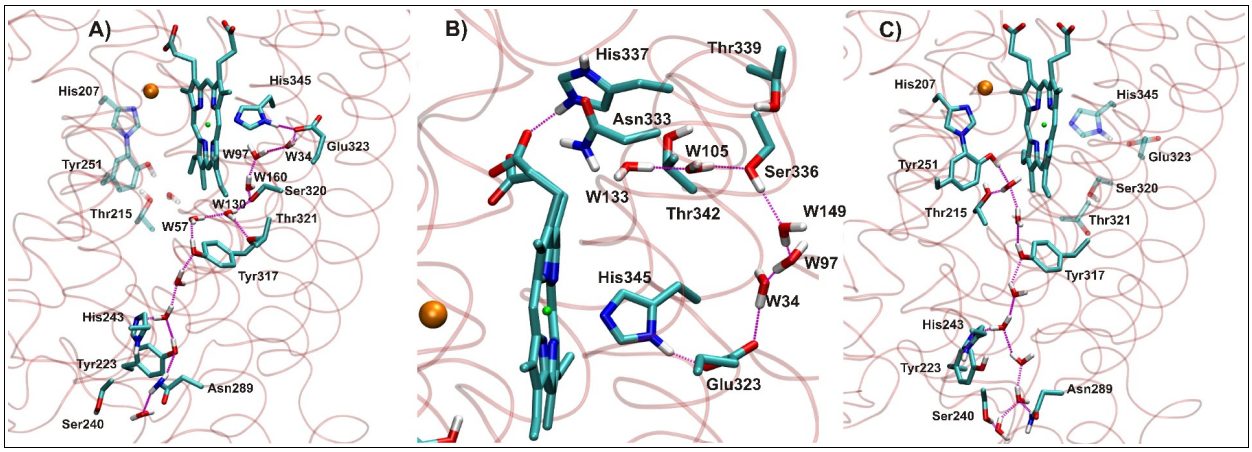

Figure 8.6 Hydrogen-bonded water chains in the C-type oxidase. A) Water-chain connecting the N-side with the proximal glutamate (CI:Glu323) and, B) proximal glutamate with the A-propionate region. C) Water chains connecting the N-side with the active site of oxygen reduction (BNC). Water molecules are marked with a prefix W. Amino acid numbering corresponds to the crystal structure of C-type oxidase from *P. stutzeri* [43]. C (cyan), H (white), N (blue), O (red), Fe (green) and Cu (orange) atoms are also shown along with hydrogen bonds as purple dotted lines. Only hydrogen atoms connected to N and O are shown. Figure 8.6 is published after permission from the publisher.

Chapter 9

Conclusions and future directions

The thesis work highlights the importance of computational methods in studying the structure and function of an enzyme. Computational methods not only help in providing a molecular basis to the experimental results, but also allows the generation of new hypotheses, which can be tested with both experimental and computational methods. However, at the same time it is important to be aware of the limitations of computational approaches.

Homology modelling successfully predicted the correct location of cross-linked tyrosine and a proton channel in the C-type oxidases, which is impressive especially considering the low levels of sequence identities (*ca.* 20%). However, it is realized in the work that manual adjustments in the alignment or other refinements in the models should be done with extra care. DFT calculations on cluster models of active-sites provided molecular level interpretation to the EPR experiment results. In addition to that, they also described the structural reasons for the high oxygen affinity shown by the C-type oxidases. Such proposals can be confirmed with further experimentation and modelling. Additionally, continuum electrostatic calculations and MD simulations on the C-type oxidase led to the identification of amino acid residues which might be important in redox-coupled proton transfer. These predictions can be analysed with biochemical methods such as site-directed mutagenesis approaches.

The recently solved crystal structure of C-type oxidase has opened up new avenues for the computational biochemistry field, such that many different computational methods can now be applied. Advanced methods including free-energy calculations or QM/MM calculations will be used to understand the electron and proton transfer reactions in the enzyme. Direct comparisons can be made to the well-studied A-type oxidases, which would help in identifying both the differences and the similarities between the two. It is of general importance to identify the minimal functional elements that would constitute a proton pumping machine.

Bibliography

1 Voet, D.; Voet J. G. Biochemistry 2004, 3rd Edition, Wiley.
2 Efremov, R. G.; Baradaran, R.; Sazanov, L. A. *Nature* **2010,** *465,* 441.
3 Sun, F.; Huo, X.; Zhai, Y.; Wang, A.; Xu, J.; Su, D.; Bartlam, M.; Rao, Z. *Cell* **2005,** *121,* 1043.
4 Lange, C.; Hunte, C. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99,* 2800.
5 Tsukihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itoh, K.; Nakashima, R.; Yaono, R.; Yoshikawa, S. *Science* **1996,** *272,* 1136.
6 Watt, I. N.; Montgomery, M. G.; Runswick, M. J.; Leslie, A. G. W.; Walker, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107,* 16823.
7 Hinkle, P. C. *Biochim. Biophys. Acta* **2005,** *1706,* 1.
8 Mitchell, P. *Nature* **1961,** *191,* 144.
9 Brandt, U. *Annu. Rev. Biochem.* **2006,** *75,* 69.
10 Wikström, M.; Morgan, J. E.; Verkhovsky, M. I. *J. Bioenerg. Biomembr.* **1998,** *30,* 139.
11 Wikström, M. *FEBS Lett.* **1984,** *169,* 300.
12 Hunte, C.; Zickermann, V.; Brandt, U. *Science* **2010,** *329,* 448.
13 Mitchell, P. *FEBS Lett.* **1975,** *56,* 1.
14 Mitchell, P. *J. Theor. Biol.* **1976,** *62,* 327.
15 Yoshida, M.; Muneyuki, E.; Hisabori, T. *Nat. Rev. Mol. Cell Biol.* **2001,** *2,* 669.
16 Schweitzer, C.; Schmidt, R. *Chem. Rev.* **2003,** *103,* 1685.
17 Malmström, B. G. *Annu. Rev. Biochem.* **1982,** *51,* 21.
18 Prabhakar, R.; Siegbahn, P. E. M.; Minaev, B. F. *Biochim. Biophys. Acta* **2003,** *1647,* 173.
19 Kharkats, Y. I.; Volkov, A. G. *Anal. Sci.* **1998,** *14,* 27.
20 Kirkinezos, I. G.; Moraes, C. T. *Semin. Cell Dev. Biol.* **2001,** *12,* 449.
21 Kaila, V. R. I.; Verkhovsky, M. I.; Wikström, M. *Chem. Rev.* **2010,** *110,* 7062.
22 Wood, P. M. *Biochem. J.* **1988,** *253,* 287.
23 Ferguson-Miller, S.; Babcock, G. T. *Chem. Rev.* **1996,** *96,* 2889.
24 Hendriks, J.; Gohlke, U.; Saraste, M. *J. Bioenerg. Biomembr.* **1998,** *30,* 15.
25 Hemp, J.; Gennis, R. *Results Probl. Cell. Differ.* **2008,** *45,* 1.
26 Pereira, M. M.; Santana, M.; Teixeira, M. *Biochim. Biophys. Acta* **2001,** *1505,* 185.
27 Sousa, F. L.; Alves, R. J.; Ribeiro, M. A.; Pereira-Leal, J. B.; Teixeira, M.; Pereira, M. M. *Biochim. Biophys. Acta* **2011** in press.
28 Ludwig, B.; Bender, E.; Arnold, S.; Hüttemann, M.; Lee, I.; Kadenbach, B. *ChemBioChem* **2001,** *2,* 392.
29 Soulimane, T.; Buse, G.; Bourenkov, G. P.; Bartunik, H. D.; Huber, R.; Than, M. E. *EMBO J.* **2000,** *19,* 1766.
30 Tiefenbrunn, T.; Liu, W.; Chen, Y.; Katritch, V.; Stout, C. D.; Fee, J. A.; Cherezov, V. *PLoS One* **2011,** *6,* e22348.
31 Koutsoupakis, C.; Pinakoulaki, E.; Stavrakis, S.; Daskalakis, V.; Varotsis, C. *Biochim. Biophys. Acta* **2004,** *1655,* 347.
32 Chang, H.; Hemp, J.; Chen, Y.; Fee, J. A.; Gennis, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106,* 16169.
33 Siletsky, S. A.; Belevich, I.; Jasaitis, A.; Konstantinov, A. A.; Wikström, M.; Soulimane, T.; Verkhovsky, M. I. *Biochim. Biophys. Acta* **2007,** *1767,* 1383.
34 Giuffrè, A.; Stubauer, G.; Sarti, P.; Brunori, M.; Zumft, W. G.; Buse, G.; Soulimane, T. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96,* 14718.
35 Saraste, M.; Castresana, J. *FEBS Lett.* **1994,** *341,* 1.
36 Castresana, J.; Saraste, M. *Trends Biochem. Sci.* **1995,** *20,* 443.
37 Forte, E.; Urbani, A.; Saraste, M.; Sarti, P.; Brunori, M.; Giuffrè, A. *Eur. J. Biochem.*

