

**Molecular Dynamics Simulations and Hydrogen-Bonded Network
Dynamics of Cytochrome *c* Oxidase from *Paracoccus denitrificans*.**

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Physik
der Johann Wolfgang Goethe – Universität
in Frankfurt am Main

von
Elena Olkhova
aus Moskau

Frankfurt am Main 2003

vom Fachbereich Physik der
Johann Wolfgang Goethe – Universität als Dissertation angenommen

Dekan:

Gutachter: Prof. Dr. Werner Mäntele
Prof. Dr. Hartmut Michel

Datum der Disputation: x. x. 200x

CONTENTS

Deutschsprachige Zusammenfassung.	i - vi
SUMMARY	5
CHAPTER 1. Cytochrome <i>c</i> oxidase.	
1.1 Structure and function of cytochrome <i>c</i> oxidase.	7-10
1.2 Catalytic reaction of cytochrome <i>c</i> oxidase.	10-13
1.3 Coupling of electron and proton transfer.	13-14
CHAPTER 2. Water networks in cytochrome <i>c</i> oxidase from <i>Paracoccus denitrificans</i> .	
2.1 Role of the water in the function of cytochrome <i>c</i> oxidase.	15-16
2.2 Water in protein cavities: computational methods to identify internal water.	17-20
2.3 Previous computational work on cytochrome <i>c</i> oxidase.	20-22
2.4 Computer simulation of internal water in COX using the GRID method.	
2.4.1 'GRID' program.	23
2.4.2 Coordinates.	23-25
2.4.3 Internal water modeling.	25-27
2.4.4 Thermodynamic properties of protein cavities.	27
2.4.4 Predicted water binding regions.	28-31
CHAPTER 3. Computer simulation methodology.	
3.1 Statistical mechanics.	31-33
3.2 Force-field and potential energy function.	34-36
3.3 Energy minimization methods.	36-37
3.4 Equations of motion.	37-38
3.5 Application of constraints in simulations.	39
3.5.1 Constant temperature.	40
3.5.2 Constant pressure.	41-42
3.6 Special techniques.	43-45

3.7 Interfaces.	45
3.7.1 Constant normal pressure and surface area.	45-46
3.7.2 Constant normal pressure and surface tension.	46
 CHAPTER 4. Dynamic water networks in COX investigated by MD simulations.	
4.1 Overview of molecular dynamics simulations of membrane proteins.	47-50
4.2 The previous computational work on cytochrome <i>c</i> oxidase.	51
4.3 Details of the simulation protocol.	
4.3.1 Initial set up of protein-membrane-water system.	52-56
4.3.2 Equilibration and dynamics.	57-58
4.3.3 Analysis of the trajectories from the simulations	
4.3.3.1 Hydrogen bonds.	58-59
4.3.3.2 Root mean square deviations and atomic fluctuations.	59
4.3.3.3 Diffusion coefficient.	59
4.3.3.4 Computation details.	59
4.4 Results of the simulations.	
4.4.1 Average structural properties.	60-62
4.4.2 Water distribution in COX and its dynamic properties.	62-67
4.4.3 The distribution of water in the K- and D- pathways.	67-76
 CHAPTER 5. MD study of COX in different stages of the catalytic cycle.	
5.1 Details of the simulation protocol.	77-78
5.2 Results of the simulations.	79-83
 Overall conclusions and future work.	83-86
 Appendix A. Programs.	88-91
 References	92-103
 Acknowledgements	104
 Lebenslauf	105

Deutschsprachige Zusammenfassung.

Cytochrom-*c*-Oxidase (COX) ist das terminale Enzym der Atmungskette von Mitochondrien und aeroben Bakterien. Dieses Enzym katalysiert den Elektronentransfer von Cytochrom *c* auf molekularen Sauerstoff und die Translokation von Protonen über die innere Mitochondrien- bzw. Bakterienmembran. Der genaue Ablauf dieses Prozesses ist bis jetzt noch nicht verstanden. Die Bioenergetik beschäftigt sich hauptsächlich mit den Mechanismen, die der Sauerstoffaktivierung, dem Elektronentransfer und der Protonenverschiebung zugrunde liegen.

Die Cytochrom-*c*-Oxidase, die in die innere Membran der Mitochondrien und vieler Bakterien integriert ist, ist eines der am intensivsten untersuchten Membranproteine. Sie katalysiert den terminalen Schritt der Zellatmung, den Transfer von vier Elektronen von Cytochrom *c* zum Sauerstoffmolekül (Babcock, G. T., M. Wikström. 1992. Oxygen activation and the conservation of energy in cell respiration. *Nature* 356(6367):301-309). Die Reduktion des Sauerstoffmoleküls zu Wasser ist mit einer Verschiebung ("Pumpen") von vier Protonen über die innere Mitochondrienmembran, bzw. bakterielle Cytoplasmamembran verbunden (Wikström, M. K. F. 1977. Proton pump coupled to cytochrome *c* oxidase in mitochondria. *Nature* 266:271-273). Der daraus entstehende Protonengradient, $\Delta\mu^{\text{H}^+}$, kann von der ATP-Synthase zur Erzeugung von ATP genutzt werden (Abbildung 1). Die O₂-Reduktion findet in einem binuklearen Häm-*a*₃/Cu_B-Zentrum statt.

Diese Reaktion hat eine sehr hohe Aktivierungsenergie; sie läuft nicht spontan ab. Die Erzeugung von Sauerstoff ist stark exergon; eine große Energiemenge wird während dieses Prozesses frei. Die Cytochrom-*c*-Oxidase hat die Aufgabe, das Sauerstoffmolekül zu aktivieren und die Energie aufzufangen, die während dieser Reaktion anfällt. Die Energie wird in Gestalt eines elektrochemischen Protonengradienten über die Membran gespeichert, $\Delta\mu^{\text{H}^+}$. Die Cytochrom-*c*-Oxidase ermöglicht den Aufbau eines Protonengradienten nach zwei Prinzipien: Vektorielle Chemie und Protonenpumpen. Das erste Prinzip ist eine direkte Konsequenz von orientierter Chemie im Inneren des Proteins. Nativ sitzt das Enzym in der Lipidmembran von Mitochondrien und prokaryontischen Zellen, wobei die Elektronen in das Enzym vom Cytochrom *c* aus von einer Membranseite gelangen, während die Protonen, die zur Wassererzeugung erforderlich sind, von der anderen Membranseite kommen. Der

Transport von vier Ladungen durch die Lipidmembran geschieht allein durch chemische Prozesse.

Die atomaren Strukturen der Cytochrom-*c*-Oxidase von *Paracoccus denitrificans* Mitochondrien aus Rinderherz, *Thermus thermophilus* und *Rhodobacter sphaeroides* wurde mittels Röntgenkristallographie bestimmt (Abbildung 3). Die Untereinheit II enthält ein bimetallisches Cu_A-Zentrum, und Untereinheit I enthält zwei redox-aktive Cofaktoren: eine Häm *a* mit niedrigem Spin und ein binukleares Metallzentrum, das durch das Häm *a*₃ und Cu_B gebildet wird. Beide Häme liegen im Inneren des Transmembranteils des Enzyms; der Abstand zwischen den beiden Eisenatomen beträgt etwa 13 Å, wobei beide etwa 20 Å vom Periplasma entfernt liegen und 35 Å vom Cytoplasma. Diese relativ kompakte Gruppierung von Metallzentren in der Untereinheit I liegt nahe an der cytoplasmatischen Oberfläche, wobei es durch einen Protonenkanal mit der negativen wässrigen Phase verbunden ist, so dass entsprechende Protonenreaktionen in Einklang mit den Elektronentransportprozessen ablaufen können. Elektronen vom Cytochrom *c* werden auf das Cu_A-Zentrum übertragen und dann über Häm *a* weiter in das binukleare Zentrum (Michel, H. 1998. The mechanism of proton pumping by cytochrome *c* oxidase. *Proc. Natl. Acad. Sci. USA* 95:12819-12824; Mills, D. A., and S. Ferguson-Miller. 2003. Understanding the mechanism of proton movement linked to oxygen reduction in cytochrome *c* oxidase: lessons from other proteins. *FEBS Letters* 545:47-51). Im binuklearen Zentrum wird Sauerstoff an das Häm-*a*₃-Eisen gebunden, dann reduziert, wobei letztlich zwei Wassermoleküle entstehen. Es existieren zwei zusätzliche, nicht redox-aktive Metallzentren in der Cytochrom-*c*-Oxidase von *Paracoccus denitrificans*: ein Mg/Mn-Bindungszentrum an der Grenzfläche zwischen den Untereinheiten I und II und eine Ca-Bindungsstelle. Die Ergebnisse von ortsgerichteter Mutagenese (Thomas, J. W., A. Puustinen, J. O. Alben, R. B. Gennis, and M. Wikström. 1993. Substitution of asparagine for aspartate-135 in subunit I of the cytochrome *bo* ubiquinol oxidase of *Escherichia coli* eliminates proton-pumping activity. *Biochemistry* 32:10923-10928; Brzezinski, P., P. Adelroth. 1998. Proton-controlled electron transfer in cytochrome *c* oxidase: functional role of the pathways through Glu 286 and Lys 362. *Acta Physiol. Scand. Suppl.* 643:7-16; Mills, D. A., and S. Ferguson-Miller. 1998. Proton uptake and release in cytochrome *c* oxidase: separate pathways in time and space? *Biochim. Biophys. Acta* 1365(1-2):46-52; Aagaard, A., G. Gilderson, D. A. Mills, S. Ferguson-Miller, P. Brzezinski. 2000. Redesign of the proton-pumping machinery of cytochrome *c* oxidase: proton pumping does not require Glu(I-286). *Biochemistry* 39(51):15847-15850; Pfitzner, U., K. Hoffmeier, A. Harrenga, A. Kannt, H. Michel, E. Bamberg, O. M. Richter, B. Ludwig. 2000. Tracing the D-pathway in reconstituted site-

directed mutants of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Biochemistry* 39(23):6756-62) und die Analyse der Strukturdaten (Iwata, S., C. Ostermeier, B. Ludwig, H. Michel. 1995. Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature* 376:660-669) deuten auf die Existenz zweier verschiedener Protonenwege hin (Abbildung 2). Der K-Weg für den Protonentransfer führt zum aktiven Zentrum über den essentiellen Lysinrest Lys354, während der nach Asp124 benannte D-Weg vom Asp124 über einen lösungsmittelgefüllten Hohlraum nach Glu278 führt. Der Protonenweg jenseits von Glu278 ist bisher unklar. Er könnte über eine temporär gebildete Wasserkette direkt in das binukleare Zentrum führen oder alternativ zum Häm- a_3 -Propionat, entweder durch direkte Kontakte aufgrund von Konformationsänderungen zu Glu278 oder durch in der Electronendichtekarte nicht sichtbare Wassermoleküle (Michel, H. 1998. The mechanism of proton pumping by cytochrome *c* oxidase. *Proc. Natl. Acad. Sci. USA* 95:12819-12824). Es wurde vorgeschlagen, dass diese beiden Protonenpfade mit verschiedenen Teilen des katalytischen Zyklus funktional assoziiert sein könnten. Trotz aller Strukturkenntnisse, bleibt der Mechanismus der Kopplung zwischen Redoxprozess und Protonenpumpen weiterhin weitgehend unbekannt.

Ziel dieser Dissertation war die Untersuchung der Kopplung von Elektronenübertragung und Protonentransport in der Cytochrom-*c*-Oxidase von *Paracoccus denitrificans* mittels elektrostatischer und molekulardynamischer Berechnungen. Diese Dissertation beschreibt Modellrechnungen an der Cytochrom-*c*-Oxidase von *Paracoccus denitrificans*. Die gemeinsame Verwendung der hier vorgestellten unterschiedlichen Methoden führte zu überzeugenden Ergebnissen, die ein besseres Verständnis von Experimenten erlauben könnten, und bieten Einsicht in den Mechanismus des Protonentransports in solch komplexen Systemen. Die Ergebnisse dieser theoretischen Arbeit können auch dazu dienen zu verstehen, welche einfachen theoretischen Ansätze fruchtbar sind und welche nicht, und führen zur Entwicklung einiger neuer und genauerer Modelle.

Zunächst wurden in dieser Arbeit die Positionen der Wassermoleküle im Enzym vorhergesagt und ein Modell erstellt. Dann wurden molekulardynamische Simulationen an dem System durchgeführt, in welchem die Cytochrom-*c*-Oxidase im vollständig oxidierten Zustand in eine komplett hydratisierte Dimyristoylphosphatidylcholin-Lipid-Doppelschicht eingebettet ist. Aus der Kombination dieser beiden Ansätze wurde ein dynamisches Modell entwickelt, mit dem sich die Wasserstoffbrücken im Inneren der Cytochrom-*c*-Oxidase beschreiben lassen. Dazu wurden parallel zwei molekulardynamische Modellrechnungen jeweils für die Dauer von 1,125 ns bei konstantem Druck und konstanter Temperatur

durchgeführt. Dabei wurde der unterschiedliche Grad der Hydratisierung des Proteins unter dreidimensional periodischen Randbedingungen bei vollständiger Berücksichtigung der elektrostatischen Wechselwirkung mit einbezogen. Aus diesem Modell konnte die mittlere Anzahl der Wassermoleküle in den für die Protonenleitung wichtigen des K- und des D-Wegen berechnet werden.

Die beobachteten leicht veränderlichen Wasserstoffbrückennetzwerke schaffen zusammen mit der deutlichen Diffusion von einzelnen Wassermolekülen eine Grundlage, den Protonentransport in der Cytochrom-*c*-Oxidase und den Vorgang des Protonenpumpens besser zu verstehen. Es konnte einerseits die Bedeutung des Wasserstoffbrückennetzwerkes belegt und andererseits eine mögliche Verknüpfung von lokalen Strukturänderungen mit Konformationsänderungen in übergeordneten Bereichen der Cytochrom-*c*-Oxidase während des katalytischen Zyklusses gezeigt werden.

Durch Röntgenstrukturanalyse ermittelte Proteinstrukturen können das Protonierungs- / Deprotonierungsgleichgewicht nur in einem begrenzten pH-Intervall zutreffend beschreiben. Das in dieser Arbeit berechnete Modell hingegen ist in der Lage, den Einfluss von pH-Änderungen auf einzelne Aminosäuren zu simulieren und die Auswirkung auf das ganze System zu ermitteln. Am Beispiel der Aminosäure Lys-II-191 konnte gezeigt werden, dass durch die Änderung der Orientierung der Seitenkette einer einzigen Aminosäure sich der Protonierungszustand des gesamten Proteins dramatisch ändert.

Da das Vorkommen interner Wassermoleküle sowie lokale und großräumige Änderungen der Konformation von Aminosäuren mit Redox-Änderungen gekoppelt sein könnte, wurde eine Anzahl molekulardynamischer Modellrechnungen (Kapitel 4 & 5) und Berechnungen zu möglichen Positionen interner Wassermoleküle (Kapitel 2) durchgeführt. Die Rechnungen lieferten eine weit größere mittlere Anzahl von internen Wassermolekülen als in den Strukturbestimmungen gefunden wurde.

Unser System umfasst 100 000 Atome und ist damit viel grösser als typische Simulationen von Membranproteinen, wie Bacteriorhodopsin oder der KcsA-Kanal. Im Folgenden wird beschrieben, wie für die Cytochrome-*c*-Oxidase durch Konstruktion einer DMPC (Dimyristoylphosphatidylcholin) Doppellipidschicht eine geeignete Membranumgebung geschaffen wurde. Wir benutzten ein allgemeines Protokoll für molekulardynamische Simulationen um die Ausgangskonfiguration des Protein-Membran-Wasser Systems zu konstruieren. Das mikroskopische System besteht aus der Cytochrom-*c*-Oxidase (2 Untereinheiten, bestehend aus je 549 und 254 Aminosäuren, 181 Lipiden, 88 internen Wassermolekülen, die in der Kristallstruktur der Oxidase nachgewiesen werden konnten und

24323 Wassermolekülen in der Umgebung der Cytochrom-c-Oxidase (Olkhova, E., M. C. Hutter, M. A. Lill, V. Helms, and H. Michel. Dynamic water networks in cytochrome *c* oxidase from *Paracoccus denitrificans* investigated by molecular dynamics simulations. *Biophys J.*, accepted). Zusätzlich wurden 176 bzw. 755 Wassermoleküle, die mit Hilfe der GRID Methode konstruiert wurden, in das System integriert. 42 Na⁺ und 30 Cl⁻ Ionen wurden eingefügt, um eine 100 mM Salzlösung zu simulieren. (Die Ionen wurden nicht weiter als bis auf 6Å an die Lipiddoppelschicht angenähert. Eine Bildung von Ionenpaaren wurde nicht erlaubt). Nach Solvatisierung bestand das System aus insgesamt 101852 Bzw. 103589 Atomen. Die hier vorgestellten Molekulardynamiksimulationen führen zu einem stabilen System. Die Molekulardynamik Rechnungen erreichten nach etwa 1 ns ein dynamisches Gleichgewicht wegen der Protein – Membran-, als auch der Protein – Wasser-Wechselwirkungen. Ähnliches Verhalten wurde auch von anderen Systemen berichtet (Berneche, S., M. Nina, and B. Roux. 1998. Molecular dynamics simulation of melittin in a dimyristoylphosphatidylcholine bilayer membrane. *Biophys. J.* 75:1603-1618). Das Netzwerk der Wasserstoffbrücken in der Cytochrom-c-Oxidase ist nicht gleichmäßig verteilt und die Anordnung der Wassermoleküle ist veränderlich. Das Protein – Membran – Wasser-Modell erlaubt eine genaue Beschreibung der Struktur und der Beweglichkeit der Wasserstoffbrücken, so dass eine Anzahl permanenter Positionen für Wassermoleküle in den K- und D-Wegen erkannt wurden. Das Netzwerk der Wasserstoffbrücken konnte in viele Richtungen bis oberhalb von Häm *a* und Häm *a*₃ beobachtet werden. In einigen Positionen in der Nähe des binuclearen Zentrums haben Wassermoleküle Lebensdauern von etwa 1 ns, im Allgemeinen aber sind die Wassermoleküle viel mobiler. Die Ergebnisse zeigen bewegliche Wassermoleküle auch in Bereichen in denen mit kristallographischen Methoden kein Wasser gefunden wurde. Eine Erklärung des großen Unterschieds in der Anzahl der internen Wassermoleküle zwischen den Modellrechnungen und den kristallographischen Strukturbestimmungen liegt in deren hohen Beweglichkeit und Austauschrate zwischen inneren und äußeren Positionen.

Während die Ergebnisse unserer Modellrechnungen mit früheren theoretischen Untersuchungen im Allgemeinen übereinstimmen (Hofacker, I., and K. Schulten. 1998. Oxygen and proton pathways in cytochrome *c* oxidase. *Proteins* 30(1):100-107; Riistama, S., G. Hummer, A. Puustinen, R. B. Dyer, W. H. Woodruff, M. Wikström. 1997. Bound water in the protein translocation mechanism of the heme-copper oxidases. *FEBS Letters* 414:275-280; Zheng, X., D. M. Medvedev, J. Swanson, A. A. Stuchebrukhov. 2003. Computer simulation of water in cytochrome *c* oxidase. *Biochim. Biophys. Acta* 1557:99-107; Wikström,

M., M. I. Verkhovsky, G. Hummer. 2003. Water-gated mechanism of proton translocation by cytochrome *c* oxidase. *Biochim. Biophys. Acta* 1604(2):61-5), gibt es doch einige Unterschiede. Der wesentliche betrifft die deutliche Diffusion der einzelnen Wassermoleküle in den D- und K- Wegen, wie auch in dem Bereich, der sie mit dem binuclearen Zentrum verbindet. Die Art des Protonentransports entlang dieser Wege könnte sich von der Art unterscheiden, wie sie über Wasserketten in Gramicidin A und Nanoröhrchen beschrieben wurde. Wir hatten anfangs erwartet, eine kleine Anzahl definierter Netze von Wasserstoffbrücken zu finden, wie beispielsweise für Vitamin B₁₂ beschrieben (Savage, H. 1986. Water structure in vitamin B12 coenzyme crystals. I. Analysis of the neutron and x-ray solvent densities. *Biophys. J.* 50(5):947-956). Jedoch stellte sich heraus, dass in Cytochrome-*c*-Oxidase das System der Wasserstoffbrücken so beweglich und komplex ist, dass es nur mit Abbildungen und Tabellen nicht hinreichend genau zu beschreiben ist. Jeder Pfad entlang Wasserstoffbrücken ist kurzlebig. Protontransport findet eher durch viele wechselnde und unterschiedliche Pfade statt, die sich aus der Umorientierung und dem Austausch von Wassermolekülen ergeben, als über einen einzigen Pfad von Wasserstoffbrücken. Dennoch muss ein Triebkraft für den Protonentransport in eine bestimmte Richtung gegeben sein.

Die MD Modellrechnungen liefern erste Einsichten wie während des katalytischen Zyklus lokale Änderungen in den aktiven Stellen von Cytochrom-*c*-Oxidase mit Konformationsänderungen des ganzen Proteins verknüpft sind, insbesondere des Netzwerks der Wasserstoffbrücken.

SUMMARY

Cytochrome *c* oxidase is the terminal enzyme in the respiratory chain of mitochondria and aerobic bacteria. This enzyme ultimately couples electron transfer from cytochrome *c* to an oxygen molecule with proton translocation across the inner mitochondrial and bacterial membrane. This reaction requires complicated chemical processes to occur at the catalytic site of the enzyme in coordination with proton translocation, the exact mechanism of which is not known at present. The mechanisms underlying oxygen activation, electron transfer and coupling of electron transfer to proton translocation are the main questions in the field of bioenergetics.

The major goal of this work was to investigate the coupling of electron transfer and proton translocation in cytochrome *c* oxidase from *Paracoccus denitrificans*. Different theoretical approaches have been used to investigate the coupling of electron and proton transfer. This thesis presents an internal water prediction scheme in the enzyme and a molecular dynamics study of cytochrome *c* oxidase from *Paracoccus denitrificans* in the fully oxidized state, embedded in a fully hydrated dimyristoylphosphatidylcholine lipid bilayer membrane. Two parallel molecular dynamics simulations with different levels of protein hydration, 1.125 ns each in length, were carried out under conditions of constant temperature and pressure using three-dimensional periodic boundary conditions and full electrostatics to investigate the distribution and dynamics of water molecules and their corresponding hydrogen-bonded networks inside cytochrome *c* oxidase. The average number of solvent sites in the proton conducting K- and D- pathways was determined. The highly fluctuating hydrogen-bonded networks, combined with the significant diffusion of individual water molecules provide a basis for the transfer of protons in cytochrome *c* oxidase, therefore leading to a better understanding of the mechanism of proton pumping. The importance of the hydrogen bonding network and the possible coupling of local structural changes to larger scale changes in the cytochrome *c* oxidase during the catalytic cycle have been shown.

CHAPTER 1.

Cytochrome *c* oxidase.

1.1 Structure and function of cytochrome *c* oxidase.

Cytochrome *c* oxidase (COX), located in the inner membrane of mitochondria and many bacteria, is one of the most intensively studied membrane proteins. It catalyzes the terminal step in cellular respiration, a transfer of four electrons from cytochrome *c* to dioxygen (see Babcock and Wikström, 1992, for review). The reduction of dioxygen to water

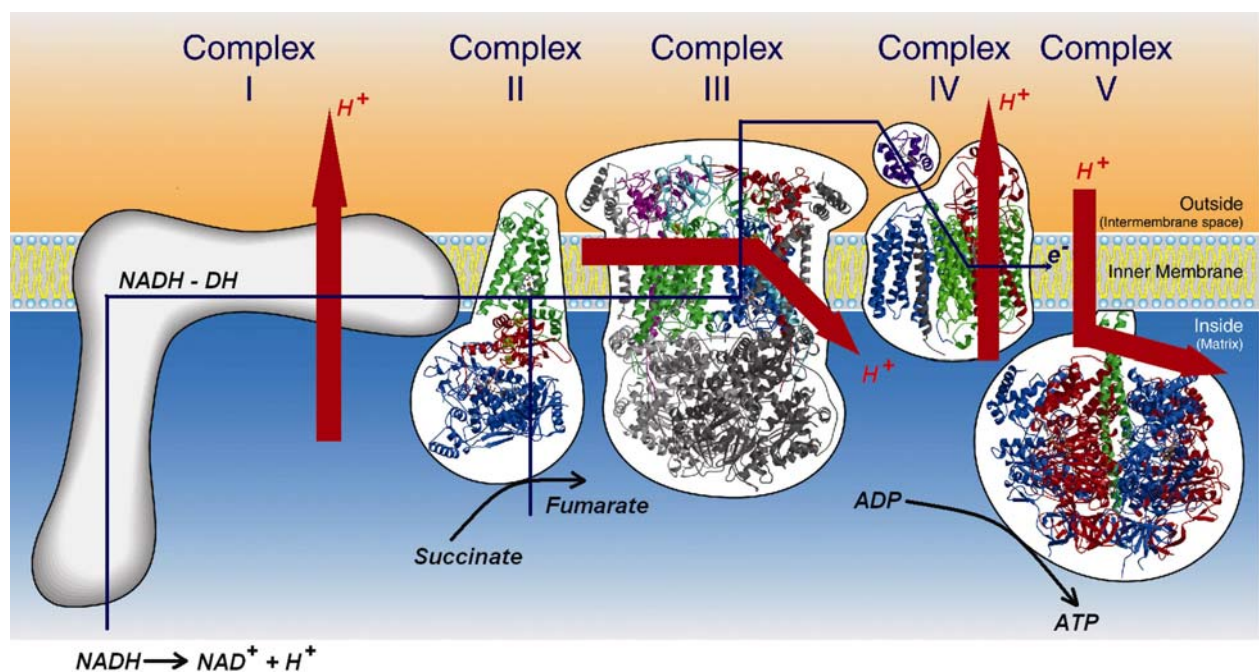


FIGURE 1. Schematic representation of the respiratory chain indicating the pathway of electron transfer from NADH/H⁺ to molecular oxygen via complexes I – IV. This electron transfer is coupled to proton translocation across the membrane resulting in an electrochemical proton gradient.