2001, *268*, 6486.
38 Ducluzeau, A.; Ouchane, S.; Nitschke, W. *Mol. Biol. Evol.* **2008,** *25*, 1158.
39 Sassera, D.; Lo, N.; Spis, S.; et al. *Mol. Biol. Evol.* **2011,** *28*, 3285.
40 Brochier-Armanet, C.; Talla, E.; Gribaldo, S. *Mol. Biol. Evol.* **2009,** *26*, 285.
41 Sousa, F. L.; Alves, R. J.; Pereira-Leal, J. B.; Teixeira, M.; Pereira, M. M. *PloS One*, **2011,** *6*, e19117.
42 Iwata, S.; Ostermeier, C.; Ludwig, B.; Michel, H. *Nature* **1995,** *376*, 660.
43 Buschmann, S.; Warkentin, E.; Xie, H.; Langer, J. D.; Ermler, U.; Michel, H. *Science* **2010,** *329*, 327.
44 Hino, T.; Matsumoto, Y.; Nagano, S.; Sugimoto, H.; Fukumori, Y.; Murata, T.; Iwata, S.; Shiro, Y. *Science* **2010,** *330*, 1666.
45 Brzezinski, P.; Gennis, R. *J. Bioenerg. Biomembr.* **2008,** *40*, 521.
46 Wikström, M. *Encyclopedia of Inorg. Chem.* **2006,** John Wiley and Sons.
47 Tsukihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itoh, K.; Nakashima, R.; Yaono, R.; Yoshikawa, S. *Science* **1995,** *269*, 1069.
48 Svensson-Ek, M.; Abramson, J.; Larsson, G.; Törnroth, S.; Brzezinski, P.; Iwata, S. *J. Mol. Biol.* **2002,** *321*, 329.
49 Ludwig, B.; Schatz, G. *Proc. Natl. Acad. Sci. U.S.A.* **1980,** *77*, 196.
50 Muramoto, K.; Hirata, K.; Shinzawa-Itoh, K.; Yoko-o, S.; Yamashita, E.; Aoyama, H.; Tsukihara, T.; Yoshikawa, S. S. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104*, 7881.
51 Qin, L.; Hiser, C.; Mulichak, A.; Garavito, R. M.; Ferguson-Miller, S. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103*, 16117.
52 Han, S.; Ching, Y.; Rousseau, D. L. *Nature* **1990,** *348*, 89.
53 Han, S. W.; Ching, Y. C.; Rousseau, D. L. *J. Biol. Chem.* **1989,** *264*, 6604.
54 Wikström, M. *Biochim. Biophys. Acta* **2011,** in press (doi:10.1016/j.bbabio.2011.10.010)
55 Aoyama, H.; Muramoto, K.; Shinzawa-Itoh, K.; Hirata, K.; Yamashita, E.; Tsukihara, T.; Ogura, T.; Yoshikawa, S. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106*, 2165.
56 Suga, M.; Yano, N.; Muramoto, K.; Shinzawa-Itoh, K.; Maeda, T.; Yamashita, E.; Tsukihara, T.; Yoshikawa, S. *Acta Crystallographica Section D* **2011,** *67*, 742.
57 Ostermeier, C.; Harrenga, A.; Ermler, U.; Michel, H. *Proc. Natl. Acad. Sci. U.S.A.* **1997,** *94*, 10547.
58 Palmer, G.; Babcock, G. T.; Vickery, L. E. *Proc. Natl. Acad. Sci. U.S.A.* **1976,** *73*, 2206.
59 Song, Y.; Michonova-Alexova, E.; Gunner, M. R. *Biochemistry* **2006,** *45*, 7959.
60 Sakaguchi, M.; Shinzawa-Itoh, K.; Yoshikawa, S.; Ogura, T. *J. Bioenerg. Biomembr.* **2010,** *42*, 241.
61 Pratt, D. A.; Pesavento, R. P.; van der Donk, W. A. *Org. Lett.* **2005,** *7*, 2735.
62 Proshlyakov, D. A.; Pressler, M. A.; DeMaso, C.; Leykam, J. F.; DeWitt, D. L.; Babcock, G. T. *Science* **2000,** *290*, 1588.
63 Proshlyakov, D. A.; Pressler, M. A.; Babcock, G. T. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** *95*, 8020.
64 Peter, B. *Trends Biochem. Sci.* **2004,** *29*, 380.
65 Brzezinski, P.; Larsson, G. *Biochim. Biophys. Acta* **2003,** *1605*, 1.
66 Brändén, G.; Gennis, R. B.; Brzezinski, P. *Biochim. Biophys. Acta* **2006,** *1757*, 1052.
67 Artzatbanov, V. Y.; Konstantinov, A. A.; Skulachev, V. P. *FEBS Lett.* **1978,** *87*, 180.
68 Capitanio, G.; Martino, P. L.; Capitanio, N.; Papa, S. *Biochim. Biophys. Acta* **2011,** *1807*, 1287.
69 Tsukihara, T.; Shimokata, K.; Katayama, Y.; Shimada, H.; Muramoto, K.; Aoyama, H.; Mochizuki, M.; Shinzawa-Itoh, K.; Yamashita, E.; Yao, M.; Ishimura, Y.; Yoshikawa, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** *100*, 15304.
70 Wikström, M.; Verkhovsky, M. I. *Biochim. Biophys. Acta* **2007,** *1767*, 1200.
71 Siegbahn, P. E. M.; Blomberg, M. R. A. *Biochim. Biophys. Acta* **2007,** *1767*, 1143.
72 Kaila, V. R. I.; Verkhovsky, M. I.; Hummer, G.; Wikström, M. *Proc. Natl. Acad. Sci.*

U.S.A. **2008,** *105,* 6255.
73 Pisliakov, A. V.; Sharma, P. K.; Chu, Z. T.; Haranczyk, M.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105,* 7726.
74 Fetter, J. R.; Qian, J.; Shapleigh, J.; Thomas, J. W.; García-Horsman, A.; Schmidt, E.; Hosler, J.; Babcock, G. T.; Gennis, R. B.; Ferguson-Miller, S. *Proc. Natl. Acad. Sci. U.S.A.* **1995,** *92,* 1604.
75 Brändén, G.; Pawate, A. S.; Gennis, R. B.; Brzezinski, P. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103,* 317.
76 Pawate, A. S.; Morgan, J.; Namslauer, A.; Mills, D.; Brzezinski, P.; Ferguson-Miller, S.; Gennis, R. B. *Biochemistry* **2002,** *41,* 13417.
77 Pfitzner, U.; Hoffmeier, K.; Harrenga, A.; Kannt, A.; Michel, H.; Bamberg, E.; Richter, O. H.; Ludwig, B. *Biochemistry* **2000,** *39,* 6756.
78 Garcia-Horsman, J. A.; Puustinen, A.; Gennis, R. B.; Wikstrom, M. *Biochemistry* **1995,** *34,* 4428.
79 Han, D.; Namslauer, A.; Pawate, A.; Morgan, J. E.; Nagy, S.; Vakkasoglu, A. S.; Brzezinski, P.; Gennis, R. B. *Biochemistry* **2006,** *45,* 14064.
80 Han, D.; Morgan, J. E.; Gennis, R. B. *Biochemistry* **2005,** *44,* 12767.
81 Mitchell, D. M.; Fetter, J. R.; Mills, D. A.; Ädelroth, P.; Pressler, M. A.; Kim, Y.; Aasa, R.; Brzezinski, P.; Malmström, B. G.; Alben, J. O.; Babcock, G. T.; Ferguson-Miller, S.; Gennis, R. B. *Biochemistry* **1996,** *35,* 13089.
82 Konstantinov, A. A.; Siletsky, S.; Mitchell, D.; Kaulen, A.; Gennis, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1997,** *94,* 9085.
83 Riistama, S.; Hummer, G.; Puustinen, A.; Brian Dyer, R.; Woodruff, W. H.; Wikström, M. *FEBS Lett.* **1997,** *414,* 275.
84 Hofacker, I.; Schulten, K. *Proteins* **1998,** *30,* 100.
85 Zheng, X.; Medvedev, D. M.; Swanson, J.; Stuchebrukhov, A. A. *Biochim. Biophys. Acta* **2003,** *1557,* 99.
86 Wikström, M.; Verkhovsky, M. I.; Hummer, G. *Biochim. Biophys. Acta* **2003,** *1604,* 61.
87 Liu, J.; Qin, L.; Ferguson-Miller, S. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** *108,* 1284.
88 Verkhovskaya, M. L.; Garcìa-Horsman, A.; Puustinen, A.; Rigaud, J.; Morgan, J. E.; Verkhovsky, M. I.; Wikström, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997,** *94,* 10128.
89 Ädelroth, P.; Svensson Ek, M.; Mitchell, D. M.; Gennis, R. B.; Brzezinski, P. *Biochemistry* **1997,** *36,* 13824.
90 Aagaard, A.; Gilderson, G.; Mills, D. A.; Ferguson-Miller, S.; Brzezinski, P. *Biochemistry* **2000,** *39,* 15847.
91 Hellwig, P.; Behr, J.; Ostermeier, C.; Richter, O. H.; Pfitzner, U.; Odenwald, A.; Ludwig, B.; Michel, H.; Mäntele, W. *Biochemistry* **1998,** *37,* 7390.
92 Tuukkanen, A.; Kaila, V. R. I.; Laakkonen, L.; Hummer, G.; Wikström, M. *Biochim. Biophys. Acta* **2007,** *1767,* 1102.
93 Kaila, V. R. I.; Verkhovsky, M. I.; Hummer, G.; Wikström, M. *Biochim. Biophys. Acta* **2009,** *1787,* 1205.
94 Pomès, R.; Hummer, G.; Wikström, M. *Biochim. Biophys. Acta* **1998,** *1365,* 255.
95 Xu, J.; Voth, G. A. *Biochim. Biophys. Acta* **2008,** *1777,* 196.
96 Dürr, K. L.; Koepke, J.; Hellwig, P.; Müller, H.; Angerer, H.; Peng, G.; Olkhova, E.; Richter, O. H.; Ludwig, B.; Michel, H. *J. Mol. Biol.* **2008,** *384,* 865.
97 Namslauer, A.; Aagaard, A.; Katsonouri, A.; Brzezinski, P. *Biochemistry* **2003,** *42,* 1488.
98 Wikström, M.; Verkhovsky, M. I. *Biochim. Biophys. Acta* **2006,** *1757,* 1047.
99 Wikström, M.; Verkhovsky, M. I. *Biochim. Biophys. Acta* **2011,** *1807,* 1273.
100 Mills, D. A.; Ferguson-Miller, S. *Biochim. Biophys. Acta* **1998,** *1365,* 46.
101 Tomson, F. L.; Morgan, J. E.; Gu, G.; Barquera, B.; Vygodina, T. V.; Gennis, R. B. *Biochemistry* **2003,** *42,* 1711.
102 Cukier, R. I. *Biochim. Biophys. Acta* **2005,** *1706,* 134.