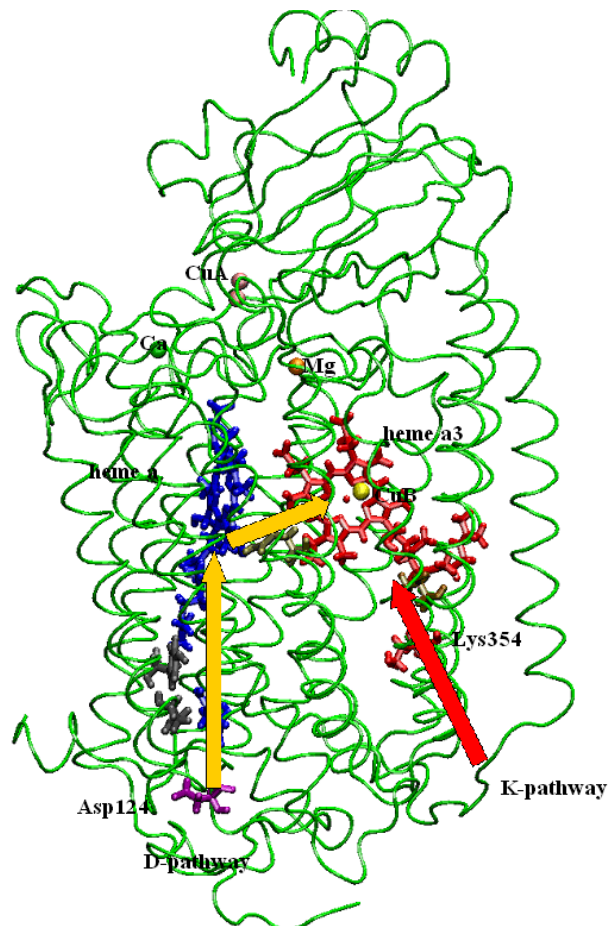
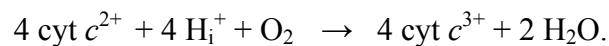


FIGURE 2. Proton transfer pathways in cytochrome *c* oxidase from *Paracoccus denitrificans*. The D-pathway starts at Asp124 and, via a number of polar residues, leads towards Glu278. The K-pathway leads to the binuclear center involving Lys354. At the top, the Mg, Ca, and Cu_A center are shown.

is accompanied by the translocation (“pumping”) of four protons across the inner mitochondrial membrane or the bacterial cytoplasmic membrane (Wikström, 1977). The resulting electrochemical proton gradient, $\Delta\mu_{\text{H}}^+$, can be used by ATP synthase to generate ATP (FIGURE 1). O₂ reduction takes place at a heme *a*₃/Cu_B – binuclear center. Water is produced and protons are consumed according to the following equation:



This reaction has a very high activation barrier and does not occur spontaneously. The reduction of oxygen is highly exergonic and a large amount of energy are released during this process. The function of cytochrome *c* oxidase is to activate the oxygen molecule and to conserve the energy released during this reaction. The energy is conserved in the form of an electrochemical proton gradient across the membrane, $\Delta\mu_{\text{H}}^+$. There are two principles by which COX generates $\Delta\mu_{\text{H}}^+$ - vectorial chemistry and proton pumping. The first principle is a direct consequence of the oriented chemistry inside the protein. Natively, the enzyme is in the lipid membrane of mitochondria or a prokaryotic cell where electrons enter the enzyme from

cytochrome *c* from one side of the membrane and the protons required to make water are taken from the other side of the membrane. Only chemical processes contribute to the transfer of four charges across the lipid membrane.

Atomic structures of cytochrome *c* oxidase from *Paracoccus denitrificans* (Iwata et al., 1995; Ostermeier et al., 1997), bovine heart mitochondria (Tsukihara et al., 1996), *Thermus thermophilus* (Soulimane et al., 2000), and *Rhodobacter sphaeroides* (Svensson-Ek et al., 2002) have been determined by X-ray crystallography (FIGURE 3). Subunit II of COX contains a bimetallic Cu_A center and subunit I contains two redox-active cofactors: a low spin heme *a* and a binuclear metal center formed by the heme *a*₃ and Cu_B (Iwata et al., 1995; reviewed by Ferguson-Miller and Babcock, 1996; Michel, 1998). Both hemes are buried in the transmembrane part of the enzyme; the distance between the two iron atoms is ~13 Å, both being about 20 Å from the periplasm and 35 Å from the cytoplasm. This relatively compact grouping of the metal centers in subunit I is situated towards the cytoplasmic surface, with a proton channel connecting it to the negative aqueous phase so that appropriate protonation reactions can occur in concert with the electron transfer processes. Electrons from cytochrome *c* are transferred to the Cu_A center and then further via heme *a* to the binuclear center (see Babcock and Wikström, 1992; Michel, 1998; Mills and Ferguson-Miller, 2003; Brzezinski and Larsson, 2003 for reviews). Within the binuclear center oxygen is bound to the heme *a*₃ iron, then reduced, and two molecules of water are formed. There are two additional, non-redox active, metal sites in the cytochrome *c* oxidase from *Paracoccus denitrificans*: a Mg/Mn binding site at the interface of subunits I and II, with the metal being ligated by the side chains of His403, Asp404 and Glu-II-218 and one water molecule, and a putative calcium site with the Ca²⁺ being coordinated to the backbone carbonyl oxygen atoms of residues 56, 59 and 61 and the sidechain carbonyl oxygen atoms of Glu56 and Gln63 (Ostermeier et al., 1997).

The results of site-directed mutagenesis experiments (Thomas et al., 1993; Fetter et al., 1995; Garcia-Horsman et al., 1995; Brzezinski and Ädelroth, 1998; Mills and Ferguson-Miller, 1998; Aagaard et al., 2000; Zaslavsky and Gennis, 2000; Pfitzner et al., 2000) and the analysis of the structural data (Iwata et al., 1995) indicate the existence of two different proton transfer pathways (FIGURE 2). The K-pathway of proton transfer leads to the active site through the essential lysine residue Lys354, and the D-pathway (for pumped protons) named after Asp124 leads from the Asp124 via a solvent filled cavity to Glu278. Beyond



FIGURE 4. X-ray structure of cytochrome *c* oxidase from *Paracoccus denitrificans* (picture from Ostermeier *et al* (1997). Structure at 2.7 Å resolution of the *Paracoccus denitrificans* two-subunit cytochrome *c* oxidase complexed with an antibody Fv fragment. (Proc Natl Acad Sci USA. 1997, 94:10547).

Glu278 the pathway of protons is unclear. It may lead directly to the binuclear site via a temporarily established chain of water molecules or to the heme a_3 propionate (Iwata *et al.*, 1995), either by direct contacts upon conformational changes of Glu278, or via unresolved intervening water molecules (reviewed by Michel, 1998). It has been suggested that these two proton pathways may be functionally associated with different parts of the catalytic cycle (Konstantinov *et al.*, 1997). In spite of recent advances in our structural knowledge (Ostermeier *et al.*, 1997; Tsukihara *et al.*, 1996; Svensson-Ek *et al.*, 2002), the mechanism of coupling of the redox processes to proton pumping in COX is largely unknown.

1.2 Catalytic reaction of cytochrome *c* oxidase.

Dioxygen reduction in COX occurs in four steps, one for each electron being transferred to the binuclear center (Siegbahn, 2003). A large amount of spectroscopic information (resonance Raman and optical spectroscopy) has been gathered over the past decades characterizing the intermediates of the binuclear centers appearing during this cycle (for reviews, see Ferguson-Miller *et al.*, 1996; Kitagawa and Ogura, 1997; Michel, 1998; Michel *et al.*, 1998; Morgan *et al.*, 2001). The kinetics of their interconversion has been studied mainly by application of the flow-flash technique to CO-ligated cytochrome *c* oxidase (Gibson and Greenwood, 1963; Greenwood and Gibson, 1967) using differently reduced forms, such as the mixed-valence, the three-electron reduced or the fully reduced form of the enzyme as starting points. An overview of the catalytic cycle of cytochrome *c* oxidase is shown in FIGURE 4.

The scheme starts with an **O**-state (fully oxidised enzyme) containing the heme a_3 iron, a water ligand, an OH^- bound to Cu_B , and the covalently crosslinked Tyr280 to His276 in the active site. Uptake of one electron by the fully oxidase enzyme leads to the formation

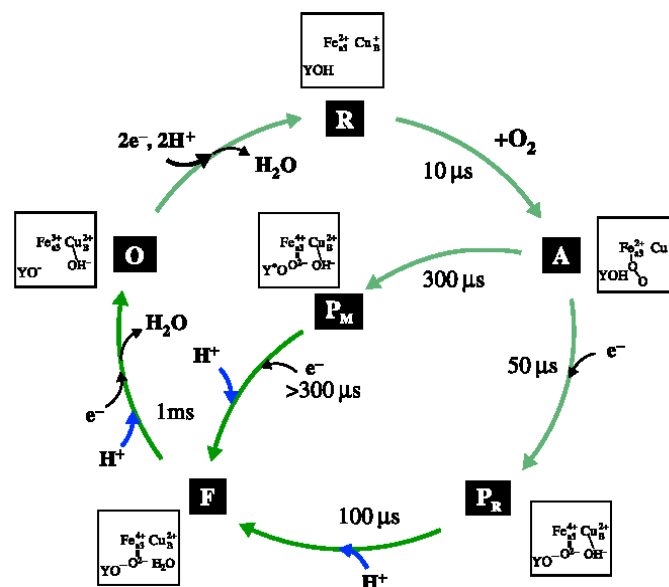
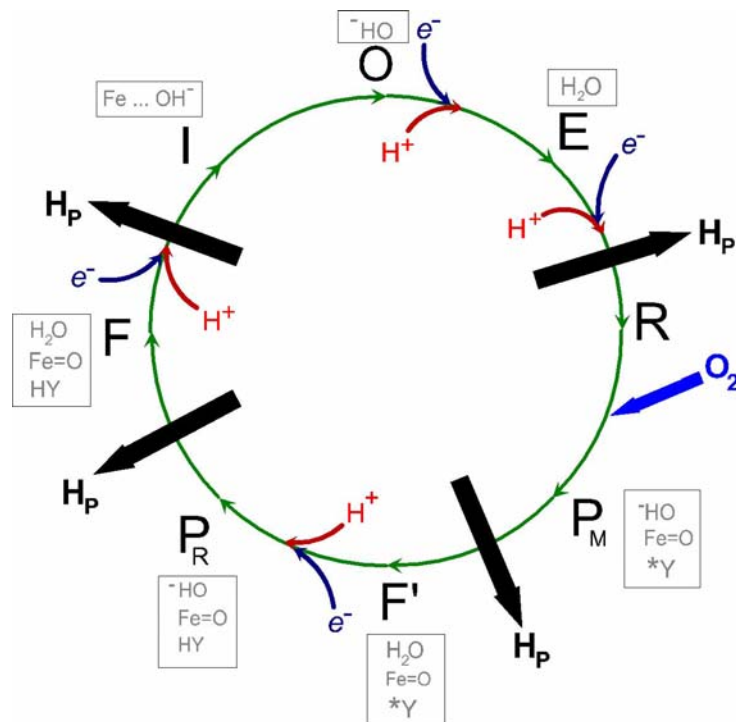


FIGURE 3. A. Catalytic cycle of the cytochrome *c* oxidase (from Michel, 1999). B. The catalytic cycle of cytochrome *c* oxidase (after Brzezinski and Larsson, (2003)).

of the **E**-state (one-electron reduced enzyme), where the electron equilibrates between Cu_A , heme *a* and the binuclear centre (Moody et al., 1991; for review see Michel, 1998). Further electron transfer and net uptake of one proton (Michel, 1999) generates the two-electron reduced state (**R**), where dioxygen can bind to the reduced heme a_3 . A reduction of heme a_3 by the second electron transfer would be impossible, due to the electrostatic repulsion from the already reduced Cu_B , unless this repulsion would be cancelled by protonation of the OH^- ligand of Cu_B . Next, molecular oxygen binds to the reduced binuclear site and compound **A** is formed (Chance, 1975). Compound **A** spontaneously rearranges without net uptake or release of protons to the **P_M** state (oxoferryl state with an OH^- bound to Cu_B). According to Michel, **P_M** will spontaneously form the state **F'** by uptake of a proton via the D-pathway, provided the connection from Glu278 to the binuclear site is established. Next, the third electron transfer to heme *a* could pull in a proton (via the D-pathway) for charge compensation. The proton is probably stored around the heme propionates. Transfer of the electron to the tyrosine radical could convert the latter to a tyrosinate. Therefore, immediate transfer of a proton from the water to the tyrosinate would occur, and the **P_R** state is formed. A negative charge will be stabilized by Cu_B at the Cu_B ligand. OH^- could receive a proton via the D- or K- pathway; the proton would electrostatically repel and pump the proton at the heme propionates. As result the **F** state is formed. The transition starts again with the transfer of an electron to heme *a*, a charge-compensating proton uptake via the D-pathway, and storage of this proton around the heme propionates. This proton will be pumped later. Upon transfer of the electron to heme a_3 another proton will be pulled in via the D- or K-pathway simultaneously and convert the oxoferryl group to a ferric hydroxy group. The incoming proton again would electrostatically repel (and thus pump) the proton around the heme propionates. This intermediate should possess an OH^- group at the heme a_3 iron (Han et al., 1990) and water at Cu_B (**I** state), which then forms the **O**-state again by proton transfer from the water to the OH^- .

A major controversy exists concerning the number of protons being translocated in each part of this cycle (Gennis, 1996). For a long time, it was generally agreed that all four protons are translocated during the oxidative phase, two between **P_M** and **F**, and two between **F** and **O** (Wikström, 1989). More recently this conclusion has been questioned, and it was proposed that only one proton is translocated between **F** and **O** (Michel, 1998). The fourth proton should then be translocated during the reductive phase (Ruitenberget al., 2002). It has also been suggested that only two protons once translocated in the oxidative phase, and the energy needed for translocation of all four protons is made available during this phase

(Verkhovsky et al., 1999). The question of the number of protons being translocated in each step is still not settled.

1.3 Coupling of electron and proton transfer in COX.

We focus here on the question of how oxygen reduction is coupled to proton translocation. There are two leading suggestions for the proton pumping mechanism in cytochrome oxidase (Siegban et al., 2003). The first one is the so-called histidine cycle of Wikström and co-workers (Morgan et al., 1994; Wikström et al., 1998; Wikström, 2000). A key feature of this mechanism is that a gate is present that passes the protons for either pumping or consumption (oxygen reduction). Such a gate is suggested to be needed to prevent the protons from being used directly from the inside in the strongly exergonic O₂ reduction chemistry, and instead be directed toward the endergonic pumping. One cycle of this mechanism in the oxidative phase can be briefly described as follows. An electron is first transferred from heme *a* to the binuclear center. This leads to a large driving force for proton transfer to the binuclear center from the inner side of the membrane. However, these protons cannot reach the binuclear center but are hindered on their way by the conserved His291(325) (bovine numbering) ligand of Cu_B, which is initially in an imidazolate state (Im⁻). When two protons have reached this histidine, leading to an imidazolium group (ImH₂⁺), this group can swing out from the Cu_B centre toward the propionates. This means that the proton gate is open, resulting in proton flow from the inside into the binuclear centre where the charge is annihilated and the chemistry completed. As a consequence, the imidazolium group is destabilized, and protons are released to the outside. When the histidine returns as an imidazolate, the cycle is closed. The net result is thus a translocation of two protons across the membrane and a reduction of the binuclear center. In this mechanism, all the energy used for proton translocation is generated in the oxidative phase. It should be added that it is also possible to translocate only one proton with this mechanism. A detailed scheme assigning all intermediates observed, as well as the motion of all translocated and consumed protons, has been suggested using this mechanism for translocation.

The second leading mechanism of proton translocation in cytochrome oxidase is that proposed by Michel and co-workers, based on electrostatics arguments (Michel, 1998; Michel, 1999). In this model, each of the four reductions of heme *a* during the catalytic cycle is coupled to the uptake of one proton. These protons (not more than two at a time) are temporarily stored in the regions of the heme *a* and heme *a*₃ propionates and are driven to the outside by electrostatic repulsion from protons entering the binuclear centre. A detailed

scheme showing the proton and electron transfers, involving all intermediates observed, has been suggested. An important point of this scheme in comparison to the histidine cycle is that the reductive phase plays a bigger role in proton translocation. Also, heme *a* is more important. The question of how the protons are gated is not clearly addressed in this mechanism (Gennis, 1998). For example, the protons on the propionates have to be prevented from moving to the binuclear centre as the electron moves to this site, otherwise a large amount of available energy could be dissipated.

CHAPTER 2.

Water networks in cytochrome *c* oxidase from *Paracoccus denitrificans*.

2.1 Role of the water in the function of cytochrome *c* oxidase.

Water plays a crucial role in the function of cytochrome *c* oxidase. First of all, water molecules are produced in the enzyme's catalytic site, which is located deeply in the interior of the membrane protein complex. Second, while the electrons required for oxygen reduction can be transferred between redox co-factors without any "assistance" by quantum mechanical tunneling in the protein medium over large distances, the protons, required for the oxygen reduction reaction, as well as those that are pumped by the enzyme, by themselves can move only short distances limited to the length of one hydrogen bond. Chains of water molecules, and hydrogen bonded side chains, are therefore required to assist proton translocation over large distances. Thus, without detailed data about water in the enzyme, the structural and dynamic characterisation of the enzyme is incomplete.

In the X-ray structure of both mammalian and bacterial cytochrome *c* oxidase, large cavities without any clear electron density peak inside, but large enough for trapping more than one water molecule, have been detected, suggesting that mobile waters are located inside the cavities (FIGURE 5). These cavities could provide very effective proton transfer pathways

between protonatable groups. Such cavities are found in the X-ray structure of *Paracoccus denitrificans* cytochrome *c* oxidase, the structure of which has been determined at 2.7 Å resolution (Ostermeier et al., 1997).

Since the crystal structures of cytochrome *c* oxidase from *Paracoccus denitrificans* do not contain all water molecules, that could be involved into the proton transfer pathways, it is necessary to use computer algorithms to determine the likely positions and model internal water molecules. The questions that we address are: 1) How many water molecules are there in the enzyme (only subunit I and II) and how do they move in the protein? 2) What are the dynamic characteristics of water chains for proton transfer? What is their role in proton translocation? The answers to these questions have considerable implications for understanding the function of COX.

Previous computational work that addressed related issues in cytochrome *c* oxidase is described in the next part of Chapter 2, as well as the method which was used for our own computer simulation of water in the *Paracoccus denitrificans* cytochrome *c* oxidase. The dynamic characteristics of the water chains are described in Chapter 4.1 and in Chapter 5.2.

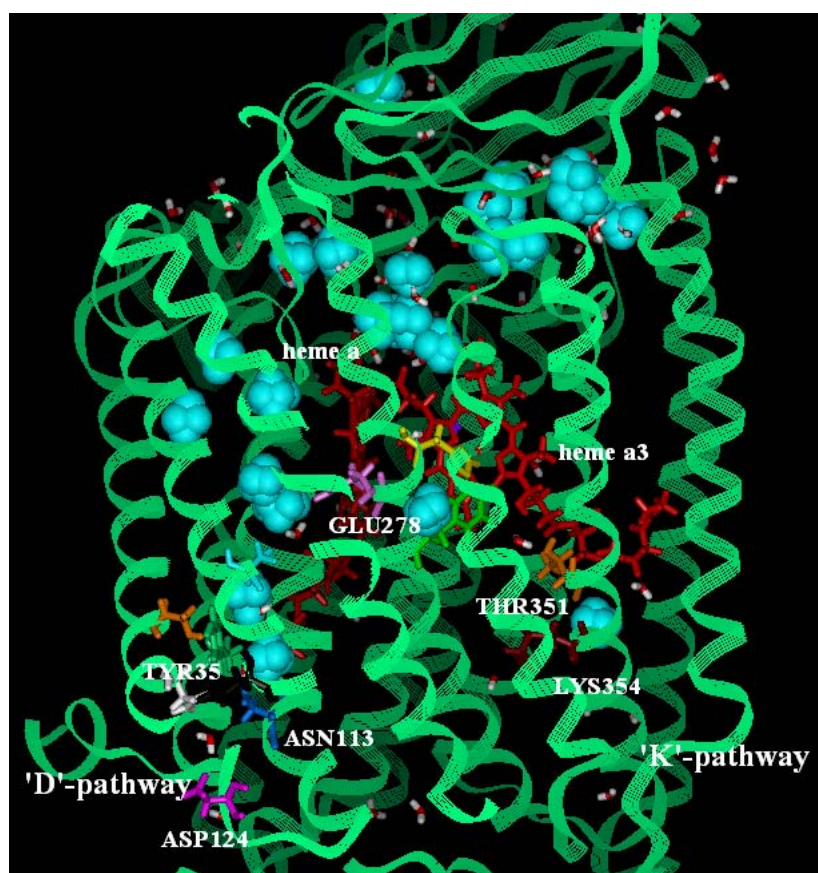


FIGURE 5. Large cavities without any clear electron density peak inside, but large enough for trapping more than one water molecule, can be detected, suggesting that mobile waters are located inside these cavities. Blue surfaces indicate the cavities (cavities were calculated using the program GRASP).

2.2 Water in protein cavities: computational methods to identify internal water.

The interaction of water molecules with proteins plays an important role in biomolecular structure, dynamics, and function. Binding of water molecules at specific sites on the protein surface or in its interior can have an influence on the specificity and function of proteins. The hydration of internal protein cavities has been the subject of a number of experimental, statistical, and theoretical studies. Water molecules in protein solution may be broadly classified into three categories (Finney, 1977; Denisov and Halle, 1996; Bizzarri and Cannistrato, 2002): (i) strongly bound internal water, (ii) water molecules that interact with the protein surface, (iii) bulk water. Internal waters, which can be identified crystallographically and are conserved in homologous proteins, are extensively hydrogen bonded and comprise an integral part of the protein structure (Rupley and Careri, 1991). They have residence times ranging from ~ 10 ns to ms, and their exchange with the bulk solvent requires local unfolding to occur (Denisov and Halle, 1995). Surface water molecules are much less well defined structurally than internal water molecules (in the sense that surface binding sites identified crystallographically are not highly conserved among different crystal forms of the same protein), and are much more mobile, with residence times on the order of tens of picoseconds (Denisov and Halle, 1996). In addition to being important for protein stability, and in the energetics and specificity of ligand binding, surface waters also have a profound influence on the dynamics of protein molecule as a whole.

To date, most knowledge of water properties in proteins has come from experiments (Rupley and Careri, 1991; Denisov et al., 1999). In particular, high-resolution X-ray and neutron diffraction methods can provide detailed information on the presence of well-ordered “structural” water molecules, called bound water, interacting with the protein in the crystal (Carugo and Bordo, 1999). Such water molecules, which are extensively involved in the protein-solvent H-bonding, are believed to be important for protein stability and protein function. However, little information on their lifetime has been obtained, and information on the occupancy of interaction sites can also be difficult to access (Schoenborn et al., 1995). Hydration of proteins can be directly investigated by nuclear magnetic resonance dispersion (NMRD) (Halle et al., 1998). This technique however cannot provide high-resolution information where specific protein-water interaction sites are located. High-resolution NMR spectroscopy can provide information on the presence of water molecules at the atomic level, as well as on their lifetimes (Otting et al., 1997). An issue of particular controversy is the extent to which protein cavities are hydrated. For example, Ernst et al. (1995) have found that

hydrophobic cavities in human interleukin-1 β contained two to four water molecules that reside within the protein for times longer than 1 ns. These water molecules were not observed in high resolution crystal structures, since water with large mean square displacement (MSD) fluctuations ($>1\text{\AA}$) makes negligible contributions to high resolution X-ray diffraction (Yu et al., 1999). However, careful analysis of the low resolution diffraction data by Yu et al. shows that disordered water molecules observed by Ernst et al. are indeed present. Ernst et al. have suggested that hydrophobic cavities commonly observed in protein structures may be filled with disordered water. Disordered water may also play an important role in protein function and dynamics, and may occupy transiently opened cavities inside proteins (see Garcia and Hummer, 2000 for review).