103 Tuukkanen, A.; Verkhovsky, M. I.; Laakkonen, L.; Wikström, M. *Biochim. Biophys. Acta 1757*, 1117.
104 Maneg, O.; Malatesta, F.; Ludwig, B.; Drosou, V. *Biochim. Biophys. Acta* **2004,** *1655*, 274.
105 Witt, H.; Malatesta, F.; Nicoletti, F.; Brunori, M.; Ludwig, B. *J. Biol. Chem.* **1998,** *273*, 5132.
106 Mitchell, R.; Rich, P. R. *Biochim. Biophys. Acta* **1994,** *1186*, 19.
107 Rich, P. R.; Meunier, B.; Mitchell, R.; John Moody, A. *Biochim. Biophys. Acta* **1996,** *1275*, 91.
108 Capitanio, N.; Vygodina, T. V.; Capitanio, G.; Konstantinov, A. A.; Nicholls, P.; Papa, S. *Biochim. Biophys. Acta* **1997,** *1318*, 255.
109 Hallen, S.; Nilsson, T. *Biochemistry* **1992,** *31*, 11853.
110 Kannt, A.; Lancaster, C. R. D.; Michel, H. *Biophys. J.* **1998,** *74*, 708.
111 Chance, B.; Leigh Jr., J. S. *Proc. Natl. Acad. Sci. U.S.A.* **1977, 7***4*, 4777.
112 Clore, G. M.; Andréasson, L.; Karlsson, B.; Aasa, R.; Malmström, B. G.; *Biochem J.* **1980,** *185*, 155.
113 Han, S.; Ching, Y. C.; Rousseau, D. L. *Biochemistry* **1990,** *29*, 1380.
114 Collman, J. P.; Sunderland, C. J.; Berg, K. E.; Vance, M. A.; Solomon, E. I. *J. Am. Chem. Soc.* **2003,** *125*, 6648.
115 Blomberg, M. R. A.; Siegbahn, P. E. M.; Babcock, G. T.; Wikström, M. *J. Am. Chem. Soc.* **2000,** *122*, 12848.
116 Kaukonen, M. *J Phys Chem B* **2007,** *111*, 12543.
117 Yoshioka, Y.; Satoh, H.; Mitani, M. *J. Inorg. Biochem.* **2007,** *101*, 1410.
118 Keiji, S. *Prog. Biophys. Mol. Biol.* **2006,** *91*, 83.
119 Wittenberg, J. B.; Wittenberg, B. A.; Peisach, J.; Blumberg, W. E. *Proc. Natl. Acad. Sci. U.S.A.* **1970,** *67*, 1846.
120 Rovira, C.; Ballone, P.; Parrinello, M. *Chem. Phys. Lett.* **1997,** *271*, 247.
121 Proshlyakov, D. A.; Ogura, T.; Shinzawa-Itoh, K.; Yoshikawa, S.; Appelman, E. H.; Kitagawa, T. *J. Biol. Chem.* **1994,** *269*, 29385.
122 Fabian, M.; Palmer, G. *Biochemistry* **1995,** *34*, 13802.
123 Hosler, J. P.; Ferguson-Miller, S.; Calhoun, M. W.; Thomas, J. W.; Hill, J.; Lemieux, L.; Ma, J.; Georgiou, C.; Fetter, J.; Shapleigh, J.; Tecklenburg, M. M. J.; Babcock, G. T.; Gennis, R. B. *J. Bioenerg. Biomembr.* **1993,** *25*, 121.
124 Thomas, J. W.; Calhoun, M. W.; Lemieux, L. J.; Puustinen, A.; Wikström, M.; Alben, J. O.; Gennis, R. B. *Biochemistry* **1994,** *33*, 13013.
125 Mogi, T.; Minagawa, J.; Hirano, T.; Sato-Watanabe, M.; Tsubaki, M.; Uno, T.; Hori, H.; Nakamura, H.; Nishimura, Y.; Anraku, Y. *Biochemistry* **1998,** *37*, 1632.
126 Hemp, J.; Christian, C.; Barquera, B.; Gennis, R. B.; Martinez, T. J. *Biochemistry* **2005,** *44*, 10766.
127 Buse, G.; Soulimane, T.; Dewor, M.; Meyer, H. E.; Blüggel, M. *Protein Sci.* **1999,** *8*, 985.
128 Blomberg, M. R. A.; Siegbahn, P. E. M.; Babcock, G. T.; Wikström, M. *J. Inorg. Biochem.* **2000,** *80*, 261.
129 Blomberg, M. R. A.; Siegbahn, P. E. M.; Wikström, M. *Inorg. Chem.* **2003,** *42*, 5231.
130 Wikström, M. *Proc. Natl. Acad. Sci. U.S.A.* **1981,** *78*, 4051.
131 Verkhovsky, M. I.; Morgan, J. E.; Wikström, M. *Biochemistry* **1994,** *33*, 3079.
132 Belevich, I.; Borisov, V. B.; Verkhovsky, M. I. *J. Biol. Chem.* **2007,** *282*, 28514.
133 Belevich, I.; Verkhovsky, M. I.; Wikström, M. *Nature* **2006,** *440*, 829.
134 Belevich, I.; Bloch, D. A.; Belevich, N.; Wikström, M.; Verkhovsky, M. I. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104*, 2685.
135 Wikström, M.; Krab, K.; Saraste, M. Cytochrome oxidase : A synthesis **1982,** Academic Press.

136 Wikström, M.; Verkhovsky, M. I. *Biochim. Biophys. Acta* **2007,** *1767*, 1200.
137 Siegbahn, P. E. M.; Blomberg, M. R. A.; Blomberg, M. L. *J Phys Chem B* **2003,** *107*, 10946.
138 Daskalakis, V.; Farantos, S. C.; Guallar, V.; Varotsis, C. *J Phys Chem B* **2011,** *115*, 3648.
139 Kadenbach, B.; Hüttemann, M.; Arnold, S.; Lee, I.; Bender, E. *Free Radical Biology and Medicine* **2000,** *29*, 211.
140 Susanne, A. *Mitochondrion* in press.
141 Kadenbach, B.; Arnold, S. *FEBS Lett.* **1999,** *447*, 131.
142 Pitcher, R. S.; Watmough, N. J. *Biochim. Biophys. Acta* **2004,** *1655*, 388.
143 Toledo-Cuevas, M.; Barquera, B.; Gennis, R. B.; Wikström, M.; García-Horsman, J. A. *Biochim. Biophys. Acta* **1998,** *1365*, 421.
144 Han, H.; Hemp, J.; Pace, L. A.; Ouyang, H.; Ganesan, K.; Roh, J. H.; Daldal, F.; Blanke, S. R.; Gennis, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** .
145 Thöny-Meyer, L.; Beck, C.; Preisig, O.; Hennecke, H. *Mol. Microbiol.* **1994,** *14*, 705.
146 Preisig, O.; Zufferey, R.; Thöny-Meyer, L.; Appleby, C. A.; Hennecke, H. *J. Bacteriol.* **1996,** *178*, 1532.
147 Arslan, E.; Kannt, A.; Thöny-Meyer, L.; Hennecke, H. *FEBS Lett.* **2000,** *470*, 7.
148 Preisig, O.; Zufferey, R.; Hennecke, H. *Arch. Microbiol.* **1996,** *165*, 297.
149 Visser, J. M.; De Jong, G. A. H.; De Vries, S.; Robertson, L. A.; Kuenen, J. G. *FEMS Microbiol. Lett.* **1997,** *147*, 127.
150 Stevens, J. M.; Daltrop, O.; Allen, J. W. A.; Ferguson, S. J. *Acc. Chem. Res.* **2004,** *37*, 999.
151 Oh, J.; Kaplan, S. *J. Biol. Chem.* **2002,** *277*, 16220.
152 Varotsis, C.; Babcock, G. T.; Garcia-Horsman, J.; Gennis, R. B. *J. Phys. Chem.* **1995,** *99*, 16817.
153 Wang, J.; Gray, K. A.; Daldal, F.; Rousseau, D. L. *J. Am. Chem. Soc.* **1995,** *117*, 9363.
154 Pitcher, R. S.; Brittain, T.; Watmough, N. J. *Biochemistry* **2003,** *42*, 11263.
155 Stavrakis, S.; Koutsoupakis, K.; Pinakoulaki, E.; Urbani, A.; Saraste, M.; Varotsis, C. *J. Am. Chem. Soc.* **2002,** *124*, 3814.
156 Hemp, J.; Han, H.; Roh, J. H.; Kaplan, S.; Martinez, T. J.; Gennis, R. B. *Biochemistry* **2007,** *46*, 9963.
157 Rauhamäki, V.; Baumann, M.; Soliymani, R.; Puustinen, A.; Wikström, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103*, 16135.
158 Hemp, J.; Robinson, D. E.; Ganesan, K. B.; Martinez, T. J.; Kelleher, N. L.; Gennis, R. B. *Biochemistry* **2006,** *45*, 15405.
159 Rauhamäki, V.; Bloch, D. A.; Verkhovsky, M. I.; Wikström, M. *J. Biol. Chem.* **2009,** *284*, 11301.
160 Goodin, D. B.; McRee, D. E. *Biochemistry* **1993,** *32*, 3313.
161 Huang, Y.; Reimann, J.; Lepp, H.; Drici, N.; Ädelroth, P. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105*, 20257.
162 Huang, Y.; Reimann, J.; Singh, L. M. R.; Ädelroth, P. *Biochim. Biophys. Acta* **2010,** *1797*, 724.
163 Smith, M. A.; Finel, M.; Korolik, V.; Mendz, G. L. *Arch. Microbiol.* **2000,** *174*, 1.
164 Watmough, N. J.; Field, S. J.; Hughes, R. J. L.; Richardson, D. J. *Biochem. Soc. Trans.* **2009,** *37*, 392.
165 Fujiwara, T.; Fukumori, Y. *J. Bacteriol.* **1996,** *178*, 1866.
166 Field, S. J.; Thorndycroft, F. H.; Matorin, A. D.; Richardson, D. J.; Watmough, N. J. *Methods Enzymol.* **2008,** 437, 79.
167 Hino, T.; Matsumoto, Y.; Nagano, S.; Sugimoto, H.; Fukumori, Y.; Murata, T.; Iwata, S.; Shiro, Y. *Science* **2010,** *330*, 1666.
168 Timóteo, C. G.; Pereira, A. S.; Martins, C. E.; Naik, S. G.; Duarte, A. G.; Moura, J. J. G.; Tavares, P.; Huynh, B. H.; Moura, I. *Biochemistry* **2011,** *50*, 4251.

169 Flock, U.; Lachmann, P.; Reimann, J.; Watmough, N. J.; Ädelroth, P. *J. Inorg. Biochem.* **2009,** *103,* 845.
170 Flock, U.; Thorndycroft, F. H.; Matorin, A. D.; Richardson, D. J.; Watmough, N. J.; Ädelroth, P. *J. Biol. Chem.* **2008,** *283,* 3839.
171 Maizel, J. V.; Lenk, R. P. *Proc. Natl. Acad. Sci. U.S.A.* **1981,** *78,* 7665.
172 Needleman, S. B.; Wunsch, C. D. *J. Mol. Biol.* **1970,** *48,* 443.
173 Smith, T. F.; Waterman, M. S. *J. Mol. Biol.* **1981,** *147,* 195.
174 Notredame, C. *PLoS Comput. Biol.* **2007,** *3,* e123.
175 Pei, J. *Curr. Opin. Struct. Biol.* **2008,** *18,* 382.
176 Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. *Nucleic Acids Res.* **2002,** *30,* 3059.
177 Edgar, R. C. *Nucleic Acids Res.* **2004,** *32,* 1792.
178 Wheeler, T. J.; Kececioglu, J. D. *Bioinformatics* **2007,** *23,* i559.
179 Notredame, C.; Higgins, D. G.; Heringa, J. *J. Mol. Biol.* **2000,** *302,* 205.
180 Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994,** *22,* 4673.
181 Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990,** *215,* 403.
182 O'Sullivan, O.; Suhre, K.; Abergel, C.; Higgins, D. G.; Notredame, C. *J. Mol. Biol.* **2004,** *340,* 385.
183 Yang, Z. *Curr. Opin. Struct. Biol.* **2009,** *19,* 145.
184 Yang, Z. *Curr. Opin. Struct. Biol.* **2008,** *18,* 342.
185 Zhang, Y.; DeVries, M. E.; Skolnick, J. *PLoS Comput. Biol.* **2006,** *2,* e13.
186 Battey, J. N. D.; Kopp, J.; Bordoli, L.; Read, R. J.; Clarke, N. D.; Schwede, T. *Proteins* **2007,** *69,* 68.
187 Tramontano, A.; Morea, V. *Proteins* **2003,** *53,* 352.
188 Cozzetto, D.; Tramontano, A. *Proteins* **2005,** *58,* 151.
189 Lipman, D. J.; Pearson, W. R. *Science* **1985,** *227,* 1435.
190 Bowie, J. U.; Lüthy, R.; Eisenberg, D. *Science* **1991,** *253,* 164.
191 Jones, D. T.; Taylor, W. R.; Thornton, J. M. *Nature* **1992,** *358,* 86.
192 Dunbrack Jr. R. L. *Curr. Opin. Struct. Biol.* **2006,** *16,* 374.
193 Lemer, C. M. R.; Rooman, M. J.; Wodak, S. J. *Proteins* **1995,** *23,* 337.
194 Šali, A.; Blundell, T. L. *J. Mol. Biol.* **1993,** *234,* 779.
195 Chung, S. Y.; Subbiah, S. *Structure* **1996,** *4,* 1123.
196 Ohlson, T.; Wallner, B.; Elofsson, A. *Proteins* **2004,** *57,* 188.
197 Wang, G.; Dunbrack, R. L. *Protein Sci.* **2004,** *13,* 1612.
198 Michael, L. *J. Mol. Biol.* **1992,** *226,* 507.
199 Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. *Nucleic Acids Res.* **2003,** *31,* 3381.
200 Petrey, D.; Xiang, Z.; Tang, C. L.; Xie, L.; Gimpelev, M.; Mitros, T.; Soto, C. S.; Goldsmith-Fischman, S.; Kernytsky, A.; Schlessinger, A.; Koh, I. Y. Y.; Alexov, E.; Honig, B. *Proteins* **2003,** *53,* 430.
201 Forrest, L. R.; Tang, C. L.; Honig, B. *Biophys. J.* **2006,** *91,* 508.
202 Wallner, B.; Elofsson, A. *Protein Sci.* **2005,** *14,* 1315.
203 Nayeem, A.; Sitkoff, D.; Krystek, S. *Protein Sci.* **2006,** *15,* 808.
204 Moult, J. *Curr. Opin. Struct. Biol.* **2005,** *15,* 285.
205 Moult, J.; Fidelis, K.; Kryshtafovych, A.; Rost, B.; Hubbard, T.; Tramontano, A. *Proteins* **2007,** *69,* 3.
206 Jones, D. T. *Pharmacogenomics J* **2001,** *1,* 126.
207 Venclovas, C.; Margelevicius, M. *Proteins* **2009,** *77,* 81.
208 Sternberg, M. J.; Bates, P. A.; Kelley, L. A.; MacCallum, R. M. *Curr. Opin. Struct. Biol.* **1999,** *9,* 368.
209 Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *Journal of Applied Crystallography* **1993,** *26,* 283.
210 Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray,