Very often both protein NMR and crystallography provide incomplete information about protein cavity hydration. This can be supplemented by theoretical calculations employing molecular mechanics force fields (Wade et al., 1990; Helms and Wade, 1995; Roux et al., 1996; Hermans and Shankar, 1986; Resat and Mezei, 1994; Oprea et al., 1997). In addition to assisting in the identification of energetically favorable hydration sites (Zhang and Hermans, 1996; Wade and Goodford, 1993), such calculations have also shown that it is sometimes thermodynamically favorable for protein cavities to be empty rather than being filled with solvent (Wade et al., 1991).

Internal cavities are large enough to accommodate one or several water molecules (Bakowies and van Gunsteren, 2002). Detailed characterization of protein cavity hydration becomes more difficult as cavity size increases. The water molecules in large cavities can form extended hydrogen-bonded networks, but these are often rather disordered. The experimentally most well-characterized example of such a network for a biomolecule is bacteriorhodopsin, where water molecules visible in the crystals seem to be involved in a hydrogen-bonded network (Le Coutre et al., 1995; reviewed by Luecke, 2000).

The interaction between the polypeptide chain and the solvent is crucial for both the biological function and the thermodynamic stability of proteins. For example, the dramatic influence of solvent accessibility on the redox potential of metal cofactors has been shown (Langen et al., 1992; Battistuzzi et al., 1997; Springs et al., 2002). Furthermore, metalloproteins present the possibility of redox-state dependent variations in the interaction with the solvent (Bertini et al., 1999; Banci et al., 2000). For example, structural water in the interior of both oxidized and reduced horse-heart cytochrome *c* in solution has been studied by nuclear magnetic resonance (NMR) (Qi et al., 1994) and X-ray crystallography (Bushnell et al., 1990) show deeply buried water molecules that are apparently conserved across the

cytochrome *c* family. In the NMR study by Qi et al., six water molecules with residence times greater than 300ps were located in ferrocycytochrome *c* and five in ferricytochrome *c*. Two of these water molecules are located near the heme. One of these water molecules undergoes a large change in position upon a change of oxidation state, suggesting that buried structural waters in cytochrome *c* may have a role in the solvent reorganization energy associated with electron transfer.

Some theoretical studies have addressed the energetics of hydration of cavities large enough to accommodate more than one water molecule. Roux et al. (Roux et al., 1996) studied a column of four ordered water molecules in bacteriorhodopsin by MD simulations. Wade and Goodford (1993) studied a cavity in human lysozyme with the GRID program that can accommodate five ordered water molecules. Wade (1990) also applied the same technique to examine the hydration of the active site cavity of cytochrome P450_{cam}. Although six water sites have been assigned crystallographically to this cavity (Poulos et al., 1986), they are partly disordered, indicating that a full characterization of its hydration requires an MD simulation approach. The thermodynamically most favorable number of water molecules in the active site of cytochrome P450_{cam} was calculated by Helms and Wade (Helms and Wade, 1998). It was concluded that occupation of the active site by five or six water molecules is favored over occupation by seven or eight water molecules.

A number of methods have been developed in the past to identify interior cavities by analyzing the protein geometry and topology rather than monitoring ligands or solvent in the protein interior (Bakowies and van Gunsteren, 2002). Many of them are based on the concept of solvent-accessible (Lee and Richards, 1971) or molecular contact surfaces (Richards, 1977), which are obtained by rolling spherical probes over the molecule (Connolly, 1983). Cai et al. (1998) suggest approximately Connolly's analytical description by a triangular mesh built around the embracing ellipsoid, which is iteratively deflated until it touches the original surface. Alpha-shape theory (Liang et al., 1998) and related methods (Sastry et al., 1997) also represent a molecule by van der Waals spheres, but they do not depend on auxiliary rolling probes. Instead, they use Delaunay tessellations (O'Rourke, 1995) of the molecule and distinguish between fully and partially occupied tetrahedra to identify both surface pockets and interior cavities (Liang et al., 1998). Alard and Wodak (Alard and Wodak, 1991) present a method to detect cavities by analyzing surfaces built from sets of interpenetrating spheres. Voorintholt et al (1989) and Levitt and Banaszak (1992) proposed procedures that first map the protein structure onto a grid and then use simple distance information between protein atoms and grid points to detect cavities. A modification using pattern recognition techniques

and cellular logic operations has been proposed by Delaney (1992). Although many of these approaches have been used in computer graphics and related applications, they are of limited utility only. In the absence of a closed surface, it appears impossible to discriminate between “interior” and “exterior”.

Recently Bakowies and Van Gunsteren (2002) performed long MD simulations of rat intestinal fatty acid-binding protein (FABD) in its *apo*-form and in complex with palmitate. One of the major objectives was to study both the energetic and dynamic aspects of internal water and to analyze differences between the *apo*- and *holo*- forms of the protein. Such an analysis obviously requires a suitable procedure to distinguish between internal and external solvent. Garcia and Hummer (2000) proposed a classification scheme based on the coordination sphere of water molecules. The procedure is based on the observation that interior or surface water molecules have zero to three water molecules in their first hydration shell, whereas bulk water maintains larger coordination numbers, typically around five. Although this analysis can identify isolated interior water molecules and small interior water clusters, it is not applicable to proteins with water aggregates in large interior cavities. Recently a new computational procedure based on Delaunay triangulation has been presented to define the bounding surface of the protein, to identify internal water molecules, and to determine portals of exchange with bulk water (Bakowies and van Gunsteren, 2002).

2.3 Previous computational work on cytochrome *c* oxidase.

The previous articles dealing with internal hydration of both bacterial and mitochondrial cytochrome *c* oxidase focused on the prediction of bound water molecules within the D- and K-pathways. Another critical point is the role of a water molecule not observed in the crystal structures, which may be a heme a_3 -iron ligand, and/or a hydroxy ion as an additional Cu_B ligand, whose exact position and network of interactions with the protein moiety is thought to be a function of the oxidation state of the enzyme. Riistama et al. (1997) have applied the statistical-mechanical ‘potential of mean force’ (PMF) methodology of Hummer et al. (1996) to predict bound water molecules within proton pathways, and used mutagenesis experiments to show that breaking this water structure impedes proton translocation. The calculations were based on the crystal coordinates of the mitochondrial cytochrome aa_3 from beef heart mitochondria. This approach predicted the existence of several bound water molecules within the D-channel. Three water molecules connect the D-channel to the conserved glutamic acid residue 242 (bovine numbering), and at least three

more water molecules are predicted by the PMF approach to link Glu242 further to one of the Cu_B histidine ligands.

Hofacker and Schulten (1998) positioned water molecules in cytochrome *c* oxidase using the program Dowser (Zhang and Hermans, 1996) and refined the water positions using energy minimization and molecular dynamics simulations (FIGURE 6). They found more than one hundred likely water sites, which give rise to an extended network of hydrogen bonds suitable for proton conduction. The large number of waters suggests an important contribution to the proton pathways suggested for transport of oxygen and protons by predicting the distribution of water molecules in the protein and by molecular dynamics (MD) simulation of oxygen diffusion for COX from *Paracoccus denitrificans*. These calculations

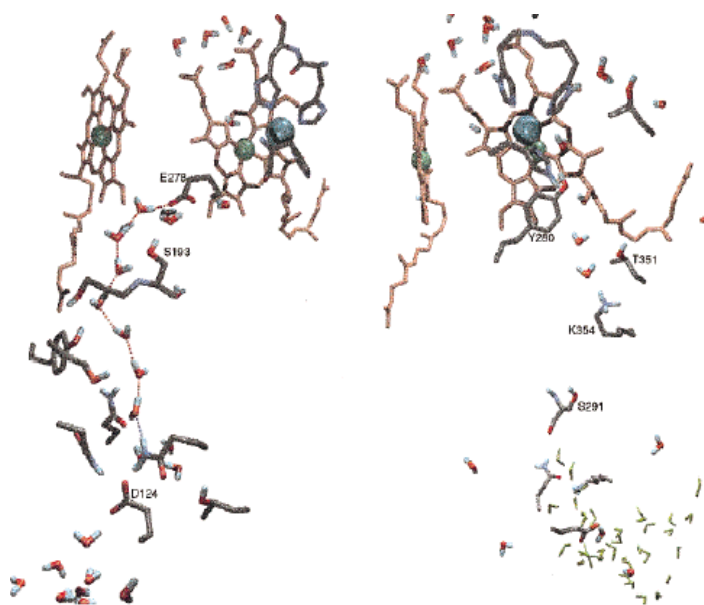


FIGURE 6. (picture from Hofacker I., Schulten K. (1998)). Two channels suitable for proton transfer. Left: A chain of hydrogen bonds (broken lines) involving seven water molecules connects Asp124 to Glu278. Right: Protons could be transferred from Lys354 to the binuclear center via two waters, the heme hydroxyl group and Tyr280. The lysine is flexible and has swung upwards from its original position close to Ser291, bridging the gap to Thr351. An open cavity connects Ser 291 (S291) to bulk water on the inside of the membrane. Water molecules filling this cavity are shown in green to distinguish them from buried waters placed by DOWSER. The two views are rotated by approximately 36° relative to each other.

predicted 130 water molecules within subunits I and II with an interaction energy of less (more favorable) than -12 kcal/mol, but still show large and small cavities, where DOWSER did not place water. A large number of water molecules are predicted in the interface of subunits I and II and near the propionate groups of the hemes, permitting proton exchange with the bulk solvent on the periplasmic side (intermembrane space). The results of these calculations also indicate the existence of two pathways for proton uptake, in good agreement with experimental results. One pathway is formed by a chain of water molecules leading from

Asp124 at the cytoplasmic side of the membrane to Glu278. These simulations show that the hydrogen bonded chain in this channel is formed almost exclusively by water molecules placed by the DOWSER method. The polar residues lining the channel are thought to be important only for stabilizing the water molecules. Another possible proton pathway leads from Ser291, which is in contact with external water molecules, via Lys354 to Tyr280. Two water molecules were placed by DOWSER within this channel, which could contribute to the transport of protons consumed in the reduction of oxygen.

Zheng et al. (2003) placed water molecules in the hydrophilic pockets of the enzyme using the program DOWSER. In the calculations, the chemical potential of bulk water was

varied in the range -10 to -5 kcal/mol. After each DOWSER run the newly placed water molecules were fixed, and the “dowsing” was repeated until no new water was found. After four “dowsings”, a total of 165 water molecules were located in subunits A and B with a chemical potential of -5 kcal/mol (FIGURE 7). The distribution of water in both D- and K-proton channels was not continuous, but was concentrated in the protein cavities in these two channels. Thus, it was concluded that such a water distribution suggests that protein residues play the role of intermediate re-translation sites in the proton translocation process. It was found that the heme pockets, and the catalytic site are hydrophobic, because no waters are predicted at -5 kcal/mol and even higher energy criteria. The calculations also predict a (quasi-) stable chain of water molecules, formed from water produced by the enzyme, connecting Glu242 (bovine numbering), catalytic site, and the propionate *d* of heme *a*₃.

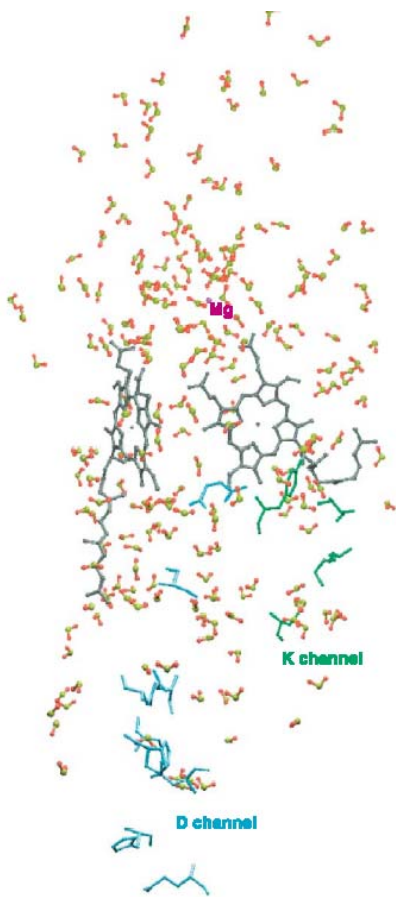


FIGURE 7. Computer simulation of water in cytochrome *c* oxidase. (H. Zheng, D. M. Medvedev, J.Swanson, A.A.Stuchebrukhov, BBA, **2003**, 1557, 99-107).

2.4 Placement of internal water molecules in cytochrome *c* oxidase from *Paracoccus denitrificans* using the GRID method.

2.4.1 ‘GRID’ program.

In order to place water molecules in the two-subunit enzyme of cytochrome *c* oxidase we used the program GRID (Goodford, 1985). This method has already been successful in locating water positions in a number of crystal structures, for example, in cytochrome P450_{cam} (Wade, 1990; Helms and Wade, 1995), and in acetylcholinesterase (Henchman et al., 2002). The program implements a computationally fast method of determining energetically favorable ligand binding sites on molecules of known structure. The probe molecule is moved through the protein matrix on a grid, and at each point, energies are calculated as a sum of Lennard-Jones, electrostatic, and hydrogen-bonding terms with explicit modeling of hydrogen bond geometries, given by the following equations:

$$E = \sum E_{lj} + \sum E_{elec} + \sum E_{hb} \quad (2.1)$$

$$E_{lj} = A / r^{12} - B / r^6 \quad (2.2)$$

$$E_{elec} = q_p q_t / K \zeta \left\{ 1/r + [(\zeta - \epsilon) / (\zeta + \epsilon)] / (r^2 + 4s_p s_t)^{1/2} \right\} \quad (2.3)$$

$$E_{hb} = E_r \times E_t \times E_p \quad (2.4)$$

Energy contour surfaces for the various probes delineate regions of attraction between probe and protein. The total energy, E , is computed as the sum of the pairwise interactions of the probe group with each atom in the target molecule. r is the distance between the probe group and an atom in the target, q_p is the electrostatic charge of the probe, q_t is the electrostatic charge of the target atom, K is a combination of geometrical factors and natural constants, ϵ is the dielectric constant of the solvent, ζ is the dielectric constant of the target phase, s_p is the depth of the probe in the target phase in Å, s_t is the depth of the target atom in the target phase in Å, A and B are parameters describing the types of interacting atoms, and E_r , E_t , and E_p are functions of r , t , and p , respectively, where the angle made by the hydrogen bond at the target is t and at the probe is p (Boobbyer et al., 1989).

2.4.2 Coordinates.

The starting configuration of the fully-oxidized two-subunit COX from *Paracoccus denitrificans* at 2.7 Å resolution was the Protein Data Bank (PDB) entry 1AR1, designated post-1AR1 after further refinement (Harrenga and Michel, unpublished). This structure was

obtained under oxidizing conditions. This set of coordinates contains a Mg^{2+} ion, located at the interface between subunits I and II, that is coordinated by His403, Asp404, Glu(II)218, a Ca^{2+} ion bound to His59, Gly61, Gln63 and Glu56, and 88 water molecules. A hydroxide ion, OH^- , was modeled at a distance of 2.05 Å from the Cu_B , in agreement with previous theoretical calculations (Kannt et al., 1998). Hydrogens were added to the structure by using CHARMM's hbuild function (Brünger and Karplus, 1988).

Charges on the protein atoms and ionizable groups in different protonation states were taken from the CHARMM22 force field (MacKerell et al., 1995; MacKerell et al., 1998, see Chapter 3.2 for an introduction into force fields). Lys354 was assigned as neutral because of its hydrophobic environment. Partial atomic charges for neutral Lys354 as well as for protonated Asp399, and for protonated Glu278 were taken from the AMBER96 force field. Because all titrating protons have been positioned in the structure, we can describe the deprotonation/protonation of the particular residue as the addition of an explicit hydrogen atom.

Electrostatic calculations (in these calculations water molecules were not included explicitly) showed that Lys354 is in its neutral state over a wide pH range (pH 4 – 11.5) (Kannt et al., 1998). However, this has been questioned recently because water molecules hydrogen bonded to Lys362 and Ser299 were found in the *Rhodobacter sphaeroides* COX structure (Svensson-Ek et al., 2002). This observation has led to the postulation that the homologous Lys362 (*Rhodobacter sphaeroides* numbering) is protonated at pH=7 (Svensson-Ek et al., 2002). One heme a_3 propionate shares a proton with Asp399; all other propionates are most likely deprotonated at pH=7 (Kannt et al., 1998).

The atomic partial charges for the redox centers – the heme a and heme a_3 sites, the His276 - Tyr280 crosslink and for Cu_B and its three histidine ligands (His276, His325 and His326) for the fully oxidized enzyme – were kindly provided by Dr. Michael Hutter who obtained them by quantum chemical calculations (FIGURE 8, A – E). Atom centered point charges for the heme a and heme a_3 and the copper center were derived from restraint electrostatic potential fits on model systems using the ESP module of NWChem 4.1 employing the 6-31G* basis set (NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.1. (2002)). Computing partial atomic charges for unusual cofactors by the electrostatic potential (ESP) fit method is not a standard procedure in the CHARMM community and is therefore a potential source of error. However, ESP charges combined with CHARMM gave good agreement with experimental pK_a values in electrostatic continuum calculations for bacterial photosynthetic reaction centers (Rabenstein et al., 1998). For protoporphyrin IX (heme) a 3-

21G optimized structure as well as a high resolution crystal structure geometry were used. However, the charges obtained for the hemes showed only little dependence on the molecular geometry employed. In particular a partial charge of 0.36 only for Fe(III) was found; the ligating nitrogens compensate most of the positive charge. The partial charge of NE2 decreases from -0.70 in a neutral histidine (HSD) to -0.14 in the case of a five-coordinated Fe(III) and to -0.42 for a six-coordinated Fe(III), while the partial charges of the nitrogens on the porphyrin ring change to -0.10 for Fe(III). The calcium ion was modeled with its formal $-2e$ charge. This is not believed very crucial because the ion is strongly coordinated by four residues (see above, Ostermeier *et al.*, 1997). Therefore, its partial atomic charge is shielded from the environment and it is bound in a stable geometry.

2.4.3 Internal water modeling.

To run GRID, a table of parameters (table GRUB) is needed in order to evaluate the Lennard-Jones and other empirical energy function, and program GRIN appends these parameters to the atomic coordinates of the protein. An “extended” water molecule is treated as an electrically neutral group that has no dipole but can donate up to two hydrogen bonds and can also accept up to two. A sufficiently close GRID spacing of 0.5 \AA between GRID points was used to ensure that no appreciable cavity in the matrix of the protein would be missed. The output of GRID provides information on favorable positions for several functional groups.

GRID (Goodford, 1985) was run with a water probe over the two subunit structure of cytochrome *c* oxidase in order to identify possible water binding sites. Due to the large size of $\sim 13\,000$ atoms, the system was split into 8 parts, and GRID was run for each part independently. The largest region with a favourable binding energy in the GRID energy map was in the internal cavity containing Glu278 where a structural water molecule W_S (structural water molecule) was detected. Although this water molecule is clearly defined in the electron density map, GRID was used to position it in the cavity without using the information from the experimental coordinates. GRID calculations over the whole internal cavity region gave an energy of -14 kcal/mol for this water probe, which indicates favorable hydration. After locating the energy minimum its position was refined and a water molecule W_G (internal GRID water molecule) was then assigned to it. GRID was then run again with input coordinates containing this water molecule. A second energy minimum was found and a second water molecule assigned to it. The criteria for water placement in a cavity were: the GRID energy was lower than -8 kcal/mol for the first set of coordinates (‘W8’ water

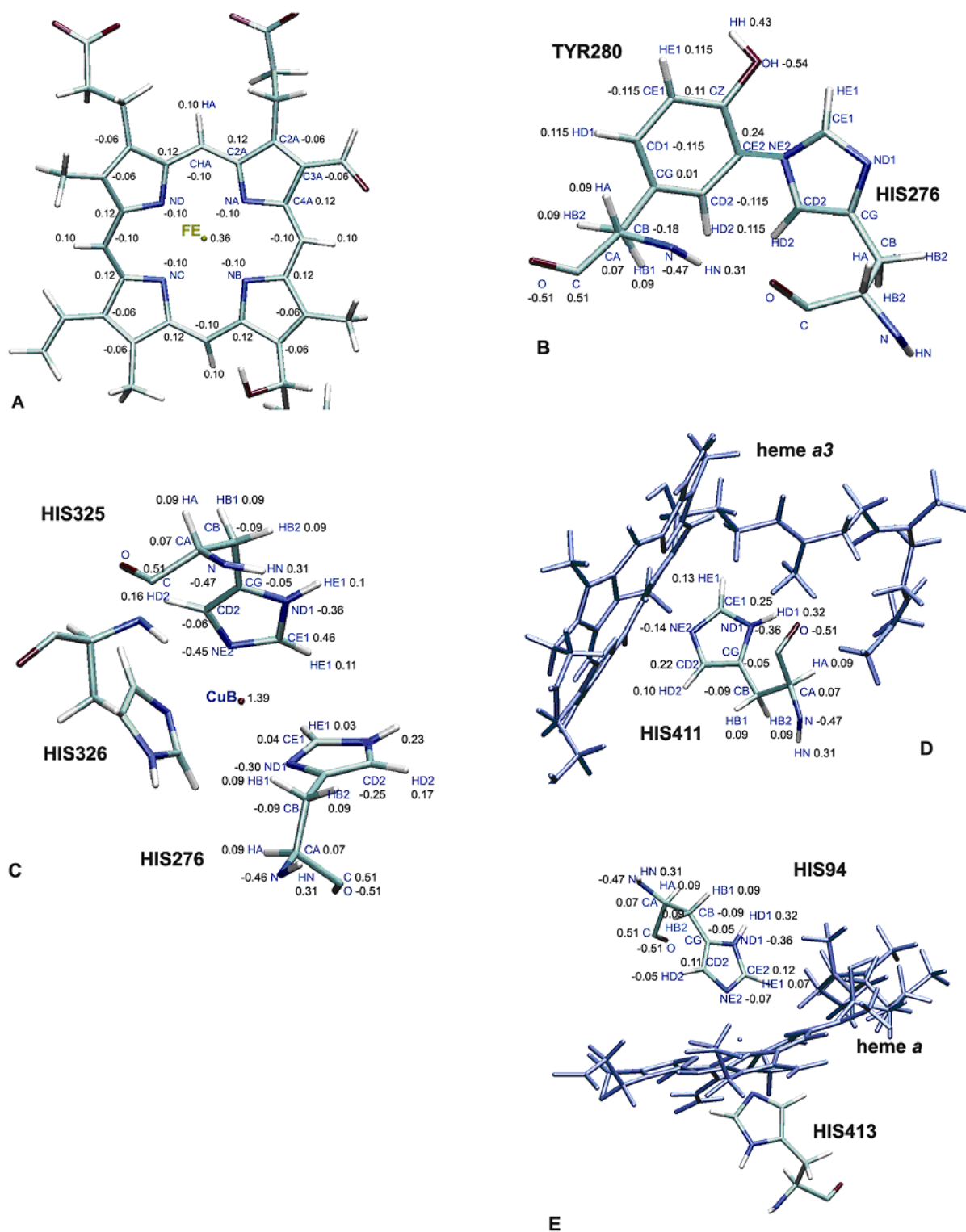


FIGURE 8. A) partial atomic charges for the fully oxydized heme *a* / heme *a*₃ derived from quantum chemical calculations; B) partial atomic charges for the Tyr280 – His276 crosslink (charges for His276 are listed in FIGURE 6.C.); C) partial atomic charges for the histidine ligands to Cu_B. The charges on His326 were restrained to be identical to those of His325 and therefore not shown; D) partial atomic charges for the histidine ligand to the Fe atom of heme *a*₃; E) partial atomic charges for histidine ligands to the Fe atom of heme *a*; the two His ligands carry the same charges.

molecules), and lower than -12 kcal/mol for the second set of coordinates ('W12' water molecules), and each water molecule forms at least two hydrogen bonds (Helms and Wade, 1995). The GRID procedure was repeated four times keeping all previously located water molecules until no suitable deep energy minimum (lower than -8 kcal/mol or lower than -12 kcal/mol) was found, indicating that all cavities were fully solvated for the chosen energy criteria (Wade and Goodford, 1993). Internal water modeling was done on a one processor Alpha machine. The GRID energy function is very short-ranged and electrostatic interactions are computed using a cut-off of only 4.5 Å. Therefore, although potentially of significance, the order of the parts is irrelevant because water molecules in a boxes hardly interact with water molecules of other boxes.

2.4.4 Thermodynamic properties of protein cavities.

Water molecules at molecular interfaces influence the thermodynamics of intramolecular interactions as well as having structural and functional importance (Wade et al., 1991). Because GRID energies are only effective energies, not true Gibbs free enthalpies, it is not quite clear which value to choose for deciding whether predicted hydration sites are occupied or not. Most theoretical work has focused on water in internal protein cavities and on the free energy only (Roux et al., 1996). Dunitz (1994) provided an empirical upper bound of 7 kcal/mol K (about 2 kcal/mol at 300 K) for the entropy cost of transferring a water molecule from the bulk to a binding site. Similar estimates for the enthalpy of water ordering gave -3.8 kcal/mol (Ladbury, 1996). From the simulations for the pure water system, the excess chemical potential of water was computed to be -6.4 ± 0.4 kcal/mol, in accord with experiment (Ben-Naim and Marcus, 1984). For the protein cavity containing an experimentally observed water molecule, the free energy change on hydrating it with one water molecule was calculated as -10.0 ± 1.3 kcal/mol, indicating the high probability that this cavity is occupied by a water molecule. By contrast, for the cavity in which no water molecules were experimentally observed, the free energy change on hydrating it with one water molecule was calculated as 0.2 ± 1.5 kcal/mol, indicating its low occupancy by water. Based on a comparison with free energy perturbation calculations, GRID energies lower than -6 kcal/mol were suggested as a criterion in a previous study on cytochrome P450_{cam} (Helms and Wade, 1995).

2.4.5 Predicted water binding regions.

We have applied the procedure outlined above to identify water molecules in the interior of COX. Here, we tested two levels of hydration, a “looser” criterion of accepting all GRID waters with energies of less than - 8 kcal/mol and a “tighter” criterion by requiring all waters to be more favorable than - 12 kcal/mol. After four GRID cycles for the two subunit structure of COX a total of 755 water molecules were found: 176 molecules with interaction energies below -12 kcal/mol, 267 molecules with interaction energies between -12 kcal/mol and -10 kcal/mol and 314 molecules with interaction energies between -10 kcal/mol and - 8 kcal/mol. We refer to the 176 water molecules with interaction energies below -12 kcal/mol as the ‘W12’ set of water molecules, and to the 755 water molecules with interaction energies below - 8 kcal/mol as the ‘W8’ set. A comparison showed that all the 88 crystallographically observed water molecules are located in positions, formed as energetically favourable by GRID calculations.

The general distributions of internal water molecules detected by X-ray crystallography and found by GRID are in a good agreement. Fewer water molecules were found near the center of the protein, more were identified in the regions close to the surface. The results of the GRID calculations were also compared with the positions of the observed water molecules in the crystal structure (Svensson-Ek et al., 2002) of cytochrome *c* oxidase from *Rhodobacter sphaeroides*. Because the structure has been solved by X-ray crystallography at 2.3 Å resolution, the water positions in the structure of COX from *Rhodobacter sphaeroides* provides a model system for testing the GRID method. In total the structure contains 436 water molecules (O atoms). The comparison showed that positions of water molecules resolved in the crystal structure are similar to positions of water molecules from the ‘W12’ set of coordinates. The number of water molecules from ‘W12’ set also compare reasonably well with the number of well-ordered internal water molecules identified in the crystal structures of the two forms of COX (1M56.pdb and 1M57.pdb) from *Rhodobacter sphaeroides*.

The properties of the predicted water molecules are listed in TABLE 1. They were predicted at a mean distance of 1.2 ± 0.3 Å from the experimentally observed water sites (in agreement with previous studies of Wade and Goodford, 1993). The distances between the predicted and experimentally observed water sites were all less than the mean displacement of 0.7 – 0.9 Å of the waters, computed from the experimental B-factors.

TABLE 1. Comparison of water binding positions between ‘W12’ set of water coordinates predicted by GRID and crystallographically observed water molecules from cytochrome *c* oxidase from *Rhodobacter sphaeroides*.

predicted water molecule	No of predicted water molecules included in the target	closest observed water molecule	distance from observed water (Å)	binding energy at predicted position molecule (kcal/mol)	neighbour partners at predicted position	at observed position
W _G 15	0	W71	1.25	-10.83	HEMA3:O2D W _G 339 W _G 27	HEMA3:O1D HEMA3:OD2 W55
W _G 32	1	W42	0.86	-14.06	PHE116:O SER205:OG W _G 36 W _G 24	PHE108:O W44 W26
W _G 228	0	W83	0.90	-10.54	THR545:O	
W _G 236	0	W45	0.94	-9.02	GLY228:O	ASN121:O
W _G 44	2	W50	1.32	-14.08	ASP(II)193 W _S 86 W _S 12 W _G 424	ASP(II)229:OD2 GLU(II)254:OE1 W209

TABLE 1 shows that the experimentally observed water sites are in energetically favourable positions. For all these water sites, the calculated binding energy is better than -9 kcal/mol. The water molecule W_G44 was found in the large region where the probe was calculated to have a weak attractive binding energy. Four water molecules have been located within this large cavity; it seems that these molecules could not be resolved by the X-ray experiments. Displacement of these tightly bound water molecules would be favourable as it would result in a gain in entropy. It has been found that the inclusion of the water molecule W_G69 in the active site of the protein provides a strong connection between the Fe atom of the heme *a*₃ and Cu_B; the calculated binding energy of this water molecule is -13.69 kcal/mol.

All water molecules from the ‘W12’ and ‘W8’ sets of coordinates are examined to the possible formation of hydrogen bonding network requires for the proton pumping in cytochrome *c* oxidase. We began by constructing proton-conducting networks based on the atomic structure of the protein obtained from X-ray crystallography. As input information, we used the positions of heavy atoms such as oxygen, nitrogen, and sulfur that can participate in proton transfer, as listed in TABLE 2 .

TABLE 2 . Donor and acceptor atoms in proton transfer networks (from Taraphder and Hummer, 2003):

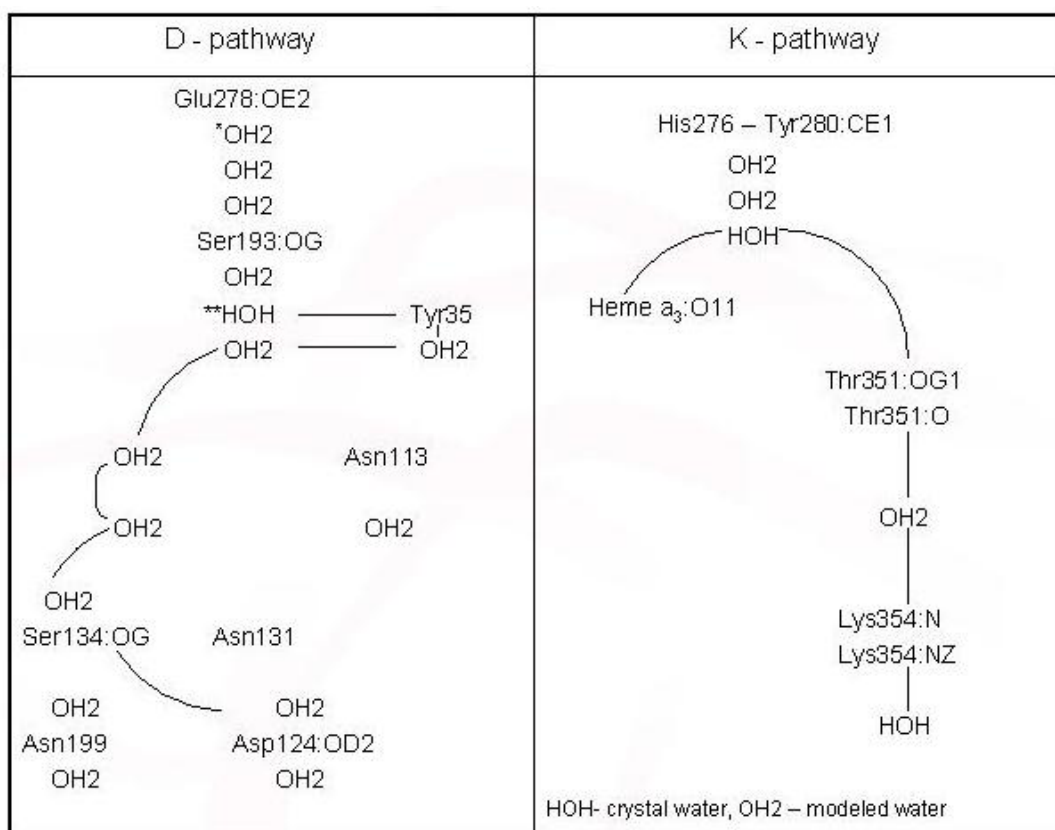
Residue	atom
ASP	OD1, OD2
GLU	OE1, OE2
HIS	ND1, NE2
CYS	SG
TYR	OH
SER	OG
THR	OG1
ARG	NH1, NH2, NE
HOH	O
HEM	O1A, O2A, O1D, O2D

In the simplest model that we investigated prior to the MD simulations, the coordinates of hydrogen atoms were not included. A pair of donor and acceptor atoms within a distance of 3.5 Å was examined. This distance is chosen to be larger than the hydrogen bonding distance of about 2.5 Å to account for molecular flexibility and dynamic effects. For the K- and D- proton pathways we used a shorter distance of ~ 3 Å to analyse conformational changes which then would be possible to obtain from MD simulations. This procedure provided us with a list of water molecules that may form a long living hydrogen bonded network in COX corresponding to the static structure of the protein. For example, the pathway leading to Glu278 is shown in TABLE 3. The short distances (≤ 2.5 Å) between “inner” oxygens (representing a water molecule) in the D-pathway indicate that, if these atoms do not share a proton, they should form a hydrogen bonds either with protein residues or with each other. Possibly, these water molecules can show a large mobility during dynamic simulations, both forming and breaking hydrogen bonds. In contrast, the hydrogen bonded network in the K-pathway is fragmented and does not extend from the surface of the protein into a buried active site.

Several water molecules also have been found close to the heme propionates. This cluster was suggested to serve as a possible exit channel for water molecules released from the active site upon electron and proton transfer (Puustinen and Wikstrom, 1999).

The network analysis of the static structure may miss critical links between fragmented parts of the hydrogen bonded networks. Depending on the resolution of the protein structure, difficulties may arise in particular when the water molecules and side chains involved in proton transfer are relatively mobile or alternate site chain positions have low occupancies (≤ 0.3) (Taraphder and Hummer, 2003). To bridge the resulting gaps, we carried out an energy minimization and molecular dynamics simulations using the package CHARMM (Brooks et al., 1983) (see Chapter 4).

TABLE 3. GRID scan for water positions for the D- and K-pathways for cytochrome *c* oxidase from *Paracoccus denitrificans*.



CHAPTER 3.

Computer simulation methodology.

This chapter describes how the dynamic behaviour of a system can be simulated, how thermodynamic properties can be collected and evaluated, how simulations can be constrained to reproduce experimental conditions, and what time saving techniques exist to reduce the computational expense.

3.1 Statistical mechanics.

Let us consider the instantaneous mechanical states of a chemical system consisting of N interacting particles to be characterised by their positions in space and their momenta (assuming a classical approximation). That is, let us consider the set of $3N$ coordinates and $3N$ momenta (r^N, p^N) . Each state is a snapshot of the chemical system and is described as a location in phase space. An ensemble can be considered to be a collection of points in phase space, i.e. a collection of snapshots.

The value of a particularly interesting system property A at a point in phase space can be written as the function $A(p^N, r^N)$. As this system will evolve in real time, $A(p^N, r^N)$ will change due to the interactions between the particles. The value of A measured experimentally can be considered as an average of $A(p^N, r^N)$ over time, and is therefore referred to as a time average:

$$A_{obs} = \langle A \rangle_{time} = \lim_{t_{obs} \rightarrow \infty} \frac{1}{t_{obs}} \int_0^{t_{obs}} A(p^N(t), r^N(t)) dt \quad (3.1)$$

To calculate the average value of the properties of this system the simulation of its dynamic behaviour is required. This can be achieved (Leach, 1996) by describing an energy function that calculates the forces within the system due to interatomic interactions, and from

the force the acceleration can be determined. Integrating Newton's laws of motion then yields a trajectory which describes the positions, velocities and accelerations of the particles in the system over time, enabling the calculation of system properties directly using equation 3.1. This is exactly what is done in Molecular Dynamics (MD) simulations (Allen and Tildesley, 1987; Leach, 1996). A major problem is that in simulations the integral cannot be calculated to infinity. The value of A is then averaged over a sufficiently long time t_{MD} . In fact, the motion of the system is described on a step-by-step basis, i.e. a large finite number τ_{MD} of time steps of length $\Delta t = t_{MD} / \tau_{MD}$ are taken. In this case equation 3.1 may be rewritten in the form:

$$A_{MD} = \langle A \rangle_{time} = \frac{1}{\tau_{MD}} \sum_{\tau=1}^{\tau_{MD}} A(p^N(\tau), r^N(\tau)) \quad (3.2)$$

Because of the complexity of the time evolution of A , a different approach is to replace the single system evolving in time by a large number of replications that are considered simultaneously. The time average calculation in equation 3.1 can then be replaced by an ensemble average, where the average value of the property A is calculated over all the copies of the system:

$$A_{obs} = \langle A \rangle_{ensemble} = \iint dp^N dr^N A(p^N, r^N) \rho_{ens}(p^N, r^N) \quad (3.3)$$

The ergodic hypothesis is a fundamental axiom of statistical mechanics and assumes that the ensemble average is equal to the time average, allowing the transformation of equation 3.1 to 3.3 (Allen and Tildesley, 1987; Leach, 1996), with $\rho_{ens}(p^N, r^N)$ being the probability of finding a certain value of $A(p^N, r^N)$ in that ensemble. This is what is done in Monte Carlo (MC) simulations (Leach, 1996). A succession of a large finite number τ_{MC} of state points are generated to sample in accordance with the desired ρ_{ens} and the value of A is obtained as:

$$A_{MC} = \langle A \rangle_{ensemble} = \frac{1}{\tau_{MC}} \sum_{\tau=1}^{\tau_{MC}} A(p^N(\tau), r^N(\tau)) \quad (3.4)$$

In the equilibrium ensemble, $\rho(p^N, r^N)$ is not time-dependent, that is at each time instant the particles of the system are distributed among the different energy states according to the probability density: when one leaves a particular state, another arrives to replace it. Such a distribution of states is called a *Boltzmann distribution*. The probability density of each state i is the ratio between the number of particles in that state n_i and the total number of particles N in the system, and is calculated differently in different ensembles.

In the canonical ensemble (NVT), that is when the number of particles, the volume and the temperature of the system remain constant, this ratio is equal to:

$$\rho_{NVT}(p^N, r^N) = \frac{n_i}{N} = \frac{\exp(-H_i / k_B T)}{Q_{NVT}} \quad (3.5)$$

where H_i is the Hamiltonian of state i , k_B the Boltzmann constant, T the temperature, and Q_{NVT} is the so called partition function. The latter plays a fundamental role in statistical mechanics, and it is from this property that all thermodynamic functions can be derived. The molecular partition function of the canonical ensemble is defined as:

$$Q_{NVT} = \frac{1}{h^{3N} N!} \iint dp^N dr^N \exp(-H_i / k_B T) \quad (3.6)$$

where H_i is $H(p^N, r^N)$. Since H is always expressible as a sum of kinetic $K(p^N)$ and potential $V(r^N)$ contributions, then:

$$Q_{NVT} = \frac{1}{h^{3N} N!} \int dp^N \exp(-K_i(p^N) / k_B T) \int dr^N \exp(-V(r^N) / k_B T) = Q_{NVT}^{id} Q_{NVT}^{exc} \quad (3.7)$$

For an atomic system:

$$Q_{NVT}^{id} = \frac{V^N}{(h^2 / 2\pi m k_B T)^{3N/2} N!}, \quad Q_{NVT}^{exc} = \frac{1}{V^N} \int dr^N \exp(-V(r^N) / k_B T) \quad (3.8)$$

In Monte Carlo (MC) simulations it therefore proves possible to probe just the configurational part of phase space, with the ideal gas properties being added afterwards. Equation 3.4 can then be rewritten as:

$$A_{MC} = \frac{1}{\tau_{MC}} \sum_{\tau=1}^{\tau_{MC}} A(r^N(\tau)) \quad (3.9)$$

In the isothermal-isobaric ensemble (NPT), that is when the number of particles, the pressure and the temperature of the system remain constant, the probability density is equal to

$$\rho_{NPT}(p^N, r^N) = \frac{n_i}{N} = \frac{\exp(-(H_i + PV) / k_B T)}{Q_{NPT}} \quad (3.10)$$

where P and V are the pressure and the volume of the system, respectively. Then, the partition function is:

$$Q_{NPT} = \frac{1}{N!} \frac{1}{h^{3N}} \frac{1}{V_0} \iiint dp^N dr^N dV \exp(-(H_i + PV) / k_B T) \quad (3.11)$$

where V_0 is the unit of volume, required to render Q_{NPT} dimensionless. Again, it is possible to separate the configurational properties from the kinetic ones and to devise a Monte Carlo procedure to probe configurational space only.

3.2 Force-field and potential energy function.

To sample the positions r and momenta p of the N particles comprising the simulation system, a functional form for the calculation of the potential energy is required. This functional form is called the force-field and it assigns energetic penalties when the particles deviate from their equilibrium positions. In this approach, called molecular-mechanics (Leach, 1996), the molecular properties are considered to be dependent on the nuclei positions only and electron motion is ignored. A classical description is also employed and the Hamiltonian, H , of the simulation system is written as the sum of the particle kinetic K and potential V energies:

$$H(r, p) = K(p) + V(r) \quad (3.12)$$

In real systems the interaction of a particle with another is influenced by the presence of all the other particles, but in most computer simulations a pair wise approximation gives a good description of liquid, gas and solid phase properties.

In CHARMM (Brooks et al., 1983), one of the most common force fields, the potential energy $V(r^N)$ of a system of N particles as a function of the positions r^N is computed as the sum of the contributions from bond stretching, angle bending, dihedral torsions and non-bonded interactions (electrostatic and Van der Waals interactions):

$$V(r_N) = \sum V_{bonds} + \sum V_{angles} + \sum V_{dihedrals} + \sum V_{electrostatic} + \sum V_{Van-der-Waals} \quad (3.13)$$

The bond and angle contributions are described by harmonic potentials that give an increase in energy as the bond length b_i and the valence angle θ_i deviate from their reference values $b_{i,0}$ and $\theta_{i,0}$, by means of a force constant $k_{i,b}$ and $k_{i,\theta}$:

$$\sum V_{bonds} = \sum_{i=1}^{N_{bonds}} k_{i,b} (b_i - b_{i,0})^2 \quad (3.14)$$

$$\sum V_{angles-harmonic} = \sum_{i=1}^{N_{angles}} k_{i,\theta} (\theta_i - \theta_{i,0})^2 \quad (3.15)$$

The CHARMM force-field also includes a cross term to describe the angle terms. The angle bending, which involves the atoms 1, 2 and 3, is modelled using the Urey-Bradley potential, which considers the Van der Waals interactions between atoms 1 and 3 along their distance r_{1-3} :

$$\sum V_{angles-Urey-Bradley} = \sum_{i=1}^{N_{angles}} k_{i,UB} (r_{i,1-3} - r_{i,1-3}^0)^2 \quad (3.16)$$

Here $r_{i,1-3}^0$ is the reference distance and $k_{i,UB}$ the force constant. The torsional terms are computed as:

$$\sum V_{dihedrals} = \sum_{i=1}^{N_{dihedrals}} k_{i,\phi} [1 + \cos(n_i \phi_i - \delta_i)] \quad (3.17)$$

Here ϕ_i is the dihedral angle, $k_{i,\phi}$ the force constant, n_i is the multiplicity (which gives the number of minimum points in the function as the torsion changes from 0 to 360 degrees) and δ_i is the phase angle (which determines the location of the minimum values).

The non-bonded interactions are calculated between all pairs of atoms i and j that are in different molecules or in the same molecule but separated by at least three bonds (1,4 interactions). The electrostatic interactions are usually modelled using a Coulombic potential term:

$$\sum V_{electrostatic} = \sum^{pairs(i,j)} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (3.18)$$

where q_i and q_j are the charges on atoms i and atom j respectively, r_{ij} their separation distance, and ϵ_0 the relative permittivity of free space. The Van der Waals interactions are modelled using a Lennard-Jones (LJ) potential:

$$\sum V_{Van-der-Waals} = \sum^{pairs(ij)} \epsilon \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right] \quad (3.19)$$

where ϵ is the well depth at the energy minimum and $R_{min,ij}$ is the i - j distance at the energy minimum. In common force fields, these parameters are derived for pairs of identical atoms (Leach, 1996), for instance ii and jj . When interactions between different types of atoms are required, it is common to apply mixing rules, e.g. in CHARMM:

$$R_{min,ij} = \frac{R_{min,ii}}{2} + \frac{R_{min,jj}}{2}, \quad \epsilon_{ij} = (\epsilon_{ii} \epsilon_{jj})^{1/2} \quad (3.20)$$

The i - j distance at which the LJ potential energy is zero is usually referred to as σ_{ij} and is equal to $2^{-1/6} R_{min,ij}$.

Out of plane bending, such as that in carbonyl groups, is incorporated by relating the atoms involved using an ‘‘improper’’ torsion angle ψ_i and incorporating an additional energy term in the force-field. In the CHARMM force-field these ‘‘improper’’ torsions are described using a harmonic potential:

$$\sum V_{improper} = \sum_{i=1}^{N_{improper}} k_{i,\psi} (\psi_i - \psi_{i,0})^2 \quad (3.21)$$

Force-fields are empirically determined, that is, there are no rules that allow one to define *a priori* equilibrium values, force constants, charges and Van der Waals parameters, but rather several values are employed in an iterative fashion until the simulated system has a static and dynamic behaviour as close as possible to that of the real system. At the beginning of a force-field parametrization, equilibrium values, force constants, charges and Van der Waals parameters are usually obtained from both quantum-mechanical calculations and spectroscopic data, such as X-ray diffraction, IR and NMR. They are then modified and refined to reproduce, during a simulation, available experimental data for the system of interest (Leach, 1996).

Force-fields describe in a classical formalism the potential energy of a system of any kind of particle. In computational chemistry each atom may be explicitly represented in the model, but also a group of two or more atoms may be modelled as a single “pseudo-atom”. In the first case, the force-field is called “all-atom”, while in the second case the force-field is called “united-atom”. Usually, just the hydrogen atoms bonded to non-polar atoms such as carbon are modelled as united-atoms, while hydrogens bonded to polar atoms such as nitrogen or oxygen that are able to participate in hydrogen bonding interactions are modelled explicitly. With united-atom force-fields a considerable computational saving is possible, since the number of particles simulated is reduced and consequently the number of interactions to be calculated is smaller. However, united-atom force-fields have some drawbacks. With this kind of representation the level of detail and accuracy is lower (Leach, 1996). Moreover, chiral centers may be able to invert during the simulation because of the absence of an explicit hydrogen atom on the chiral carbon, and so an additional improper torsion must be added into the force-field to prevent the inversion (Leach, 1996).

3.3 Energy minimization methods.

Energy minimization methods are typically applied to simulation systems prior to modelling studies, but are also useful for insights into molecular structure and, occasionally, sufficient to predict the properties of a system accurately (Leach, 1996). The procedure of minimization optimises the atomic positions subject to the forces generated by the molecular force field. Minimization algorithms are designed to move atoms in such a way as to reduce the energy of the molecule until a minimum value is reached or an agreed energy convergence criterion has been met. Two common minimization algorithms are the steepest descent and the conjugate gradient methods (Leach, 1996).

The steepest descent method makes moves parallel to the direction of the force on a particle, so a move is taken directly downhill on the potential energy surface, with either a step of arbitrary size, or, from the result of a line search, to the minimum energy position along the new direction. Then the process is repeated without using any information from the previous move. Conversely, in the conjugate gradient method the previous step directions are used to refine the current direction of move.

More efficient minimization algorithms include second derivative information and perform better along flat areas of the potential energy surface. The Newton-Raphson algorithm is an example (Leach, 1996).

3.4 Equations of motion.

For any arrangement of the particles in the simulation system the force \mathbf{F}_i acting on each particle i due to its interactions with the other particles can be calculated by taking the first derivative of the energy function V_i with respect to the particle positions:

$$\mathbf{F}_i = -\nabla V_i \quad (3.22)$$

These forces allow the system to change by collective motions of particles over time in a way described by Newton's second law of motion:

$$\mathbf{F}_i = m_i \mathbf{a}_i \quad (3.23),$$

where m_i is the particle mass and \mathbf{a}_i the particle acceleration. Since it is impossible to solve the complex coupled second order differential equations required for all the particles in the system, solutions are calculated using a numerical treatment and evaluating the particle displacements over small increments of time. From position $\mathbf{r}(t)$, the force $\mathbf{F}(t)$ acting at time t is calculated for all the pairs of particles of the system, and from the force $\mathbf{F}(t)$ the acceleration $\mathbf{a}(t)$ is computed. By knowing the velocity $\mathbf{v}(t)$ and the acceleration $\mathbf{a}(t)$, the new position $\mathbf{r}(t + \Delta t)$ and the new velocity $\mathbf{v}(t + \Delta t)$ are calculated after a time interval, Δt , which is called the time step. The force is assumed to be constant during Δt , and then the calculation is repeated again. Initial velocities are obtained from Gaussian, Maxwell-Boltzmann or uniform distributions which allow the system to be at the desired temperature. There are several algorithms for integrating the equations of motion and they all assume that the positions, velocities and accelerations of the system can be approximated as Taylor series expansions. An example of one such algorithm is the leap-frog algorithm (Hockney, 1970), which solves the expansion in the following way:

$$r(t + \Delta t) = r(t) + \Delta t v\left(t + \frac{1}{2}\Delta t\right) \quad (3.24)$$

$$v\left(t + \frac{1}{2}\Delta t\right) = v\left(t - \frac{1}{2}\Delta t\right) + \Delta t a(t) \quad (3.25)$$

$$v(t) = \frac{1}{2}\left[v\left(t + \frac{1}{2}\Delta t\right) + v\left(t - \frac{1}{2}\Delta t\right)\right] \quad (3.26)$$

The choice of time step depends on the time scale of the motions in the system: it should be one order of magnitude smaller than the shortest (i. e. fastest) motion (Leach, 1996). In the case of an atomic simulation, where the simulated particles are atoms, this motion is represented by the bond stretching and so a time step of 0.5 fs is required.