L. W.; Arendall, W. B.; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. *Nucleic Acids Res.* **2007,** *35*, W375.
211 Melo, F.; Feytmans, E. *J. Mol. Biol.* **1998,** *277*, 1141.
212 Wiederstein, M.; Sippl, M. J. *Nucleic Acids Res.* **2007,** *35*, W407.
213 Fernandes, M. X.; Kairys, V.; Gilson, M. K. *Scientific World Journal* **2006,** *6*, 1542.
214 Fan, H.; Mark, A. E. *Protein Sci.* **2004,** *13*, 211.
215 Lee, M. R.; Baker, D.; Kollman, P. A. *J. Am. Chem. Soc.* **2001,** *123*, 1040.
216 Jamroz, M.; Kolinski, A. *BMC Structural Biology* **2010,** *10*, 5.
217 Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L. *Protein Sci.* **2003,** *12*, 2001.
218 Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proteins* **2000,** *40*, 389.
219 Thomas, L. H. *Mathematical Proceedings of the Cambridge Philosophical Society* **1927,** *23*, 542.
220 Fermi, E. *Rend. Accad. Naz. Lincei* **1927,** *6*, 602.
221 Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964,** *136*, B864.
222 Atkins, P.; Friedman, R. Molecular Quantum Mechanics **2005,** Oxford Press.
223 Slater, J. C. *Phys. Rev.* **1929,** *34*, 1293.
224 Koch, W.; Holthausen, M. C. A chemist's guide to Density Functional Theory, **2001,** Wiley-VCH.
225 Hoffmann, R. *J. Chem. Phys.* **1963,** *39*, 1397.
226 Kohn, W.; Sham, L. J. *Phys. Rev.* **1965,** *140*, A1133.
227 Stowasser, R.; Hoffmann, R. *J. Am. Chem. Soc.* **1999,** *121*, 3414.
228 Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996,** *100*, 12974.
229 Baerends, E. J.; Gritsenko, O. V. *J Phys Chem A* **1997,** *101*, 5383.
230 Lennard-Jones, J. E. *Trans. Faraday Soc.* **1929,** *25*, 668.
231 Slater, J. C. *Phys. Rev.* **1930,** *36*, 57.
232 Boys, S. F. *Proc. R. Soc. Lond. A* **1950,** *200*, 542.
233 Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969,** *51*, 2657.
234 Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972,** *56*, 2257.
235 Hariharan, P. C.; Pople, J. A. *Theoret. Chim. Acta* **1973,** *28*, 213.
236 Davidson, E. R.; Feller, D. *Chem. Rev.* **1986,** *86*, 681.
237 Jones, R. O.; Gunnarsson, O. *Rev. Mod. Phys.* **1989,** *61*, 689.
238 Ziegler, T. *Chem. Rev.* **1991,** *91*, 651.
239 Vydrov, O. A. *J. Chem. Phys.* **2006,** *125*, 074106.
240 Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980,** *58*, 1200.
241 Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992,** *45*, 13244.
242 Becke, A. D. *Phys. Rev. A* **1988,** *38*, 3098.
243 Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988,** *37*, 785.
244 Perdew, J. P. *Phys. Rev. B* **1986,** *33*, 8822.
245 Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003,** *91*, 146401.
246 Becke, A. D.; Roussel, M. R. *Phys. Rev. A* **1989,** *39*, 3761.
247 Perdew, J. P.; Kurth, S.; Zupan, A.; Blaha, P. *Phys. Rev. Lett.* **1999,** *82*, 2544.
248 Becke, A. *J. Chem. Phys.* **1993,** *98*, 5648.
249 Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994,** *98*, 11623.
250 Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995,** *240*, 283.
251 Von Arnim, M.; Ahlrichs, R. *J. Comp. Chem.* **1998,** *19*, 1746.
252 Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989,** *162*, 165.
253 Klamt, A.; Schuurmann, G. *J. Chem. Soc. , Perkin Trans. 2* **1993,** 799.
254 Schäfer, A.; Klamt, A.; Dattel, D.; Lohrenz, J. C. W.; Eckert, F. *Phys. Chem. Chem.*