Before a simulation starts to produce useful data, the system usually needs to be minimized. This is required since the initial structure, which may be derived either from X-ray crystallography or quantum-mechanical calculations or both, generally does not correspond to the most populated structure under the simulation condition of temperature, pressure and solvation. Moreover, bad contacts may occur very easily. Minimization algorithms find energy arrangements of the particles which correspond to stable low-energy states of the system.

After minimization, and before starting a production simulation, the system is allowed to undergo a period of equilibration. That is a period after which all memory of the initial configuration is lost and after which thermodynamic properties have ceased to show a systematic drift and have started to oscillate about steady mean values.

3.5 Application of constraints in simulations.

The MD procedure described so far samples the NVE ensemble, where the number of particles N , the volume V and the energy E of the system are constant and the temperature T and the pressure P are allowed to fluctuate. Therefore the Hamiltonian in equation 3.12 can be referred to as H_{NVE} . This is not always a useful ensemble; it would be nice to study systems of interest under conditions of constant T and P , that represent the simulation found in an experimental laboratory. Algorithms are available to constrain these degrees of freedom and NVT and NPT ensembles can be simulated.

3.5.1 Constant temperature.

In the classical molecular mechanics model, temperature is directly related to the kinetic energy of the system (Leach, 1996):

$$K = \sum_{i=1}^{N_{\text{particles}}} \frac{p_i^2}{2m_i} = \frac{k_B T}{2} (3N - N_C) \quad (3.27)$$

Here K is the kinetic energy, p_i the momentum of particle i , m_i its mass, k_B the Boltzmann constant, T the current kinetic temperature, N the number of particles simulated in the system and N_C the number of constraints. The velocities are generated during the simulation and hence the momenta and the temperature can be calculated.

One of the most rigorous algorithms to constraint the temperature is the Nosé-Hoover thermostat (Nosé and Klein, 1983; Nosé, 1984). A thermal reservoir is considered to be part of the system and it is represented as an additional degree of freedom s . The Hamiltonian H of the extended system is equal to the sum of the kinetic K and potential V energies of the particles with the kinetic K_S and potential V_S energies of the reservoir:

$$H_{NVT} = K + V + K_S + V_S = \sum_{i=1}^N \frac{p_i^2}{2m_i s^2} + V(r'^N) + \frac{p_s'^2}{2M} + (n+1)k_B T_0 \ln s \quad (3.28)$$

Here T_0 is the desired temperature, n is the number of degrees of freedom, p_s is the momentum conjugate to s , and M is the thermal inertia parameter, which can be considered the mass of extra degree of freedom. The prime indicates “extended-system” variables or virtual variables, which are related to the real variable of the physical system by:

$$\text{coordinates: } r' = r, \text{ momenta: } p' = p/s, \text{ time } t' = \int_0^t dt/s \quad (3.29)$$

The equation of motion for the real variables then modifies to:

$$\frac{d^2 r_i}{dt^2} = -\frac{1}{m_i} \frac{dV(r^N)}{dr} - \frac{sp_s}{M} \frac{dr_i}{dt} \quad (3.30)$$

The factor sp_s/M , also indicated by ξ , acts as a thermodynamic friction coefficient and is controlled by:

$$\frac{d(sp_s/M)}{dt} = \frac{1}{M} \left(\sum_{i=1}^N \frac{p_i^2}{m_i} - nk_B T_0 \right) \quad (3.31)$$

The choice of M is arbitrary but is critical to the success of a run. Too high a value results in slow energy flow between the system and the reservoir and the thermalization process is not efficient. On the other hand, if M is too low, strong energy oscillations occur, resulting in equilibration problems.

3.5.2 Constant pressure.

In simulations the pressure is related to the virial W (Allen and Tildesley, 1987). The total virial function has an internal contribution, related to the forces acting between the simulated particles, and an external contribution, related to externally applied forces:

$$W_{\text{internal}} = \frac{1}{3} \left\langle \sum_{i=1}^N x_i \cdot F_{\text{internal}} \right\rangle = -\frac{1}{3} \left\langle \sum_{i=1}^N x_i \cdot \nabla V_i \right\rangle \quad (3.32)$$

$$W_{\text{external}} = \frac{1}{3} \left\langle \sum_{i=1}^N x_i \cdot F_{\text{external}} \right\rangle = -PV \quad (3.33)$$

$$W_{\text{total}} = \frac{1}{3} \left\langle \sum_{i=1}^N x_i \cdot F_{\text{total}} \right\rangle = -Nk_B T \quad (3.34)$$

Here x_i is the position of the particle i , F_{internal} is the force acting on that particle according to the force-field and due to the interactions with the other particles of the system, F_{external} is an externally applied force acting on i , which is related to the external pressure, and F_{total} is the sum of intermolecular and external forces. The pressure P is consequently easily obtained in the following way:

$$PV = Nk_B T + \frac{1}{3} \left\langle \sum_{i=1}^N x_i \cdot F_{\text{internal}} \right\rangle \quad (3.35)$$

One of the most rigorous algorithms ensuring the simulation pressure is constant is Andersen's extended system method (Andersen, 1980). This method is based on including an extra degree of freedom mimicking the action of a piston and corresponding to the volume V of the simulation box which adjusts itself to equalize the internal and the applied pressures, respectively P_{int} and P_{ext} . The Hamiltonian of the extended system is:

$$H_{NPT} = K + V + K_\eta + V_\eta = \sum_{i=1}^N \frac{p_i'^2}{2m_i V^{2/3}} + V \left(V^{1/3} r'^N \right) \cdot \frac{p_\eta^2}{2M} + P_0 V \quad (3.36)$$

Here M is the fictitious mass of the piston (with units mass length⁻⁴), and p_η the momentum associated with V . The extended system variables are indicated with primes and relate to the real system variables by:

$$\text{coordinates: } r' = r / V^{1/3}, \quad \text{momenta: } p' = V^{1/3} p \quad (3.37)$$

$$\frac{d^2 r(t)}{dt^2} = -\frac{1}{m_i} \frac{dV(r^N)}{dr} - \frac{1}{3} \frac{dV(t)}{dt} \frac{1}{V(t)} \frac{dr(t)}{dt} \quad (3.38)$$

Here the factor $\frac{1}{3} \frac{dV(t)}{dt^2} \frac{1}{V(t)}$ is the barostat friction coefficient, and it is controlled by

(Andersen, 1980; Feller et al., 1995):

$$\frac{d^2V(t)}{dt^2} = \frac{P_{\text{int}}(t) - P_{\text{ext}}}{M} \quad (3.39)$$

Here $P_{\text{int}}(t)$ is the instantaneous internal pressure (calculated with equation 3.38) and P_{ext} the applied (desired) external pressure.

Nosé and Klein extended this method to the case of non-cubic simulation cells (Nosé and Klein, 1983):

$$H_{NPT} = H + V + H_{\eta} + V_{\eta} = \sum_{i=1}^N \frac{p_i'^2}{2m_i h' h} + V(h^{-1} r^N) \frac{p_{\eta}^2}{2M} + P_0 V \quad (3.40)$$

Here h is a matrix with three columns representing the sizes of the three cell sides, h' is its transpose, and:

$$\text{coordinates: } r' = r/h, \quad \text{momenta } p' = hp \quad (3.41)$$

The equation of motion for the real system variables thus becomes:

$$\frac{d^2 r(t)}{dt^2} = -\frac{1}{m_i} \frac{dV(r^N)}{dr} - \frac{dh(t)}{dt} \frac{1}{h(t)} \frac{dr(t)}{dt} \quad (3.42)$$

The factor $\frac{dh(t)}{dt} \frac{1}{h(t)}$ is controlled by:

$$\frac{d^2 h(t)}{dt^2} = \frac{P_{\text{int}}(t) - P_{\text{ext}}}{MVh'} \quad (3.43)$$

In this case M has the unit of a true mass.

Andersen's method has been shown to produce strong oscillations on the volume of the system. It has been improved by Feller (Feller et al., 1995), who introduced the Langevin piston method. In this case:

$$\frac{d^2V(t)}{dt^2} = \frac{P_{\text{int}}(t) - P_{\text{ext}}}{M} - \gamma \frac{dV(t)}{dt} + R(t) \quad (3.44)$$

where $R(t)$ is a random number taken from a Gaussian distribution and γ is the collision frequency. This algorithm was also implemented in the Nosé and Klein version for non-cubic cells and has been provided to minimize the effect that the choice of piston mass has on the dynamics.

3.6 Special techniques.

It has been already mentioned that the size of the time step depends on the fastest motion of the system, and in the case of atomic simulations this motion corresponds to bond stretching (Leach, 1996). Using a larger time step would speed up the calculation, since with the number of steps a greater system evolution in time would be sampled. A solution is to freeze out some vibrations, for instance bonds, by constraining them to their equilibrium value, while the rest of the degrees of freedom are allowed to vary. The normal forces \mathbf{f}_i from inter- and intra-molecular interactions defined in the force-field act on all the particles of the system, while some constraining forces \mathbf{g}_i act on the selected constrained degrees of freedom only. Consequently, the equation of motion for the constrained degrees of freedom, for instance bond lengths, can be rewritten as: $m_i \mathbf{a}_i = \mathbf{f}_i + \mathbf{g}_i$. Within the SHAKE (Ryckaert et al., 1977) algorithm, the technique is to solve the equation of motion for one time step in the absence of the constraint forces and subsequently determine their magnitude by correcting the atomic positions to ensure that the difference between the square of the new atomic distance and the square of the old atomic distance is zero at all times, within a chosen tolerance, for all the bond lengths selected to be constrained. Each constraint is considered in turn and solved. This may violate another constraint, and so it is necessary to iterate until all the constraints are satisfied.

To reduce the number of particles in the simulation system, periodic boundary conditions are employed; they are also used to overcome the problem of surface effects. The simulation box is replicated throughout space in all directions in an infinite lattice and consequently is completely surrounded by its image boxes. There are no empty spaces between the simulation control box and its images, where the first terminates the image starts, and for this reason the particles experience forces as if they were in bulk fluid. If a particle leaves the box, it is replaced by its image particle entering from the opposite side coming from an image box, so that the number of particles in the system is always constant.

The most time consuming part of a MD simulation is the calculation of the non-bonded energies and forces (electrostatic and Van der Waals). In principle the non-bonded interactions should be computed between every pair of particles in the system including the images, but this would be too time consuming, inefficient and arguably unnecessary. For this reason the calculation can be approximated using a cutoff. It is assumed that the largest contribution to the potential and forces comes from the near neighbours and a truncation can be made: all the interactions between pairs of particles further than a spherical cutoff value are set to zero. The choice of a correct cutoff radius is important. For Van der Waals interactions,

the potential becomes insignificant after a distance of about 2.5σ in the Lennard-Jones potential. This means Van der Waals interactions are short range interactions and in the case of atomistic simulations they can be truncated after 10 \AA . In contrast, electrostatic interactions are long range interactions and their effects are felt at a considerably greater distance than the cutoff commonly used for simulations. To enforce the cutoff method without losing accuracy, several techniques are available. First, two different cutoff radii can be used, a shorter one for short range Van der Waals interactions and a longer one for long range electrostatic interactions (Leach, 1996). Otherwise, more complicated methods such as the Ewald sum can be employed (Ewald, 1921) for calculating the full electrostatic energy of a unit cell. According to the Ewald sum method, each point charge present in the system is surrounded by a charge distribution of equal magnitude and opposite sign; this distribution is conveniently taken to be Gaussian. This extra distribution acts like an ionic atmosphere to screen the interactions between neighbouring charges. The screened interactions are now short-ranged and the total screened potential is calculated by summing over all the particles in the central box and all the images in real space. Then, another charge distribution of opposite sign to the first added Gaussian distribution is also added and this cancelling distribution is summed in reciprocal space. A fast implementation of the Ewald summation is the Particle Mesh Ewald (PME) method (Darden et al., 1993). In this case the reciprocal space term of the Ewald sum is approximated using Fast Fourier Transformations with convolutions on a grid where charges are interpolated to the grid points. This procedure greatly reduces the computational cost of the reciprocal space sum.

To reduce the time for the computation, a particle's neighbours list is created (Verlet, 1967). For a given particle, it contains a list of the particles within a distance slightly larger than the cutoff. To speed up the simulation the neighbour list is updated at given intervals and not at each simulation step. Between updates the program does not check through all the particles, but it calculates the distance between the particle of interest and only those particles appearing in the list, and consequently the time for the non-bonded calculation is significantly reduced.

Another problem of the cutoff method is the discontinuity in potential energy at the cutoff distance: since at this distance the potential suddenly drops from its real value to zero, at the cutoff distance interaction energies and forces fluctuate violently and this causes problems of energy and temperature conservation. A common solution consists in switching the potential between two cutoff distances (Leach, 1996): the potential has its true value until the lower cutoff and then it is multiplied by a switching function, whose values range between

1 at a distance equal to the lower cutoff and 0 at a distance equal to the upper cutoff. Force and potential are thereby smoothly reduced to zero.

3.7 Interfaces.

When simulating lipid/solvent interfaces, only the component of the pressure tensor normal to the interface P_N gives a measure of the bulk pressure of the system, while the lateral components depend on the interfacial tension. Thus, in contrast to homogeneous fluid where only one pair of thermodynamic variables (pressure P and volume V) determine the size of the simulation cell, two sets of variables determine the size and shape of the interfacial system: P_N and V , together with the surface tension γ and the surface area A (Zhang et al., 1995; Feller et al., 1995). By extension of the Andersen constant pressure method and with the inclusion of a constant temperature algorithm, two ensembles can be generated: 1) constant normal pressure and surface area (NP_NAT), and 2) constant normal pressure and surface tension ($NP_N\gamma T$).

We fix the z axis of the simulation box along the normal to the interface, and the x and y axis define the interfacial plane. In this case, the box length along z is L_z and the normal pressure is the pressure tensor along z , i.e. $P_N = P_{zz}$. At each z position along the interface normal, the lateral pressure is the average between the pressure tensors along x and y , and the average lateral pressure along the z axis is (Zhang et al., 1995):

$$P_L = \frac{1}{L_z} \int_0^{L_z} P_L(z) dz = \frac{1}{2L_z} \int_0^{L_z} [P_{xx}(z) + P_{yy}(z)] dz \quad (3.45)$$

The value of the surface tension, which is the observed γ in experiments, is obtained as:

$$\gamma = \int_{-\infty}^{+\infty} [P_N - P_L(z)] dz = L_z (P_N - P_L) \quad (3.46)$$

Only the interface contributes to the integral, since in bulk phase $P_N = P_L(z)$.

3.7.1 Constant normal pressure and surface area.

In the constant pressure and constant surface area ensemble, L_x and L_y are fixed to maintain the constant surface area $A = L_x L_y$, while L_x fluctuates to adjust the internal normal pressure of the system P_N to the applied P_N^0 . The extended system Hamiltonian for modelling in this ensemble is in practice a one dimensional Andersen's pressure model (Zhang et al., 1995). Only the equation of motion along the z axis is modified:

$$\frac{d^2 z(t)}{dt^2} = \frac{f_z(t)}{m_i} - \frac{dL_z(t)}{dt} \frac{1}{L_z(t)} \frac{dz(t)}{dt} \quad (3.47)$$

with $L_z(t)$ controlled by:

$$\frac{d^2 L_z(t)}{dt^2} = L_x L_y \frac{P_N(t) - P_N^0}{M} \quad (3.48)$$

3.7.2 Constant normal pressure and surface tension.

In the constant normal pressure and constant surface tension ensemble L_x , L_y and L_z can all vary. In what follows, P_N^0 is the reference normal pressure and γ^0 the reference surface tension. The equation of motion modifies to (Zhang et al., 1995):

$$\begin{aligned} \frac{d^2 x(t)}{dt^2} &= \frac{f_x(t)}{m_i} - \frac{dL_x(t)}{dt} \frac{1}{L_x(t)} \frac{dx(t)}{dt} \\ \frac{d^2 y(t)}{dt^2} &= \frac{f_y(t)}{m_i} - \frac{dL_y(t)}{dt} \frac{1}{L_y(t)} \frac{dy(t)}{dt} \\ \frac{d^2 z(t)}{dt^2} &= \frac{f_z(t)}{m_i} - \frac{dL_z(t)}{dt} \frac{1}{L_z(t)} \frac{dz(t)}{dt} \end{aligned} \quad (3.49)$$

Here:

$$\begin{aligned} \frac{d^2 L_x(t)}{dt^2} &= L_y \frac{\gamma^0 - L_z (P_N^0 - P_{xx}(t))}{M_x} \\ \frac{d^2 L_y(t)}{dt^2} &= L_x \frac{\gamma^0 - L_z (P_N^0 - P_{yy}(t))}{M_y} \\ \frac{d^2 L_z(t)}{dt^2} &= L_x L_y \frac{P_N(t) - P_N^0}{M_z} \end{aligned} \quad (3.50)$$

CHAPTER 4.

Dynamic water networks in cytochrome *c* oxidase from *Paracoccus denitrificans* investigated by molecular dynamics simulations.

This chapter describes and discusses the approach used to identify and characterize hydrogen bonded networks inside COX as possible pathways for proton transport by predicting water binding sites and then characterizing the hydrogen bonded networks during molecular dynamics simulations and its results.

4.1 Overview of the molecular dynamics simulations of membrane proteins.

The determination of the three-dimensional structures of cytochrome *c* oxidase of both mammalian and bacterial forms presents an opportunity for understanding the enzyme's function at the molecular level. Since certain dynamic features of cytochrome *c* oxidase cannot be captured by crystallographic techniques, MD has been used to elucidate conformational fluctuations and COX – water mobility (Hofacker and Schulten, 1998; Backgren et al., 2000; Zheng et al., 2003; Wikström et al., 2003). There are also some examples of studies where different redox states of protein systems have been analyzed using molecular dynamics (MD) procedures (Hayashi et al., 2002; Bret et al., 2002; Wikström et al., 2003). Such studies have shown that molecular dynamics simulations can provide the details

concerning individual particle motions as a function of time. Thus, MD can be used to address specific questions about the properties of a model system, often more easily than experiments on the actual system. For example, by what pathways do protons access the binuclear center?

This chapter describes a molecular dynamics study of cytochrome *c* oxidase from *Paracoccus denitrificans* in the fully oxidized state, embedded in a fully hydrated dimyristoylphosphatidylcholine lipid bilayer membrane. Two parallel simulations with different levels of protein hydration, 1.125 ns each in length, were carried out under conditions of constant temperature and pressure using three-dimensional periodic boundary conditions and full electrostatics to investigate the distribution and dynamics of water molecules and their corresponding hydrogen-bonded networks inside cytochrome *c* oxidase.

Fully atomic simulations of membrane proteins must include a lipid bilayer in order to model the natural environment. Given the irregular shape of membrane proteins, obtaining a correctly configured initial system is a non-trivial task, and yet the reliability of the subsequent simulation may depend on how carefully this is performed. In order to build these protein-lipid bilayer systems, two approaches have been reported in the literature. The first (Petrache et al, 2000; Woolf and Roux, 1994; Woolf and Roux, 1996) consists of building a bilayer around the protein lipid by lipid, each individual molecule being selected from a library of lipid conformers. The second approach (Shen et al., 1997; Tieleman and Berendsen, 1998) uses a previously equilibrated lipid bilayer, in which a cylindrical hole to accommodate the protein is created by the application of weak repulsive radial forces of the lipid atoms. In both cases the protein-lipid system is energy minimized prior to the MD simulation.

Molecular dynamics simulations have become a standard tool for the investigation of biomolecules. The approach has been summarized in two classic texts: *Computer Simulation of Liquids* by Allen and Tildesley (1987) and *Proteins: A theoretical perspective of dynamics, structure, and thermodynamics* by Brooks et al. (1988). The concept of MD dates from the early days of computers and the realization that thermodynamic quantities related to the liquid state could be calculated with a sufficiently large statistical mechanical ensemble of conformations. The earliest calculations used Van der Waals spheres to describe deviations from ideal gas behavior. The development of the methods led to the realization that proteins could be studied with the same tools. That is, the thermodynamic quantities related to the room temperature fluctuations of protein motions could be calculated on a computer. The first paper to show that protein motions could be calculated was from McCammon et al. (1977). The history of the application of MD computer calculations to proteins dates from this point.

Today we can study much more complicated systems, such as fully solvated membrane protein complexes.

Biomolecular dynamics simulations find three major areas of application today. Firstly, MD simulation is used to bring biomolecular structures alive, giving insights into the natural dynamics on different timescales of biomolecules in solution. Secondly, MD simulation affords thermal averages of molecular properties. According to the ergodic hypothesis, one can simulate a single molecule with its surroundings for a period of time and get time-averaged molecular properties that approach the experimentally measurable ensemble averages. This is used to calculate, for example, the bulk properties of fluids and the free energy differences for chemical processes such as ligand binding. Thirdly, MD can explore which conformations of a molecule or a complex are thermally accessible. This technique is used for exploring conformational space, for instance, in ligand-docking applications. Furthermore, if data from experiments are translated into restraining potentials that guide the dynamics calculations, MD can conveniently combine these experimental data with the knowledge about the general properties of molecular structure that is embodied in the hundreds of parameters of a molecular mechanics force field.

The first protein simulation (McCammon et al., 1977) 25 years ago dealt with a protein molecule in a vacuum. Here, clear progress has been made in the form of inclusion of a realistic description of the surrounding solvent. Four features of this development are the inclusion of explicit solvent molecules around the protein, the addition of counterions, a more realistic treatment of the system boundaries and a more accurate treatment of long-range electrostatic forces. Each of these points offers several alternative implementation options.

The simulation temperature can be regulated in several ways, for instance, with a simple Berendsen scaling of velocities (Berendsen et al., 1984). The use of the Nosé-Hoover approach (Nosé, 1984; Hoover, 1985) is sometimes recommended because it is said to yield a perfect canonical distribution. Tuckerman et al. (2001) have explained how the inner derivative arising from applying the chain rule for differentiation – the phase space compressibility – has been neglected in the original Nosé-Hoover argument.

With increasing computational power, larger and more complex systems have become accessible to MD simulations. Marrink et al. (2000) studied the spontaneous formation of dodecylphosphocholine surfactant micelles during simulations up to 50 ns. Their systems contain up to 54 surfactant molecules solvated in more than 20 000 water molecules. The next step towards the simulation of membrane-bound proteins was described by Tieleman and Sansom (2001), who discussed different aspects of simulating a peptide in a lipid bilayer. Gao

and Wong (2001) showed that an adrenocorticotropin peptide in a solvated micelle moves to a position parallel to the micellar surface, regardless of whether it starts inside the micelle or at the surface. Membrane simulations can model yet another aspect of natural membranes by the inclusion of sterol molecules, as was done by Smondyrev and Berkowitz (2001) using a dimyristoylphosphatidylcholine bilayer containing cholesterol, ergosterol or lanosterol.

The availability of X-ray structures of integral membrane proteins opens the way to MD simulations of such systems as well. For instance, three monomeric bacteriorhodopsin molecules, 28 lipid molecules and about 2 800 water molecules under periodic boundary conditions (PBC) were simulated for 1 ns to predict the movement of water molecules during the photocycle of the protein (Baudry et al., 2000). Structural and dynamic properties of the potassium channel were reported by Allen et al. (2000), who mimicked the lipid bilayer using constraints in the outer region of the protein. The same channel was studied by Åqvist and Luzkov (2000), in a sphere of 3.3 nm containing water molecules and hydrocarbon-like atoms representing the membrane. They calculated free energy differences between different possible ion configurations within the channel to elucidate treatment of the same channel inserted in a phospholipid bilayer in aqueous solution, using PME for a long-range electrostatic interactions. Berneche and Roux (2001) expanded their studies of the KcsA potassium channel with a set of rigorous potential of mean force calculations, arriving at conclusions similar to those of Åqvist and Luzkov.

Another explicit treatment was performed by Elmore and Dougherty (2001), who did three simulations (>1 ns) of the *Mycobacterium tuberculosis* homologue of the bacterial mechanosensitive channel of large conductance. Their very large systems consisted of a 495-residue protein immersed in 290 palmitoyl-oleoyl-phosphatidylethanolamine molecules and 17 851 water molecules, leading to a total of 73 313 atoms.

Simulations of a very large system, consisting of the tetrameric human water channel aquaporin-1 and 271 phospholipid molecules fully solvated in water, with a total system size of 87 644 atoms, as well as simulations of a similar 101 448 atom system involving the related bacterial glycerol facilitator, have been reported (De Groot and Grubmüller, 2001). Multiple events of diffusional permeation of water molecules through the pores were observed on a simulation timescale (10 ns).

4.2 Previous computational work on cytochrome *c* oxidase.

Hofacker and Schulten (Hofacker and Schulten, 1998) studied the pathways for transport of oxygen and protons by molecular dynamics (MD) simulations of oxygen diffusion as well as key proton transfer steps. A large number of water molecules were predicted within the protein using the DOWSER method. The modeled water molecules indicate two possible pathways for taking up protons from the cytoplasmic side of cytochrome *c* oxidase, in good agreement with experimental data. The shorter one, the K-pathway, leads to the binuclear center via the highly conserved residues Lys354, Thr351 and Tyr280 and the hydroxyl group of the heme a_3 hydroxyethylfarnesyl chain. In the X-ray structure, Lys354 is close to Ser291 and most likely unprotonated. The channel, consequently, exhibits a gap between Lys354 and Thr351 which is not bridged by water molecules. Molecular dynamics simulations of the *Paracoccus denitrificans* structure revealed that Lys354 is flexible enough to bridge the gap by movement of its $C_\beta - N_\epsilon$ chain. Once protonated, the lysine head group moved within 50 ps by 3.1 Å to a position close to Thr351. The second, longer pathway (D-pathway), is formed by a chain of water molecules leading from Asp124 on the inside of the membrane to Glu278. In this simulations, Glu278 was unprotonated, which is not in agreement with experimental observation; once protonated the side group forms a hydrogen bond to the backbone oxygen of Met99 while the hydrogen bonds to water molecules are weakened or broken. Although the placement of water molecules identifies an efficient pathway for protons up to Glu278, there is no obvious continuation of the pathway from there. One possibility is that the protonated glutamic acid side chain could flip upward and deliver its proton either to the binuclear center or to a heme a_3 propionate group. But within the short time span accessible to the MD simulations, such a conformational change of Glu278 was not observed. Another problem is that Hofacker and Schulten used an electrostatic cutoff at 12 Å which is not compatible with the results of electrostatic calculations (Kannt et al., 1998). These indicated the existence of electrostatic effects on amino acids chains at large distances.

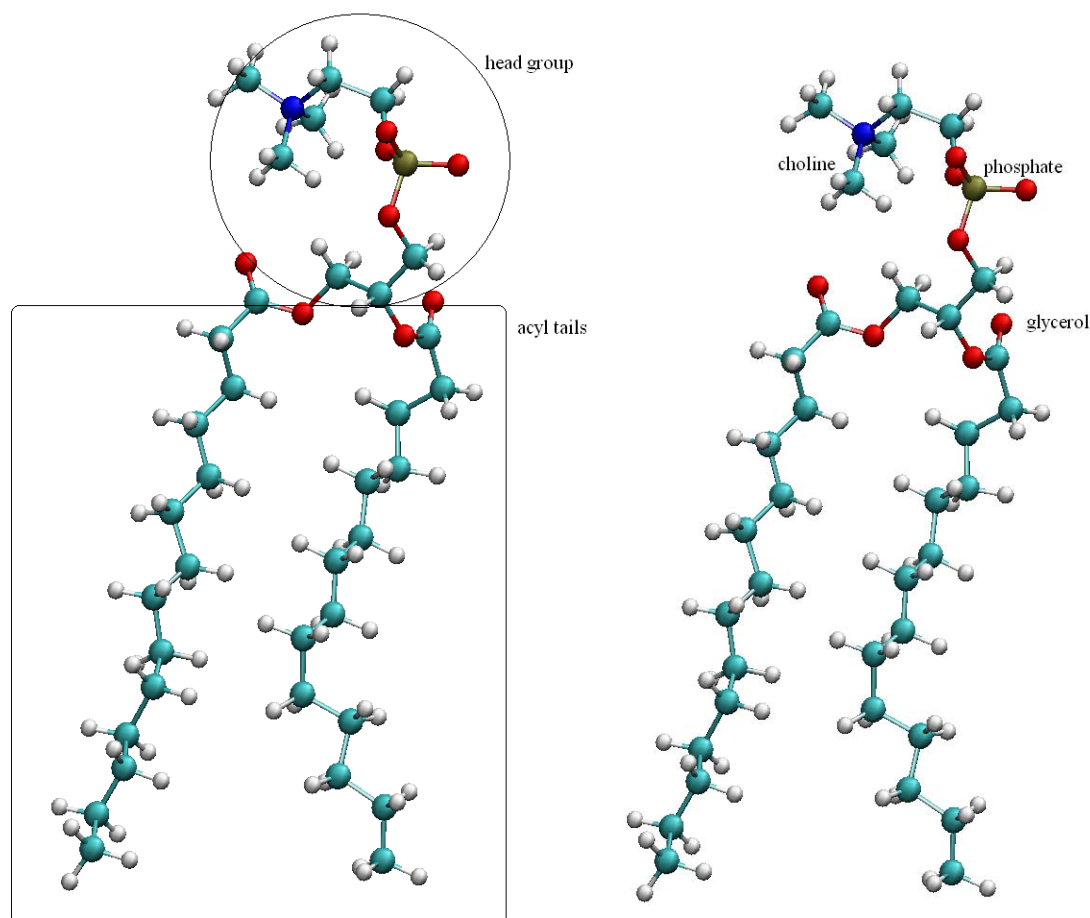


FIGURE 10. All atom model of a DMPC lipid. Graphical representations were made using program VMD.

4.3 Details of the simulation protocol.

4.3.1 Initial set up of protein-membrane-water system.

As described in the following, a proper membrane environment for the cytochrome *c* oxidase was provided by constructing a DMPC (dimyristoylphosphatidylcholine) lipid bilayer. DMPC is ubiquitous in cellular systems with fourteen carbons on each tail and a choline head (FIGURE 10). DMPC is an obvious choice as a model system for MD simulations because it is one of the simplest lipid components found in cellular membranes. In addition, the pure lipid is in the physiologically relevant liquid crystalline phase at room temperature and it has been well studied experimentally (see the review of Douliez et al., 1996). We followed the general protocol for molecular dynamics (MD) simulations of the membrane proteins (Berneche et al., 1998; Berneche and Roux, 2000) to construct the initial configuration of a protein-membrane-water system. Two ingredients were needed for our method to achieve success. The first was the choice of the appropriate cell dimensions for the system that we wanted to simulate. This required a trade-off between the ideals of a very large

system that is run for a very long time and the realities of a limited amount of CPU time that scales with the square of the number of atoms. We elected to choose a unit cell that has approximately two layers of lipids surrounding the central protein. This is not an infinitely dilute system where the lipid will reach a bulk (lipid) limit, but at more than 100,000 atoms total, this reflected the largest system that we felt could be built and run with reasonable efficiency. Furthermore, there are arguments that this type of system is that studied experimentally under the concentrations of protein and lipid usually used. A second decision concerned the appropriate algorithm for the dynamics. Ideally, a constant normal pressure and constant surface tension approach (see CHAPTER 3 for details) would be used that adjusts the cell dimensions throughout the simulation to match experimental values. Unfortunately, current molecular dynamics practice has not converged on appropriate methods for defining the appropriate value for the surface tension in such a simulation. We thus elected to perform the calculation using constant normal pressure. This means that the effective (time averaged) surface tension was determined by the choice of our lateral cell dimensions. Thus, a second ingredient was the optimal choice of cross-sectional area for lipid and cytochrome *c* oxidase molecules.

The membrane normal was oriented along the z axis, and the center of the bilayer was at $z = 0$. The protein was oriented along the z axis, in a position that left the hydrophilic residues in contact with the bulk water, and the hydrophobic residues in contact with lipid acyl chains. It was necessary to account for the *cross-section area* of two subunits of cytochrome *c* oxidase to determine the appropriate number of lipids to include in the upper and lower halves of the bilayer at the microscopic model. The estimation of the cross-sectional area of the system and the asymmetric number of lipids in the upper and lower halves of the bilayer is required, even with constant pressure algorithms in which the membrane cross-sectional area is allowed to vary. The X-ray structure does not reveal where the lipid bilayer is placed best, nor does it reveal the details of interaction between a lipid bilayer and the protein molecule. Given this starting point, the approach continued with an assumption that the experimentally measured thickness of a DMPC lipid bilayer is near to that adopted for matching to the cytochrome *c* oxidase system. We assume that part of COX where the ratio of oxygen and carbon atoms is less than 0.2 along the membrane normal for z values from -17 \AA to $+17 \text{ \AA}$ corresponds to the hydrophobic part of the protein surface (Lancaster and Michel, 1997). Periodic boundary conditions were applied in the xy directions to simulate an infinite planar layer and in the z direction to simulate a bilayer system; the periodic system has the dimensions $90 \times 90 \times 125 \text{ \AA}^3$. While the xy dimensions are kept

constant, the z dimension of the unitary cell was allowed to vary according to the constant pressure and temperature thermodynamic ensemble with fixed surface area (CPTA) (Berneche et al., 1998, Berneche and Roux, 2000). To surround the protein by a complete lipid environment, the dimensions of the system in the xy plane were set at $90 \text{ \AA} \times 90 \text{ \AA}$, corresponding to an area of 8100 \AA^2 . The total cross-sectional area for the simulation of the protein-membrane system was carefully determined because it has an important influence on the state of the bilayer. Since the average cross-sectional area of a DMPC molecule is 64 \AA^2 (Gennis, 1989; Nagle, 1993), the difference in the cross-sectional area of COX between the periplasmic side (2319 \AA^2) and the cytoplasmic side (2247 \AA^2) corresponds to ~ 1 lipid. The appropriate number of lipids was determined to be 90 for the upper layer and 91 for the lower layer. Then, a series of Van der Waals spheres were defined with a rough size near the polar head group dimensions of a DMPC (FIGURE 11, A). The position of the large spheres was obtained through molecular dynamics and energy minimization with the same periodic boundary conditions as those used in the simulation of the complete system. The spheres were constrained at $z = 17 \text{ \AA}$ and $z = -17 \text{ \AA}$ for the upper and lower layers, respectively, and were restricted to a planar motion regime. After the Van der Waals spheres had defined starting points for lipid head groups, the lipid bilayer was constructed by sampling of lipid conformations from a pre-defined lipid library containing states representative of the liquid crystalline state (Pastor et al., 1991; Venable et al., 1993) (FIGURE 11, B). In this library, the polar head groups of the DMPC are prehydrated by ~ 20 water molecules constructed on the basis of a molecular dynamics simulation of o-phosphorylcholine (o-PC) in W0 solution (Woolf and Roux, 1994). The xyz coordinates of the different spheres were used to position the center of mass of the phosphorus and nitrogen atoms of the lipid polar heads. This created a system with a large number of clashes between the alkane chains. Systematic rigid-body rotations of the lipids around the z axis and translations in the xy plane were then performed to minimize the number of unfavorable contacts and atomic overlaps. The remaining bad contacts were removed by energy minimization. The system was then fully hydrated by overlaying a pre-equilibrated box of the appropriate dimensions in x and y (FIGURE 11, C and D). The system was further refined by energy minimization before the dynamic simulation was started. In addition to the construction of water and lipid for the system, we elected to add neutralizing salt and excess salt ions to provide an electrically neutral system that was similar to experimental conditions (Ostermeier et al., 1997) for a salt concentration of 100 mM. The final microscopic system consisted of cytochrome c oxidase (2 subunits of 549 and 252 amino acids, respectively), 181 DMPC lipids (90 in the top and 91 in the bottom

layer), 88 internal crystal waters W_S , and 24,323 bulk water molecules (W_0, W_1, W_2, W_3, W_4). Additionally, 176 internal water molecules ('W12' water molecules W_G) or 755 internal water molecules ('W8' water molecules W_G) assigned by the GRID (see Chapter 2.4.3) method were included in the calculations, and 42 Na^+ and 30 Cl^- ions were

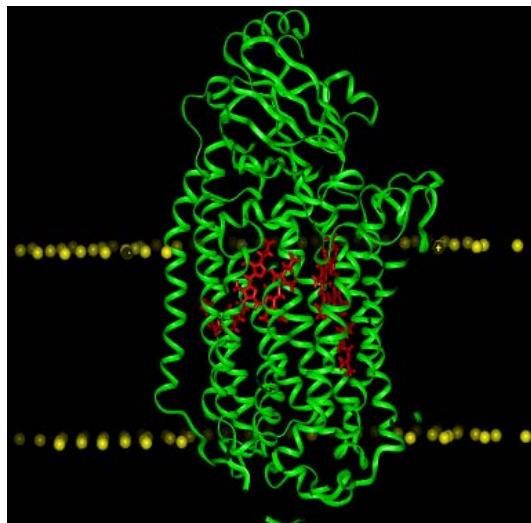


FIGURE 11 A. The low O/C ratio (less than 0.2) for Z values from -17\AA to $+17\text{\AA}$ would correspond to the hydrophobic part of the enzyme. Distributing of the Van der Waals spheres around the protein. Minimizing the energy of the Van der Waals spheres distribution.

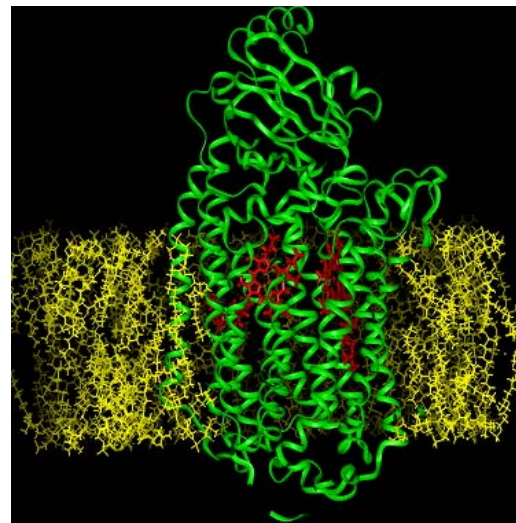


FIGURE 11 B. Dispersion the randomly selected DMPC lipids from a preequilibrated and prehydrated set around the protein

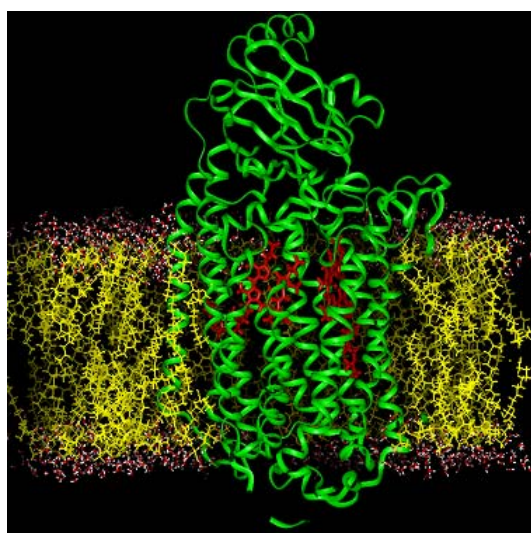


FIGURE 11 C. Addition of a water box with the same cross section as the system with a thickness of 15.6\AA .

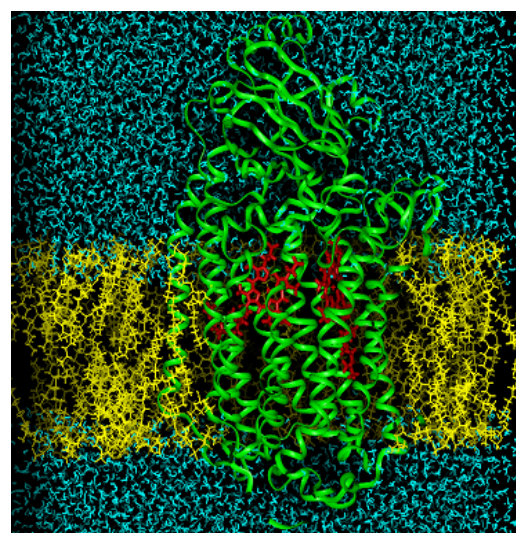


FIGURE 11 D. Generating the W_0 overlay of water for the system. An additional 176 inner water molecules, not present in the crystal structure, were included in the system.

randomly inserted (ions were positioned more than 6\AA away from the bilayer and protein, and no ion pairs were allowed to form). After solvation, the entire system consisted of

101,852 atoms ('W12' set of coordinates) and 103,589 atoms ('W8' set of coordinates). The resulting configuration of the 'W12' coordinate set is shown in FIGURE 12.

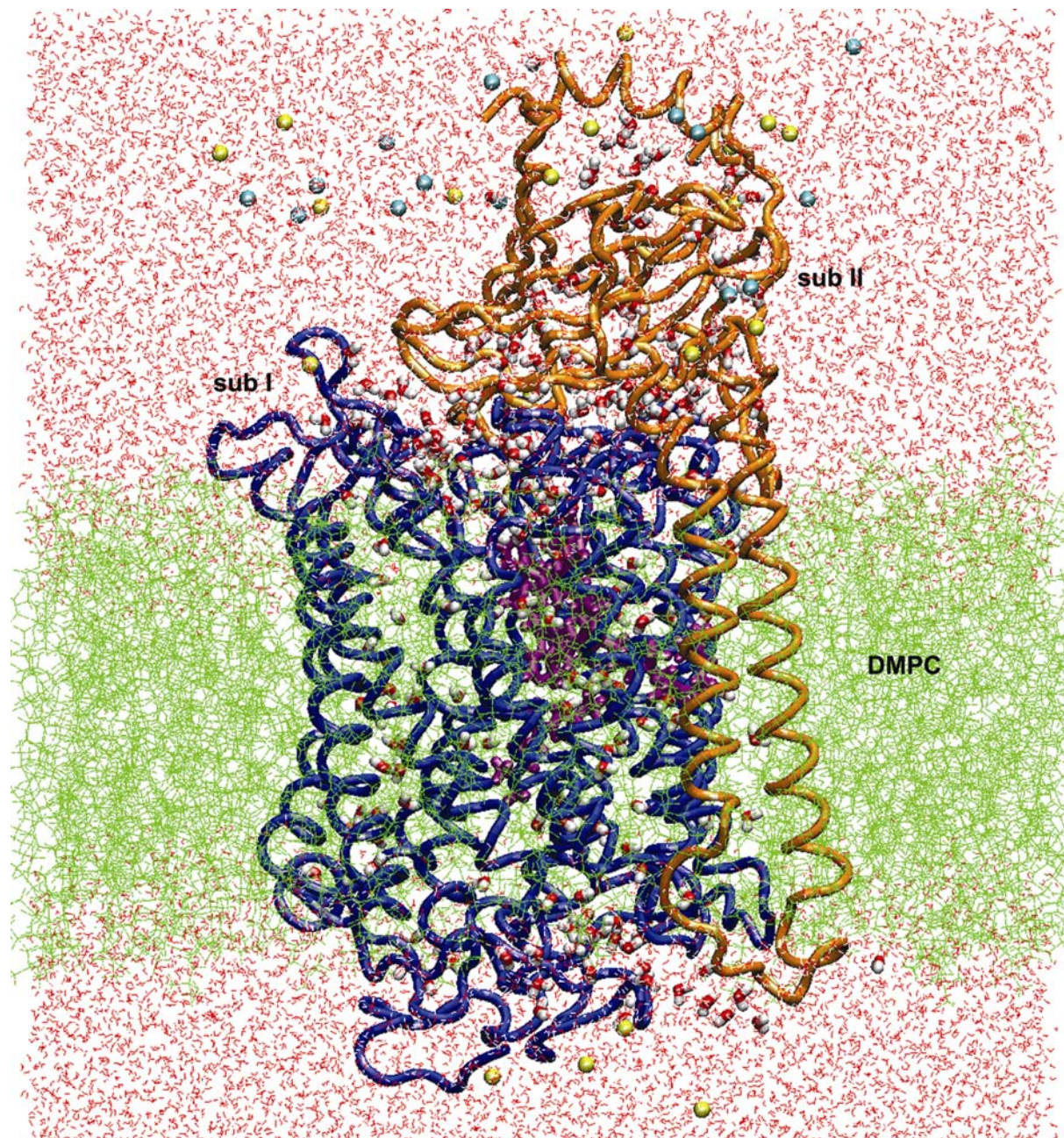


FIGURE 12 . Simulation system setup (side view): two-subunits of cytochrome *c* oxidase embedded in a DMPC membrane solvated by a 100 mM NaCl aqueous salt solution (the Na^+ ions are *yellow* and the Cl^- ions are *blue*). The lipid bilayer is depicted in green, the aqueous phase in red. The heme groups are visible in red between the helices. The snapshot was taken at the beginning of the molecular dynamics production phase after 575 ps of equilibration for the 'W12' set.

4.3.2 Equilibration and dynamics.

Starting from a system with lipid molecules, water, salt, and protein, we wanted to slowly relax the solvent (lipid and water) to the optimal state, consistent with the X-ray structure of the protein. The minimization and dynamics simulations were performed using the academic version c28b2 of the biomolecular simulation package CHARMM (Brooks et al., 1983). The protein was initially fixed and the system was minimized by 1000 steps of steepest descent (SD) followed by 1000 steps of adopted basis Newton-Raphson (ABNR) minimization. A number of energy restraints were used during the minimization and at the beginning of the equilibration period to ensure a smooth relaxation of the system toward an equilibrated configuration (Berneche et al., 1998; Berneche and Roux, 2000). A harmonic potential of $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ was applied to the backbone atoms of protein to prevent large spurious motions, the center of mass of the lipid polar heads was kept at around $z = \pm 17 \text{ \AA}$ by planar harmonic constraints to maintain the planarity of the membrane, and the penetration of water into the bilayer region was prevented by the use of planar potentials in the z direction. The protein and water constraints were then decreased to $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and were gradually reduced to obtain a free system after 575 ps of equilibration. The only remaining constraints were those on lipid headgroups (see below) and harmonic distance restraints between Cu_B and its ligands His276, His325, His326, OH^- and Cu_B , OH^- and the Fe atom of heme a_3 , the Fe atom of heme a_3 and His411, the Fe atom of heme a and His413, the Fe atom of heme a and His94, $\text{Cu}_A(\text{I})$ and Cys(II)220, $\text{Cu}_A(\text{I})$ and His(II)181, $\text{Cu}_A(\text{II})$ and Cys(II)216, $\text{Cu}_A(\text{II})$ and Met(II)227, Mg and Glu(II)78, Mg and His403, and Mg and Asp404. The simulations were performed under three-dimensional periodic boundary conditions (PBC), with a constant temperature of $T = 330 \text{ K}$ and constant pressure. The average temperature was 330 K, above the gel-liquid crystal phase transition temperature, and consistent with experimental observations (Woolf and Roux, 1996). The system was equilibrated by the use of Verlet dynamics for the first 575 ps; the time steps were 2 fs. The coordinates were saved every 5 ps and the non-bonded and image lists were updated every 20 steps. The list of nonbonded interactions was truncated at 12 \AA , using a group-based cutoff. The nonbonded Van der Waals interactions were switched off at $10 - 12 \text{ \AA}$. The electrostatic interactions were computed without truncation, using the particle mesh Ewald (PME) algorithm (Essmann et al., 1995) with an order of 4 and FFT grid points for the charge mesh per \AA were $90 \times 90 \times 125$. In the PME method implemented in CHARMM the electrostatic energy is split into a direct and a reciprocal Ewald sum. A real space Gaussian-width kappa of 0.3 \AA^{-1} was used. All bonds

involving hydrogen atoms were constrained by applying the SHAKE algorithm (Ryckaert et al., 1977).

The all-atom potential energy functions PARAM-22 for proteins (MacKerrell et al., 1995; MacKerrell et al., 1998) and phospholipids (Schlenkrich et al., 1996) were used. The TIP3 potential was used for the water molecules (Jorgensen et al., 1983). During the production trajectory, the center of mass of the protein was restrained to the center of the xy plane. The overall simulation time was about 1125 ps.

All molecular structures were drawn using the Visual Molecular Dynamic Software VMD 1.8 (Humphrey et al., 1998).

It is important to consider possible methodological limitations of our simulation protocol. The main limitation is the harmonic restraint potential constantly applied to the head groups of the DMPC lipids during the molecular dynamics production phase following equilibration to avoid the structural disorder of the lipid phase observed without restraints. Although the times of the present simulation were nearly 600 ps for equilibration and 1125 ps for the molecular dynamics production run, they seem still not long enough for a full equilibration of the lipid phase. The restraints will not, however, influence the energetics and dynamics of the internal water molecules in cytochrome *c* oxidase that was the main focus of this work.

4.3.3 Analysis of the trajectories from the simulations.

Analysis was performed in CHARMM using the abilities to interrogate trajectories and / or sets of coordinate files.

4.3.3.1 Hydrogen bonds.

The hydrogen bond patterns were analysed from the production trajectories with 0.15 ps time resolution. The criteria for a hydrogen bond (A...H – D) was that the distance between the acceptor and the hydrogen atom (A ... H) should be less than 2.5 Å with an A ... H – D angle should exceed 120° (Eriksson et al., 1995a; Eriksson et al., 1995b; Tang and Nilsson, 1999). The percentage of occupancy of a hydrogen bond was defined as the number of frames with the hydrogen bond present divided by the total number of frames used for analysis. The lifetime of a hydrogen bond was calculated as the time elapsed from its first appearance until it was first broken. The average lifetime of a hydrogen bond during the

simulation was then calculated as the average of all of its occurrences excluding those with a lifetime shorter than 1 ps.

4.3.3.2 Root mean square deviations and atomic fluctuations.

The coordinate sets from every 0.15 ps of the production run were superimposed on the initial structure of the system by minimizing the mass-weighted root mean square deviations (RMSD) of the heavy atoms from the initial structure. The average RMSD values of the C_α atoms, sidechains and some amino acids were then calculated for the entire MD trajectory.