Phys. **2000,** *2,* 2187.
255 Andzelm, J.; Kölmel, C.; Klamt, A. *J. Chem. Phys.* **1995,** *103,* 9312.
256 Orio, M.; Pantazis, D.; Neese, F. *Photosynth. Res.* **2009,** *102,* 443.
257 Ryde, U.; Greco, C.; De Gioia, L. *J. Am. Chem. Soc.* **2010,** *132,* 4512.
258 Siegbahn, P. E. M. *Chemistry.* **2008,** *14,* 8290.
259 Kaila, V. R. I.; Oksanen, E.; Goldman, A.; Bloch, D. A.; Verkhovsky, M. I.; Sundholm, D.; Wikström, M. *Biochim. Biophys. Acta* **2011,** *1807,* 769.
260 Bühl, M.; Kabrede, H. *J. Chem. Theory Comput.* **2006,** *2,* 1282.
261 Waller, M. P.; Bühl, M. *J. Comp. Chem.* **2007,** *28,* 1531.
262 Bühl, M.; Reimann, C.; Pantazis, D. A.; Bredow, T.; Neese, F. *J. Chem. Theory Comput.* **2008,** *4,* 1449.
263 Jensen, K. P. Roos, B. O.; Ryde, U. *J. Chem. Phys.* **2007,** *126,* 014103.
264 Siegbahn, P.; Himo, F. *J. Biol. Inorg. Chem.* **2009,** *14,* 643.
265 Sevastik, R.; Himo, F. *Bioorg. Chem.* **2007,** *35,* 444.
266 Hopmann, K. H.; Himo, F. *J. Chem. Theory Comput.* **2008,** *4,* 1129.
267 Hu, L.; Söderhjelm, P.; Ryde, U. *J. Chem. Theory Comput.* **2011,** *7,* 761.
268 Siegbahn, P. *J. Biol. Inorg. Chem.* **2006,** *11,* 695.
269 Curtiss, L. A.; Raghavchari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997,** *106,* 1063.
270 Himo, F. *Theor. Chem. Acc.* **2006,** *116,* 232.
271 Grimme, S.; Antony, J.; Schwabe, T.; Muck-Lichtenfeld, C. *Org. Biomol. Chem.* **2007,** *5,* 741.
272 Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991,** *95,* 9556.
273 Onufriev, A.; Case, D. A.; Ullmann, G. M. *Biochemistry* **2001,** *40,* 3413.
274 Ullmann, G. M.; Knapp, E. *Eur. Biophys. J.* **1999,** *28,* 533.
275 Rabenstein, B.; Knapp, E. *Biophys. J.* **2001,** *80,* 1141.
276 Bashford, D.; Karplus, M. *Biochemistry* **1990,** *29,* 10219.
277 Fogolari, F.; Brigo, A.; Molinari, H. *J. of Mol. Recognit.* **2002,** *15,* 377.
278 Gilson, M. Biophysics textbook online, **2000**.
279 Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins* **1986,** *1,* 47.
280 Gilson, M. K.; Honig, B. *Proteins* **1988,** *4,* 7.
281 Nicholls, A.; Honig, B. *J. Comp. Chem.* **1991,** *12,* 435.
282 Swarén, P.; Maveyraud, L.; Guillet, V.; Masson, J.; Mourey, L.; Samama, J. *Structure* **1995,** *3,* 603.
283 Bashford, D.; Case, D. A.; Dalvit, C.; Tennant, L.; Wright, P. E. *Biochemistry* **1993,** *32,* 8045.
284 Bashford, D.; Gerwert, K. *J. Mol. Biol.* **1992,** *224,* 473.
285 Kannt, A.; Lancaster, C. R. D.; Michel, H. *Biophys. J.* **1998,** *74,* 708.
286 Lancaster, C. R.; Michel, H.; Honig, B.; Gunner, M. R. *Biophys. J.* **1996,** *70,* 2469.
287 Soares, C. M.; Martel, P. J.; Carrondo, M. A. *J. Biol. Inorg. Chem.* **1997,** *2,* 714.
288 Misra, V. K.; Sharp, K. A.; Friedman, R. A.; Honig, B. *J. Mol. Biol.* **1994,** *238,* 245.
289 Karshikoff, A.; Spassov, V.; Cowan, S. W.; Ladenstein, R.; Schirmer, T. *J. Mol. Biol.* **1994,** *240,* 372.
290 Koumanov, A.; Karshikoff, A.; Friis, E. P.; Borchert, T. V. *J Phys Chem B* **2001,** *105,* 9339.
291 van Vlijmen, H. W. T.; Schaefer, M.; Karplus, M. *Proteins* **1998,** *33,* 145.
292 Wlodek, S. T.; Antosiewicz, J.; Mccammon, J. A. *Protein Sci.* **1997,** *6,* 373.
293 Engels, M.; Gerwert, K.; Bashford, D. *Biophys. Chem.* **1995,** *56,* 95.
294 Sandberg, L.; Edholm, O. *Biophys. Chem.* **1997,** *65,* 189.
295 Alexov, E. G.; Gunner, M. R. *Biophys. J.* **1997,** *72,* 2075.
296 Zhang, J.; Gunner, M. R. *Biochemistry* **2010,** *49,* 8043.
297 Gunner, M. R.; Alexov, E. *Biochim. Biophys. Acta* **2000,** *1458,* 63.

298 Baptista, A. M.; Martel, P. J.; Soares, C. M. *Biophys. J.* **1999,** *76,* 2978.
299 Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994,** *238,* 415.
300 Gilson, M. K. *Curr. Opin. Struct. Biol.* **1995,** *5,* 216.
301 Song, Y.; Mao, J.; Gunner, M. R. *Biochemistry* **2006,** *45,* 7949.
302 Yang, A.; Gunner, M. R.; Sampogna, R.; Sharp, K.; Honig, B. *Proteins* **1993,** *15,* 252.
303 Gibas, C. J.; Subramaniam, S. *Biophys. J.* **1996,** *71,* 138.
304 Becker O. M.; Mackerell, A. D.; Roux, B.; Watanabe, M. *Computational Biochemistry and Biophysics* **2001,** CRC Press.
305 Mackerell, A. D. *J. Comp. Chem.* **2004,** *25,* 1584.
306 Brooks, B. R. et al. *J. Comp. Chem.* **2009,** *30,* 1545.
307 Brooks, B. R.; Bruccoleri, R. E.; Olafson, D. J.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983,** *4,* 187.
308 Cornell, W. D.; et al. *J. Am. Chem. Soc.* **1995,** *117,* 5179.
309 Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988,** *110,* 1657.
310 Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comp. Chem.* **2005,** *26,* 1781.
311 Lamoureux, G. *J. Chem. Phys.* **2003,** *119,* 3025.
312 Patel, S.; Brooks, C. L. *J. Comp. Chem.* **2004,** *25,* 1.
313 Patel, S.; Mackerell, A. D.; Brooks, C. L. *J. Comp. Chem.* **2004,** *25,* 1504.
314 Verlet, L. *Phys. Rev.* **1967,** *159,* 98.
315 Frenkel, D.; Smit, B. *Understanding Molecular Simulations,* **2001,** Academic Press.
316 Sindhikara, D. J.; Kim, S.; Voter, A. F.; Roitberg, A. E. *J. Chem. Theory Comput.* **2009,** *5,* 1624.
317 Ryckaert, J.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977,** *23,* 327.
318 Ciccotti, G.; Ryckaert, J. P. *Computer Physics Reports* **1986,** *4,* 346.
319 Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J Phys Chem B* **2007,** *111,* 7812.
320 McQuarrie, D. *Statistical Mechanics* **2008,** University Science Books.
321 Berendsen, H. *J. Chem. Phys.* **1984,** *81,* 3684.
322 Nosé, S. *J. Chem. Phys.* **1984,** *81,* 511.
323 Hoover, W. G. *Phys. Rev. A* **1985,** *31,* 1695.
324 Grest, G. S.; Kremer, K. *Phys. Rev. A* **1986,** *33,* 3628.
325 Martyna, G. Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994,** *101,* 4177.
326 Andersen, H. *J. Chem. Phys.* **1980,** *72,* 2384.
327 Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995,** *103,* 4613.
328 Lindahl, E.; Sansom, M. S. *Curr. Opin. Struct. Biol.* **2008,** *18,* 425.
329 Forrest, L. R.; Sansom, M. S. *Curr. Opin. Struct. Biol.* **2000,** *10,* 174.
330 Norberg, J.; Nilsson, L. *Biophys. J.* **2000,** *79,* 1537.
331 Darden, T. *J. Chem. Phys.* **1993,** *98,* 10089.
332 Humphreys, D. D.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem.* **1994,** *98,* 6885.
333 Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1984,** *5,* 129.
334 Breneman, C. M.; Wiberg, K. B. *J. Comp. Chem.* **1990,** *11,* 361.
335 Hu, H.; Lu, Z.; Yang, W. *J. Chem. Theory Comput.* **2007,** *3,* 1004.
336 Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993,** *97,* 10269.
337 Mulliken, R. *J. Chem. Phys.* **1955,** *23,* 1833.
338 Hu, L.; Ryde. U. *J. Chem. Theory Comput.* **2011,** *7,* 2452.
339 Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996,** *14,* 33.
340 Frishman, D.; Argos, P. *Proteins* **1995,** *23,* 566.
341 Johansson, M. P.; Kaila, V. R. I.; Laakkonen, L. *J. Comp. Chem.* **2008,** *29,* 753.
342 Chen, J.; Brooks, C. L. *Proteins* **2007,** *67,* 922.
343 Anishkin, A.; Milac, A. L.; Guy, H. R. *Proteins* **2010,** *78,* 932.
344 Taly, J.; Marin, A.; Gibrat, J. *BMC Bioinformatics* **2008,** *9,* 6.

345 Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992,** *97,* 2571.
346 Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005,** *7,* 3297.
347 Sierka, M.; Hogekamp, A.; Ahlrichs, R. *J. Chem. Phys.* **2003,** *118,* 9136.
348 Cohen, A. J. *J. Chem. Phys.* **2007,** *126,* 214104.
349 Graèfenstein, J.; Cremer, D. *Mol. Phys.* **2001,** *99,* 981.
350 Zhang, L.; Hermans, J. *Proteins* **1996,** *24,* 433.
351 Thomas, J. W.; Lemieux, L. J.; Alben, J. O.; Gennis, R. B. *Biochemistry* **1993,** *32,* 11173.
352 Javadpour, M. M.; Eilers, M.; Groesbeek, M.; Smith, S. O. *Biophys. J.* **1999,** *77,* 1609.
353 Ortwein, C.; Link, T. A.; Meunier, B.; Colson-Corbisier, A.; Rich, P. R.; Brandt, U. *Biochim. Biophys. Acta* **1997,** *1321,* 79.
354 Wiertz, F. G. M.; Richter, O. H.; Ludwig, B.; de Vries, S. *J. Biol. Chem.* **2007,** *282,* 31580.
355 Heimdal, J.; Rydberg, P.; Ryde, U. *J. Phys. Chem.* **2008,** *112,* 2501.
356 Verkhovsky, M. I.; Morgan, J. E.; Puustinen, A.; Wikström, M. *Nature* **1996,** *380,* 268.
357 Blomberg, L. M.; Blomberg, M. R. A.; Siegbahn, P. E. M. *J. Inorg. Biochem.* **2005,** *99,* 949.
358 Sivaraja, M.; Goodin, D. B.; Hoffman, B. M.; Smith, M. B. *Science* **1989,** *245,* 738.
359 Ullmann, G. M.; Noodleman, L.; Case, D. A. *J. Biol. Inorg. Chem.* **2002,** *7,* 632.
360 Compoint, M.; Ramseyer, C.; Huetz, P. *Chem. Phys. Lett.* **2004,** *397,* 510.
361 Izrailev, S.; Crofts, A. R.; Berry, E. A.; Schulten, K. *Biophys. J.* **1999,** *77,* 1753.

Acknowledgements

This work was performed in the Helsinki Bioenergetics Group at the Institute of Biotechnology and in the Division of Biochemistry, Faculty of Biological and Environmental Sciences at the University of Helsinki. I am thankful to the staff members in these units for providing me full support throughout my PhD studies. The Academy of Finland, Viikki Doctoral Programme in Molecular Biosciences, University of Helsinki, Sigrid Jusélius Foundation and Biocentrum Helsinki are very much acknowledged for the financial support. Computational work presented in this thesis would not have been possible without the powerful computing machines available at CSC - the Finnish IT center for science. I would like to thank them for providing sufficient computing time for my research. Additionally, the computing support by the Laboratory of Instruction in Swedish in Kumpula is very much appreciated.

I would like to express my deepest gratitude to my thesis advisor, Prof. Mårten Wikström, for providing me an opportunity to work on a challenging subject and in a very motivating scientific atmosphere. Your sharp questions, their easy explanations and encouragement to try new things have always been very important in the research work. I would like to thank Dr. Liisa Laakkonen for teaching me the basics of computational biology and chemistry, and for earlier research projects. I have always enjoyed the discussions with you either scientific or non-scientific. Dr. Ville Kaila deserves a very special thanks, for being a good friend and a collaborator, and of course my TURBOMOLE mentor, without which this work would not have been possible. I wish we collaborate in future too.

I am also grateful to the Professors Arto Annila and Adrian Goldman for being the members of my thesis advisory committee, and to Prof. Ulf Ryde and Associate Prof. Pia Ädelroth for critically evaluating my thesis. It has been a rewarding experience working with Prof. Annila on certain separate projects and learning altogether new things.

The support and encouragement from the past and present HBG members is very much appreciated and a special acknowledgement to our group's former secretary Satu Sankkila for her help in all the practical and administrative matters during my earlier stay in Finland. Sharing the lobby and coffee room with the NMR group people has been great too. Special thanks to Olli Aitio, talking and discussing with you is always fun.

I would also like to take pleasure in thanking my Indian friends in Finland, Pali, Sahu, Dhaka, Nikhil, Gopal, Laxman, Tripathi, for contributing to my social life. Wolfgang, Juha, Shetal, Jani, Jussi, I have known from my Masters studies in Biochemistry department, and I have enjoyed each and every moment with them either at work or elsewhere! Wolfi, many thanks to you for being a good friend.

My parents in India have always supported and encouraged all my endeavors, and have celebrated each and every success of mine with happiness and satisfaction. All this would not have been possible without their love and care. My wife Preeti deserves a very special thanks for all the warm support and care and making my evenings full of smiles and laughs with her short funny everyday stories. Our little daughter Paivi, who has made our lives jovial, gives everyday a much more meaning than anything else.

Vivek Sharma
March 2012, Helsinki