B-factors (Debye-Waller factor) from the X-ray structure of the two-subunit structure of cytochrome c oxidase were compared with the atomic fluctuations (RMSF) in the simulation using the relation:

$$\langle \Delta r_i^2 \rangle = \frac{3B_i}{8\pi^2} \quad (4.1)$$

where Δr_i is the atomic displacement for atom i and B_i is the corresponding B-factor.

4.3.3.3 Diffusion coefficient.

The diffusion coefficients for water molecules (D_{H_2O}) in the simulated system were estimated using the Einstein relation (Eriksson et al., 1995a):

$$\lim (t \rightarrow \infty) \delta / \delta t \langle (\mathbf{R}(t) - \mathbf{R}(0))^2 \rangle = 6\mathbf{D}, \quad (4.2)$$

where $\langle \rangle$ denotes an average over the water molecules and $\mathbf{R}(t)$ the position of a water molecule at time t .

4.3.3.4 Computational details.

Internal water modelling was done on a one processor Alpha machine with 667 MHz processor using the GRID program (Goodford, 1985). Energy minimization, membrane modelling were performed on a one processor Alpha machine. Molecular dynamics simulations was performed in parallel with 32 processors, using version c28b2 of the biomolecular simulation program CHARMM (Brooks et al., 1983) on an IBM RS/6000 PS4 Regatta supercomputer at the Max Planck Society Rechenzentrum in Garching. 40 ps of simulation took 10 h CPU time on 32 Power4 processors.

4.4. Results of the simulations.

4.4.1 Average structural properties.

After constructing an initial water hydrogen network, MD simulations were carried out to study the distribution and dynamics of water molecules. The atomic system ‘W12’ at the beginning of the equilibration procedure is shown in FIGURE 12. Analysis of the deviation of

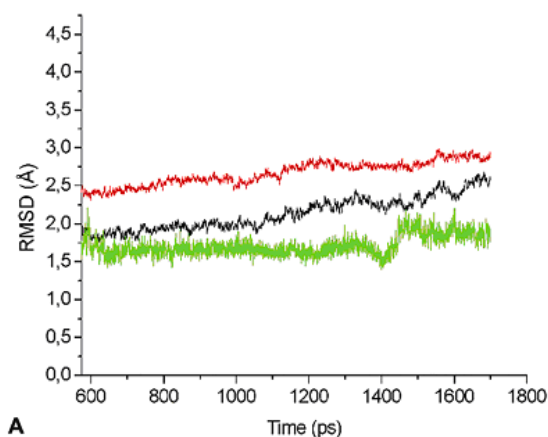
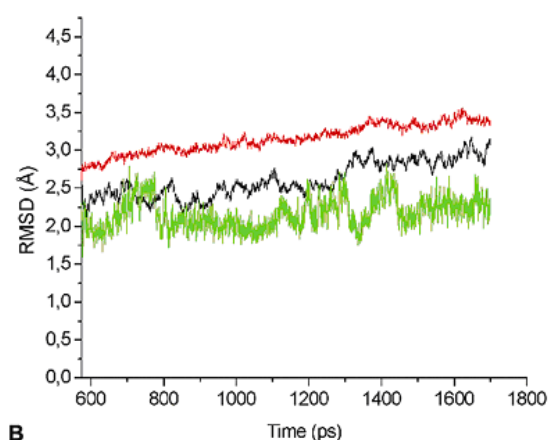
**A****B**

FIGURE 13. Root mean square deviation (RMSD) relative to the X-ray structure as a function of time, calculated over all backbone atoms (*black line*), sidechain atoms (*red line*), heme a/a_3 atoms (*green*); A) RMSD from ‘W12’ set of coordinate; B) RMSD from ‘W8’ set.

the structure from the initial crystal structure was performed to assess the stability of the simulated protein. FIGURE 13 presents the root mean square deviation (RMSD) of the C_α atoms, sidechain atoms and heme a / heme a_3 atoms from the corresponding X-ray structure with respect to the simulation time. The coordinate sets after every 0.15 ps of the production run were superimposed onto the initial structure of the system. From the beginning of the dynamics run, the heme clusters seem to have reached a rather stable state characterized by an average RMSD value from the crystal structures of 1.6 Å. The RMSD values of C_α atoms and sidechain atoms are higher. After 575 ps the average deviation remained at about 1.8 Å for the

C_{α} atoms, and at 2.5 Å for the sidechain atoms. The first 575 ps were therefore considered as equilibration and not used for analysis.

To study flexibility along the backbone, the root mean square atom-positional fluctuations of the C_{α} atoms during the production trajectories were compared with fluctuations in the crystal structure, derived from the experimental B-factors (FIGURE 14.1). The MD fluctuations are smaller than the experimental B-factors. They lack the contributions of the static lattice disorder of a 2.7 Å crystal structure. However, the MD and X-ray maxima

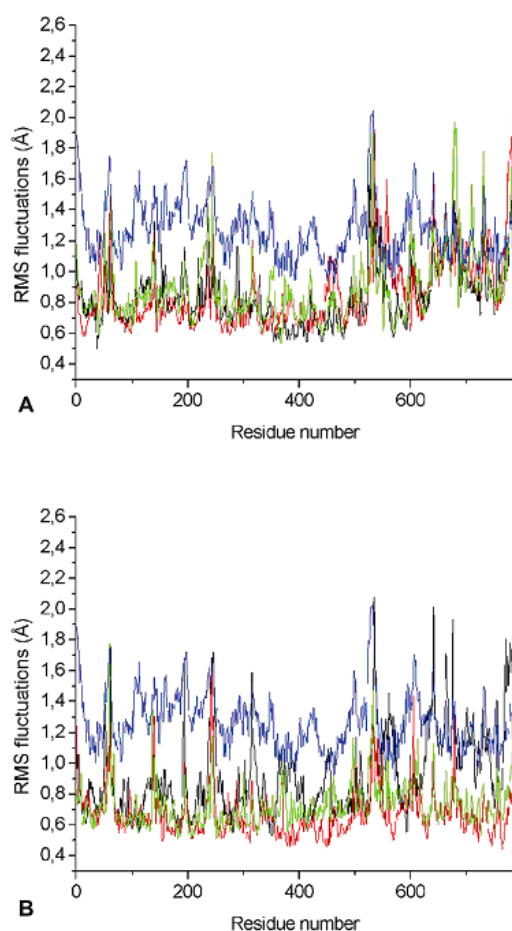


FIGURE 14.1. RMSF of the backbone atoms calculated from the molecular dynamics trajectories at 1 - 45 ps (*black line*), 450 - 495 ps (*red line*), 1080 - 1125 ps (*green line*) and from the experimental B-factor (*blue line*). All values are averaged over the individual amino acids. A) RMSF from 'W12' set. B) RMSF from 'W8' set.

and minima correlate well with one another. Regions with high flexibility are the N- and C-termini. The difference in RMSF may be due to the missing residues in the simulations. The simulated atomic system includes residues 17 - 545 for subunit I of the enzyme, and 1 - 252 for the subunit II of the enzyme (Ostermeier et al., 1997). However, the missing residues 1-16 that were not resolved in the crystallographic structure could nevertheless have an influence on the structure and dynamics of the rest of the protein. Fluctuations of the 'W12' and 'W8' systems are compared in FIGURE 14.2. No significant trend is observed and neither

simulation showed a higher fluctuation than the other; the fluctuations are ca. 0.1 Å smaller in the ‘W8’ simulations at later stages, again reflecting the denser packing.

4.4.2 Water distribution in cytochrome *c* oxidase, its dynamic properties and hydrogen-bonding network.

The identification of protein-bound water molecules and of hydrogen-bonded networks allows us to study their dynamic behaviour. We believe that changes in hydrogen bonding networks could be important for the mechanism of the electron and proton transfer coupling. Thus, we provide some insights into the changes of the hydrogen bonding network in the enzyme during simulations. Clearly, this type of analysis involves a large amount of data (trajectory files, coordinate files) and some arbitrary assumptions about what constitutes a hydrogen bonding interaction (the CHARMM potential does not use an explicit term for hydrogen bonding – see CHAPTER 3 for details). Thus, while the analysis may produce slightly different results with modified assumptions, the relative changes in the hydrogen bonding networks should be functionally relevant and are thus emphasized. The water molecules in cytochrome *c* oxidase for the ‘W12’ set at the start of the simulation are shown in FIGURE 15. Since some of the hydrogen-bonded pathways are assumed to constitute proton pathways, it is important to estimate their lifetimes, as that can provide some insight into the efficiency of proton translocation. The average hydrogen-bonding statistics for COX were computed from the 1,125 ps ‘W12’ and ‘W8’ simulations. TABLE 4 shows the time evolution of the number of hydrogen bonds between water oxygens and – protein residues. During the first 270 ps of molecular dynamics production the number increased rapidly, and then reached a steady state. This fact can be ascribed to the penetration and escape of water molecules into and out of the enzyme; an imbalance between these two processes causes the fluctuation. As the MD simulation proceeds, the system relaxes, and some water molecules gain access to the protein interior, making favourable interactions with residues.

Of the total of (176 + 88) water molecules in the ‘W12’ system, 92 % had lifetimes of hydrogen bonding of less than 100 ps, and less than 2 % had lifetimes of more than 1 ns. In the ‘W8’ system, 93 % of the total (755 + 88) water molecules had lifetimes of hydrogen bonding less than 100 ps, and less than 1 % had a lifetimes of more than 1 ns. For a further analysis of the dynamic properties of the internal water molecules in cytochrome *c* oxidase, we calculated the self-diffusion coefficient of water for both the ‘W12’ and the ‘W8’ sets of

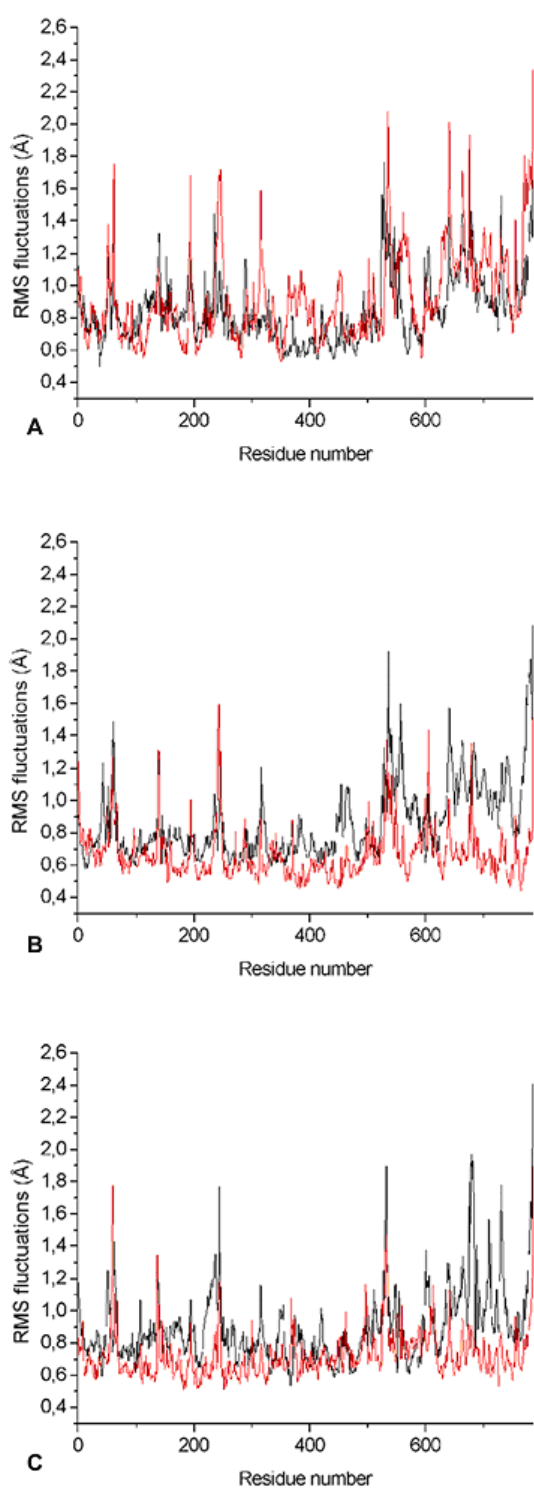


FIGURE 14.2. RMSF of the backbone atoms calculated from the molecular dynamics trajectories for the ‘W12’ and ‘W8’ sets of water molecules during 1 - 45 ps (FIGURE 14.2.A), 450 - 495 ps (FIGURE 14.2.B), 1080 - 1125 ps (FIGURE 14.2.C). All values are averaged over the individual amino acids; the backbone atoms from the ‘W12’ set of coordinates are given in *black* color, the backbone atoms from the ‘W8’ set of coordinates are shown in *red* color.

coordinates. The diffusion coefficients (D) of the water molecules were estimated from the mean square displacement obtained from the molecular dynamics trajectories (equation 4.2).

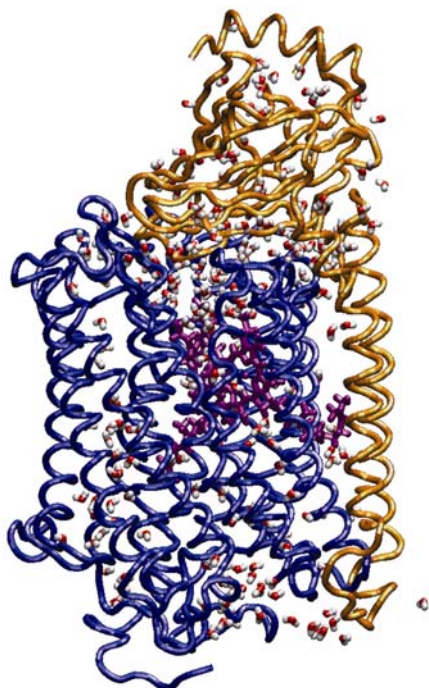


FIGURE 15. The distribution of water molecules in COX during the simulation.

In our simulations, water molecules were allowed to leave the protein during the period analysed. For the 88 water molecules found in the crystal structure plus the 176 water molecules from the GRID calculations ('W12' set of coordinates), $D_{\text{H}_2\text{O}}$ was $(3.0 \pm 0,005) * 10^{-9} \text{ m}^2/\text{s}$, and for the 88 structural water molecules plus the 755 water molecules from the 'W8' set of coordinates, $D_{\text{H}_2\text{O}}$ was $(3.1 \pm 0,005) * 10^{-9} \text{ m}^2/\text{s}$. Experimental values for $D_{\text{H}_2\text{O}}$ at 300 K are $2.3 * 10^{-9} \text{ m}^2/\text{s}$ (Mills et al., 1973), and in pure TIP3 water (Jorgensen et al., 1983) at 300 K $D_{\text{H}_2\text{O}}$ has been estimated to be $1.3 - 4.2 * 10^{-9} \text{ m}^2/\text{s}$ (Norberg and Nilsson, 1994).

The simulations show interesting features concerning the water distribution around the hemes. The hemes and the axial histidine ligands His94, His413, and His411 were constructed to be in similar conformations as in the post-1AR1 structure. In our MD simulations we find that water molecules form hydrogen bonds with the ND1 atom of the axial His413. For the whole period of the simulations, His411 formed a hydrogen bond with the O atom of Tyr391 with an occupancy of 60 % and a residence time of 691 ps, and with water molecule $\text{W}_{\text{G}114}$ with an occupancy of 46 % and a residence time of 513 ps. The water molecules bound to the axial histidines are not the only water molecules fixed inside the protein for quite a long time. The majority of water molecules had shorter than 100 ps residence times, but a few water molecules in the binuclear center had longer residence times, up to the whole simulation time (TABLES 5, 6, 7). In FIGURE 18 (A and B) and in TABLE 7 we show that the structural water $\text{W}_{\text{S}69}$ in both the 'W12' and the 'W8' sets and calculated water $\text{W}_{\text{G}7}$ in the 'W8' set

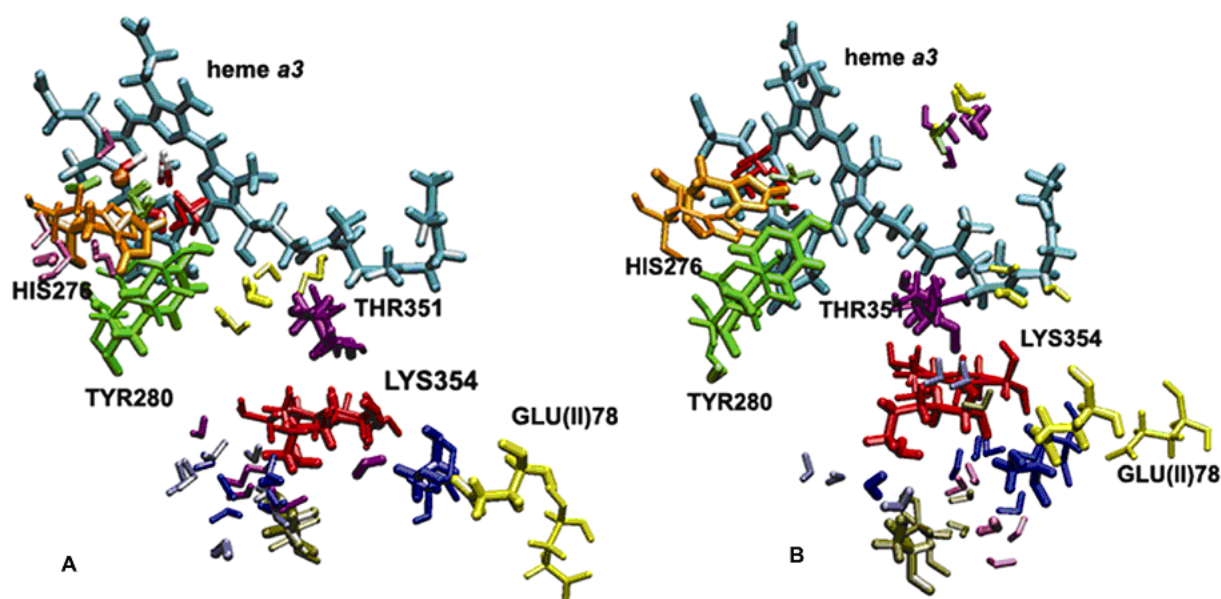


FIGURE 16 (A, B). Multiple configurations of selected residues and water molecules in the K-pathway for the ‘W12’ (A) and ‘W8’ (B) set of coordinates observed from MD trajectories. Initial coordinates are shown with the thick licorice. Positions of the selected residues after 1125 ps are shown with the thin licorice. Snapshots for selected water molecules are taken after 225 ps, 450 ps, 675 ps, 900 ps and 1125 ps of simulations.

For the ‘W12’ set of coordinates water molecule W_S6 (structural water) is represented in *blue*, W_S9 in *yellow*, W_S69 in *red*, W_S81 in *rose*, W_G7 (internal GRID water) in *light green*, W_G149 in *pink*, and W_G161 in *magenta*. For the ‘W8’ set of coordinates water molecule W_S6 is represented in *blue*, W_S69 in *red*, W_S88 in *yellow*, W_G13 in *rose*, W_G188 in *light green*, W_G194 in *gold*, W_G379 in *pink*, W_G491 in *magenta*, and W_26484 in *orange*.

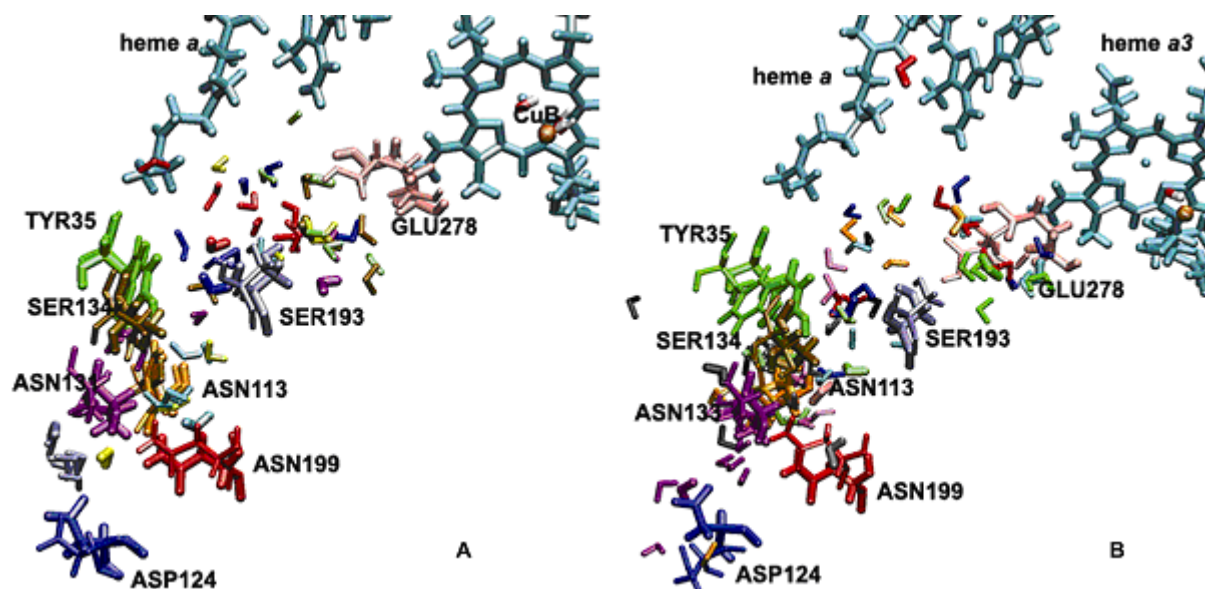


FIGURE 17 (A, B). Multiple configurations of selected residues and water molecules in the D-pathway for the ‘W12’ (A) and ‘W8’ (B) set of coordinates observed from MD trajectories. Initial coordinates are shown with the thick licorice. Positions of the selected residues after 1125 ps of MD run are shown with the thin licorice. Snapshots for selected water molecules are taken after 225 ps, 450 ps, 675 ps, 900 ps and 1125 ps of simulations. For the ‘W12’ set of coordinates water molecule W_S38 is represented in *blue*, W_S40 in *yellow*, W_S80 in *red*, W_G8 in *light green*, W_G12 in *gold*, W_G80 in *light blue*, W_G131 in *magenta*, and W_G144 in *pink*.

For the ‘W8’ set of coordinates water molecule W_S3 is represented in *rose*, W_S38 in *blue*, W_G9 in *light blue*, W_G11 in *orange*, W_G180 in *light green*, W_G198 in *magenta*, W_G220 in *green*, W_G238 in *red*, W_G540 in *grey*, and W_1232 in *black*.

remained bound to Cu_B ligated OH^- during 1125 ps of simulation. There are quite a large number of water molecules trapped within cavities, although in certain cases there is exchange with the W_0 water. The behaviour of this cavity water is like that of mobile water molecules in W_0 water with many alternating hydrogen bond partners. TABLE 7.A lists waters with 3, 4 and 5 partners, and we could only list the interactions with more than 20 % occupancy due to space constraints.

Depending on the definition of the hydrogen bond geometry used, our analysis sometimes shows that a contact between a water molecule and a residue breaks while at the same time a new interaction between the same water molecule and another neighbouring residue develops. The water molecule changes its orientation only and, as a result, remains trapped in that region for quite a long time, but the average lifetime of the hydrogen bond is smaller than the overall residence time of the water molecule. An example of such a water molecule is water W_G158 (TABLE 7.B), which can form hydrogen bonds with HIS325 and with water W_01385 .

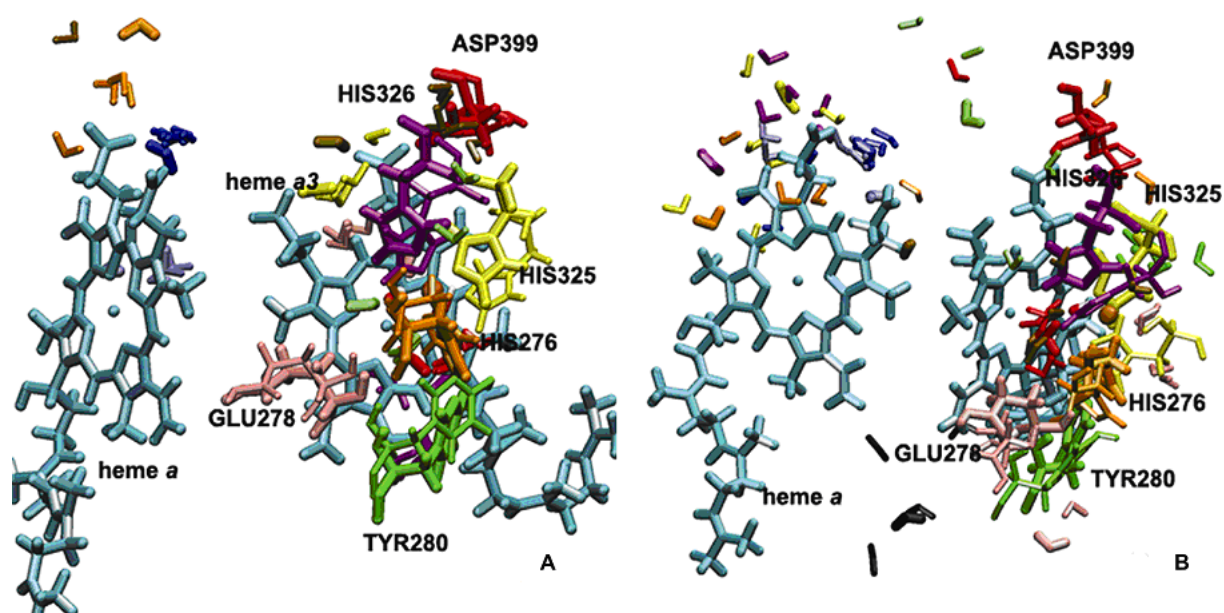


FIGURE 18 (A, B). Multiple configurations of selected residues and water molecules close to the heme a_3 – Cu_B region for the ‘W12’ (A) and ‘W8’ (A) set of coordinates observed from MD trajectories. Initial coordinates are shown with the thick licorice. Positions of the selected residues after 1125 ps of MD run are shown with the thin licorice. Snapshots for selected water molecules are taken after 225 ps, 450 ps, 675 ps, 900 ps and 1125 ps of simulations.

For the ‘W12’ set of coordinates water molecule W_S13 is represented in *blue*, W_S69 in *red*, W_G5 in *yellow*, W_G7 in *green*, W_G30 in *pink*, W_G144 in *magenta*, W_G173 in *orange*, W_4172 in *gold*, and W_4214 in *rose*.

For the ‘W8’ set of coordinates water molecule W_S13 is represented in *blue*, W_S14 in *yellow*, W_S69 in *red*, W_G31 in *light green*, W_G42 in *orange*, W_G46 in *magenta*, W_G51 in *pink*, W_G157 in *green*, W_G188 in *rose*, W_G189 in *black*, and W_G453 in *gold*.

The reason for performing two parallel simulations that only differ by the amount of internal water molecules added, is to determine the “true” level of hydration of the protein interior. All water molecules of the ‘W8’ set had GRID energies below -8 kcal/mol, indicating high occupancy. The hydrogen bond lifetimes are longer for the ‘W12’ set than for the ‘W8’ set. This is due to a greater number of hydrogen bonding possibilities in the ‘W8’ system that is filled with water more densely. The MD simulations indicate that the solvent distribution is more diffuse for higher hydration states (Helms and Wade, 1998). According to our observations, the hydrogen-bonded network in cytochrome *c* oxidase is not uniformly distributed, and the degree of water arrangement is variable, since side chains induce variable levels of the water ordering (Bret et al., 2002).

4.4.3 The distribution of water in the K- and D- pathways.

Detailed information about energetically favourable water binding sites along the K- and D-pathways is essential for understanding the mechanism of proton transfer in COX. Many water molecules exchange between the protein interior and bulk and are thus unlikely to be detected in the crystal structure experiments. By examining the MD trajectories for water molecules and using the definition of hydrogen bonding by Tang and Nilsson (1999), the number of water molecules, which are likely to form the hydrogen bonded network, can be obtained. The numbers of water molecules in the K- and D- pathways at different times are given in TABLE 8. The GRID method placed 7 water molecules in the K-pathway at the beginning of the simulations. It can be seen that the number of water molecules is not static, but ranges quite substantially from 3 to 8 with an average of 6.5 for the ‘W12’ set of coordinates, and from 5 to 15 with an average of 9.5 for the ‘W8’ set. These sites participate in the formation of a reasonable hydrogen bonded network in the K-pathway. This network is not continuous (FIGURE 16, *A* and *B*). TABLE 5 (A, B) lists the hydrogen bonds inside the K-pathway with occupancies higher than 20 %. A hydrogen-bonded connection can be seen at the beginning of the K-pathway and also at the end, leading up to heme a_3 , but there is no direct or water-mediated connection between Lys354 and Thr351 at the beginning of the ‘W12’ simulation. After 500 ps of the MD run we observe a significant reorientation of the sidechain of neutral Lys354 and formation of a hydrogen bond between the O atom of water W_G149 and the HZ1 atom of Lys354 with an occupancy of 6 % (data not shown). Due to the low occupancy, we can conclude that Lys354 has little effect on water ordering, perhaps because of its long flexible side chain. For the ‘W8’ simulations, the hydrogen-bonded network was continuous after the placement of GRID water molecules. However, after the

first 45 ps of dynamics many water molecules had left their positions and the hydrogen bonded network did not remain continuous.

For the D-pathway, the number of solvent molecules ranged from 18 to 24 with an average of 21.5 for the ‘W12’ set, and from 24 to 29 with an average of 26.33 for the ‘W8’ set. 16 sites took part in the formation of a quite stable continuous hydrogen-bonded chain leading from Asp124 to the region close to Glu278, but only 4 waters in the hydrophobic pore between Ser193 and Glu278 were found to be required to form a hydrogen-bonded chain, which could be used for proton transfer (FIGURE 17, *A* and *B*). The average lifetime of the hydrogen-bonded chain formed by these four water molecules was much shorter than the lifetime of hydrogen bonds formed by each individual pair of water molecules. Some of these waters have a high mobility during the MD simulations. The change in the number of water molecules arises both from water exchanging with W_0 water or with the interior of protein (Henchman et al., 2002). As an example, we examined the trajectory of water W_{4172} from the ‘W12’ set and found that this water molecule was able to traverse a distance of ~ 3.5 Å during ~ 0.3 ns. Although the placement of water molecules identifies an efficient pathway for protons up to Glu278, the connection of the D-pathway with other protonatable sites beyond Glu278 is not clear. One possibility is that the protonated glutamic acid side chain could flip upward and deliver its proton either to the binuclear center or to a heme a_3 propionate group as proposed (Iwata et al., 1995). The space between Glu278 and the binuclear center / heme a_3 propionate group is proposed to be a part of the oxygen diffusion channel (Svensson-Ek et al., 2002) which is hydrophobic and does not contain any crystallographically identifiable water molecules. This area may be filled with mobile water molecules as has been suggested by Riistama et al. (1997). Within the time of 1.125 ns of MD simulations, such a conformational change of the protonated Glu278 was not observed. Our data do not show that there is a connection between the protonated Glu278 and the O1A propionate group of heme a_3 for the COX in the fully oxidized state. However, an additional short simulation of COX in the fully oxidized state with deprotonated Glu278 shows a significant reorientation of the sidechain of Glu278 and the formation of a hydrogen bond chain between Glu278 via water molecules up to the O1A atom of the heme a_3 propionate (FIGURE 19, *A—D*) (these data were presented at the 12th European Bioenergetics Conference 2002; referenced by Zheng et al., 2003). This difference may be of interest for routing protons in different parts of the catalytic cycle. Locally, the hydration network follows the reorientation of particular residues that change their conformation. Examples are shown in FIGURE 16, *A* and *B*, FIGURE 17, *A* and *B*, FIGURE 18, *A* and *B*.

A large number of water molecules was observed in the gap between subunit I and II and in the hydrophilic cavity above heme a and heme a_3 . This result is in agreement with previous studies (Hofacker and Schulten, 1998; Zheng et al., 2003). These water molecules form a network of hydrogen bonds and connect the propionate groups of heme a and heme a_3 to the external aqueous phase in many ways. Since it has been suggested that the propionate

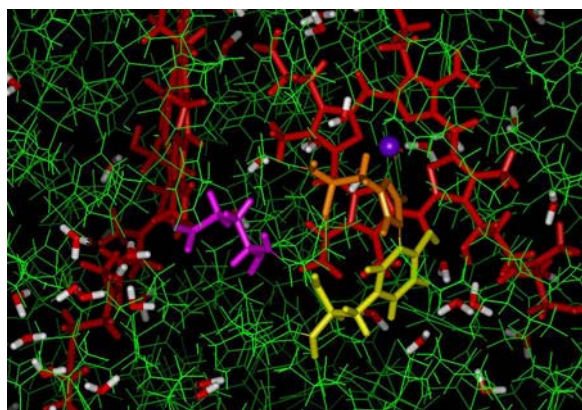


FIGURE 19 A. 0 ps.

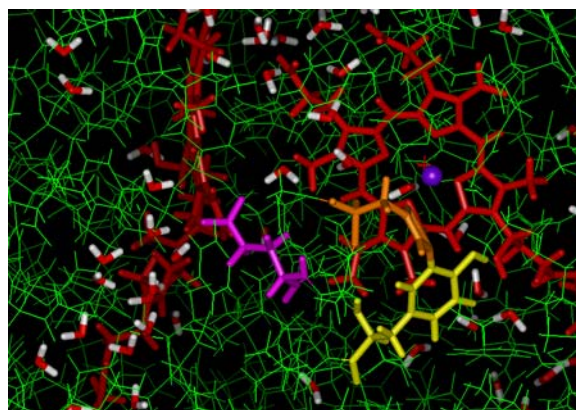


FIGURE 19 B. 200 ps.

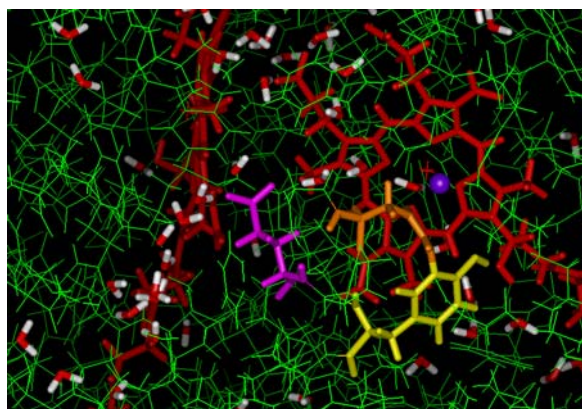


FIGURE 19 C. 235 ps.

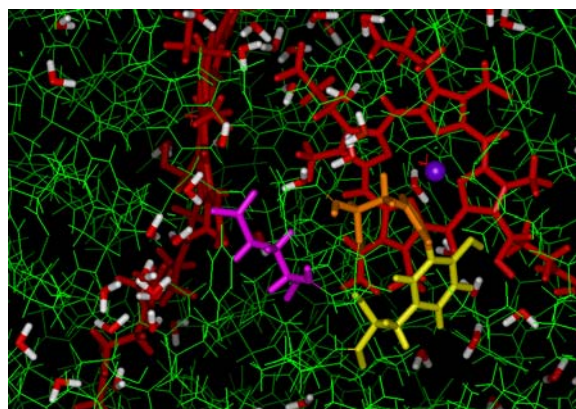


FIGURE 19 D. 270 ps.

FIGURE 19. (A—C). A and B. System after 200ps of equilibration using Langevin dynamics. C and D. Equilibration of the constructed system using Verlet algorithm for constant pressure dynamics. Pictures shows the re-orientation of internal water molecules and some of the residues (*pink* – Glu278, *yellow* – Tyr280, *orange* – His276, *cyan* sphere – copper atom, *red* – hemes) during equilibration.

groups are involved in proton pumping (Michel, 1998; Puustinen and Wikström, 1999), this network could be a possible pathway for proton exit.

In FIGURE 18, *A* and *B* we show that the cavities above heme a and heme a_3 contain a number of positionally stable, and also highly mobile water molecules with hydrogen bond occupancies of less than 30%. For example, the modelled internal water W_{G5} is present and essentially always hydrogen bonded to the O2A atom of the heme a_3 propionate for 828 ps in

the ‘W12’ set. Noteworthy are the hydrogen bonds between some water molecules and the propionate groups of the hemes, which may be of structural and energetic importance. In agreement with previous electrostatics calculations (Kannt et al., 1998), it was found that Asp399 forms a hydrogen bond with the O1A heme a_3 propionate with an occupancy of 100% and a residence time of 1123.2 ps, and also forms a hydrogen bond with water W₄₁₇₂ with an occupancy of 31 % and a residence time of 344.5 ps (FIGURE 18, *A* and *B*) for the ‘W12’ set.

Another interesting observation is that Arg473 and Arg474 seem to play an important role in directing the water network orientation in the region above heme a / heme a_3 . Their highly mobile side chains significantly influence the reorientations of water molecules in both the electron transfer and proton exit pathways (Puustinen and Wikström, 1999). The HH21 atom of Arg473 forms a hydrogen bond with the O1D atom of the heme a_3 propionate with an occupancy of 63 % and a residence time of 710 ps. The HH21 atom of Arg474 forms a hydrogen bond with the O2D atom of the heme a propionate with an occupancy of 87 % and a residence time of 990.6 ps. These results show that the Δ -propionate of the heme a is strongly stabilized by charge interactions with this arginine, in agreement with previous electrostatics calculations (Kannt et al., 1998). The simulation with the ‘W8’ set shows a significant reorientation of Tyr167 in the direction away from the hemes. Although there is no experimental evidence concerning the role of this residue, it is very likely that Tyr167, which is located close to Arg474, plays a role in the formation of a hydrogen bonded connection to Arg474 via mobile water molecules, and to the proton exit or electron transfer pathways.

TABLE 4. Statistics of the hydrogen bond network for the ‘W12’ and ‘W8’ system during the MD production run.

Time, ps	$\geq 90\%$	$\geq 75\%$	$\geq 50\%$	$\geq 25\%$	Total		$\geq 90\%$	$\geq 75\%$	$\geq 50\%$	$\geq 25\%$	Total	
45	23	47	96	266	1455		23	39	107	265	1348	
90	31	56	119	289	1417		30	53	112	278	1313	
135	29	65	122	301	1325		32	59	120	272	1285	
180	36	62	110	304	1353		27	53	110	277	1288	
225	33	53	116	313	1440		33	72	128	279	1292	
270	37	63	126	381	1400		32	62	132	278	1255	
315	37	65	132	306	1358		28	61	131	276	1248	
360	44	67	133	311	1449		24	59	125	288	1240	
405	39	63	138	292	1316		33	57	132	306	1294	
450	37	66	115	296	1406		31	57	117	291	1297	
495	34	58	132	293	1427		46	71	137	280	1228	
540	33	59	128	309	1379		33	63	129	283	1239	
585	33	61	134	310	1335		26	58	130	277	1280	
630	38	69	130	291	1335		27	62	128	281	1282	
675	37	61	133	292	1345		33	64	129	290	1249	
720	45	67	130	306	1358		30	59	127	299	1284	
765	40	72	130	290	1363		31	59	118	309	1321	
810	37	78	138	315	1321		36	57	120	288	1290	
855	37	76	142	314	1374		28	55	125	293	1270	
900	42	72	128	309	1358		36	63	133	289	1262	
945	44	75	139	302	1407		34	70	136	300	1254	
990	33	68	131	300	1411		42	67	135	280	1323	
1035	37	70	135	313	1434		37	55	118	305	1352	
1080	33	69	136	324	1447		36	75	131	281	1271	
1125	40	66	126	310	1458		40	66	128	290	1264	

TABLE 5.A. Hydrogen bonds between internal water molecules and amino acids in the K-pathway during 1125 ps of molecular dynamics production for the ‘W12’ set of water molecules binding sites (with occupancy ≥ 20 %) (all listed protein residues belong to subunit I).

Hydrogen bonds				Occup.	Aver.	Events			
				(%)	lifetime				
OH		OH2	-	WS	69	H2	100	1125.0	1
OH		OH2	-	W_G	7	H1	100	1125.0	1
TRP	358	HN	-	LYS	354	O	98	12.6	87
HIS	276	HE1	-	W _s	69	OH2	98	13.1	84
THR	361	HG1	-	SER	357	O	97	52.2	21
SER	357	HN	-	ILE	353	O	94	5.9	180
TYR	280	O	-	LEU	284	HN	92	7.0	148
TRP	358	HE1	-	SER	291	OG	86	4.8	201
SER	291	HN	-	PHE	287	O	86	4.9	197
THR	351	HG1	-	ILE	347	O	85	4.8	199
HEMA3		HO11	-	W_s	69	OH2	74	13.2	63
HIS	326	HE1	-	W_G	7	OH2	73	2.8	296
TYR	280	HD2	-	HIS	276	O	70	2.4	327
SER	357	OG	-	GLY	304	HN	68	2.6	295
THR	361	HN	-	SER	357	O	64	2.4	302
HIS	276	HE1	-	W _G	7	OH2	59	2.1	315
SER	295	HN	-	SER	291	O	58	2.8	234
TYR	280	HH	-	W _s	69	OH2	49	9.7	57
SER	291	HG1	-	PHE	287	O	47	5.5	96
HIS	325	HE1	-	W _s	69	OH2	46	2.9	177
TRP	272	O	-	HIS	276	HN	44	2.9	171
SER	295	HG1	-	SER	291	O	42	3.6	129
TYR	280	HN	-	HIS	276	O	41	2.3	205
W _s	6	OH2	-	W _G	161	H1	27	4.1	75
GLY	319	O	-	W_s	81	H2	21	79.2	3
W_G	161	OH2	-	W_G	149	H2	21	3.5	68

TABLE 5.B. Hydrogen bonds between internal water molecules and amino acids in the K-pathway during 1125 ps of molecular dynamics production for the ‘W8’ set of water molecules binding sites (with occupancy ≥ 20 %) (all listed protein residues belong to subunit I).

Hydrogen bonds				Occup.	Aver.	Events			
				(%)	lifetime				
OH		OH2	-	W_s	69	H2	100	1125.0	1
TYR	280	O	-	LEU	284	HN	93	6.1	171
TRP	358	HN	-	LYS	354	O	77	4.6	191
SER	357	HN	-	ILE	353	O	77	3.4	256
TRP	272	O	-	HIS	276	HN	75	3.5	243
TRP	358	HE1	-	SER	291	OG	66	2.9	259
SER	291	HN	-	PHE	287	O	61	2.6	266
SER	295	HG1	-	SER	291	O	59	4.8	137
SER	291	HG1	-	PHE	287	O	54	9.6	63
TYR	280	HD2	-	HIS	276	O	53	1.9	306
THR	361	HN	-	SER	357	O	53	3.4	175
THR	351	HG1	-	ALA	348	O	50	7.2	78
THR	361	HG1	-	SER	357	O	48	11.8	46
TYR	280	HN	-	HIS	276	O	43	2.1	225
SER	357	HG1	-	ILE	353	O	39	9.2	48
TYR	280	HH	-	W_G	188	OH2	33	6.9	54
SER	295	HN	-	SER	291	O	30	1.8	187
PRO	350	O	-	W_G	379	H1	28	5.1	61
VAL	349	O	-	W₂	6384	H2	27	4.0	76
THR	351	HN	-	ALA	348	O	27	2.5	121
W _s	6	H2	-	W _G	194	OH2	25	6.9	40
SER	357	OG	-	GLY	304	HN	24	2.4	114
W _s	6	OH2	-	W _G	164	H1	23	4.3	61
HIS	276	HE1	-	W _s	69	OH2	22	2.7	90
W ₂	6384	OH2	-	GLY	352	HN	22	2.4	106
VAL	349	O	-	W ₂	6384	H1	21	3.4	71
GLY	304	O	-	W _G	194	H2	21	4.0	61
W _s	88	OH2	-	W _G	491	H1	20	3.4	66

TABLE 6.A. Hydrogen bonds between internal water molecules and amino acids in the D-pathway during 1125 ps of molecular dynamics production for the ‘W12’ set of water molecules binding sites (with occupancy ≥ 20 %) (all listed protein residues belong to subunit I).

Hydrogen bonds							Occup.	Aver.	Events
							(%)	lifetime	
TYR	35	O	-	ALA	39	HN	100	40.0	28
ASN	199	O	-	THR	203	HG1	100	70.1	16
TYR	35	HN	-	ILE	31	O	98	17.3	64
SER	134	HN	-	LEU	130	O	98	15.2	73
ASN	199	O	-	THR	203	HN	97	8.8	123
SER	193	HN	-	SER	183	O	95	6.6	163
ASN	199	OD1	-	PHE	127	HN	95	5.9	181
GLE	278	HN	-	PHE	274	O	94	8.4	126
SER	193	O	-	ALA	197	HN	90	5.5	183
THR	26	O	-	W_G	144	H1	86	10.9	89
ASN	131	HD22	-	W_G	144	OH2	78	4.3	205
ASP	124	O	-	ASP	124	HN	76	3.0	281
SER	134	O	-	TYR	138	HN	74	2.6	318
SER	134	O	-	SER	192	HG1	73	6.9	118
TYR	35	HH	-	W_G	131	OH2	64	18.4	39
ASN	113	OD1	-	ASN	131	HD21	64	3.6	202
ASP	124	OD1	-	W _G	144	H2	53	85.7	7
ASP	124	OD1	-	MET	125	HN	52	6.5	90
ASP	124	OD1	-	HSD	28	HD1	51	8.4	69
ASN	113	HN	-	GLY	109	O	49	2.5	221
ASN	199	HN	-	LEU	195	O	49	2.3	242
ASN	113	HD22	-	GLY	109	O	41	3.2	147
ASN	199	HD22	-	W_G	143	OH2	41	3.5	130
ASP	124	OD2	-	MET	125	HN	40	9.9	45
ASN	113	HD21	-	ASN	131	OD1	40	5.5	82
ASN	131	O	-	TYR	135	HN	39	1.9	225
SER	134	OG	-	W_G	131	H1	38	4.2	103
SER	134	HG1	-	LEU	130	O	34	3.1	125
ASP	124	OD2	-	W _G	144	H2	34	47.4	8
ASN	131	OD1	-	W _G	131	H2	32	3.3	107
ASP	124	OD2	-	HSD	28	HD1	29	4.9	68
ASN	199	HD22	-	W_G	80	OH2	21	2.7	87
SER	134	OG	-	W_G	131	H2	20	3.5	63
ASN	131	OD1	-	W_G	131	H1	20	3.2	70
W_G	8	H1	-	W_G	12	OH2	20	10.5	22

TABLE 6.B. Hydrogen bonds between internal water molecules and amino acids in the D-pathway during 1125ps of molecular dynamics production for the ‘W8’ set of water molecules binding sites (with occupancy ≥ 20 %) (all listed protein residues belong to subunit I).

Hydrogen bonds							Occup.	Aver.	Events
							(%)	lifetime	
TYR	35	O	-	ALA	39	HN	99	22.3	50
ASN	199	O	-	THR	203	HN	96	8.6	125
TYR	35	HN	-	ILE	31	O	93	6.4	162
ASN	199	OD1	-	PHE	127	HN	93	7.8	134
SER	193	O	-	ALA	197	HN	87	4.6	215
SER	134	HN	-	LEU	130	O	81	4.7	195
ASN	199	O	-	THR	203	HG1	78	11.6	76
SER	134	O	-	TYR	138	HN	65	2.7	275
SER	193	HN	-	SER	189	O	50	2.8	202
ASN	131	HN	-	PHE	127	O	46	2.2	234
TYR	35	HH	-	W_G	540	OH2	43	10.0	49
TYR	114	O	-	W_G	198	H1	41	5.0	91
ASP	124	O	-	ASP	124	HN	38	1.9	225
ASP	124	OD1	-	MET	125	HN	37	4.8	86
PHE	274	O	-	GLE	278	HN	33	3.8	99
GLE	278	HN	-	PHE	274	O	33	3.8	99
ASN	113	HD22	-	GLY	109	O	32	2.6	140
ASP	124	OD2	-	MET	125	HN	31	4.4	79
ASN	199	HN	-	LEU	195	O	27	1.7	177
ASN	113	HN	-	GLY	109	O	24	2.4	110
SER	193	HG1	-	W_G	220	OH2	23	9.2	28
SER	134	HG1	-	LEU	130	O	23	4.8	55
TYR	35	HH	-	W₁	232	OH2	22	15.4	16
ASN	131	OD1	-	W_G	540	H1	20	3.0	76

TABLE 7.A. Hydrogen bonds between internal water molecules and amino acids in the heme a_3 / Cu_B region during 1125 ps of molecular dynamics production for the ‘W12’ set of water molecules binding sites (with occupancy $\geq 20\%$) (all listed protein residues belong to subunit I).

Hydrogen bonds				Occup.	Aver.	Events			
				(%)	lifetime				
HEMA3	O1A	-	ASP	399	HD2	100	140.4	8	
OH ⁻	OH2	-	W _s	69	H2	100	1125.0	1	
OH ⁻	OH2	-	W _G	7	H1	100	1125.0	1	
HIS	276	HE1	-	W _s	69	OH2	98	13.1	84
ARG	473	HE	-	W _s	13	OH2	98	15.3	72
HEMA3	O1A	-	HSD	403	HD1	96	12.6	86	
HIS	413	O	-	SER	417	HN	91	4.5	230
ASP	399	O	-	HSD	403	HN	89	5.3	189
ASP	399	HN	-	GLN	395	O	83	4.3	218
HEMA3	O2A	-	W _G	5	H1	74	23.0	36	
HEMA3	HO11	-	W _s	69	OH2	74	13.2	63	
HIS	326	HE1	-	W _G	7	OH2	73	2.8	296
HIS	413	HN	-	VAL	409	O	70	2.9	271
HIS	276	O	-	TYR	280	HD2	70	2.4	327
HEMA3	O2D	-	ARG	473	HH12	69	17.9	43	
HEMA3	O2A	-	HSD	403	HD1	65	2.8	261	
HIS	94	O	-	MET	98	HN	64	3.0	243
HIS	94	HD1	-	SER	46	O	61	2.7	252
HIS	411	HN	-	TYR	407	O	60	2.6	266
HIS	276	HE1	-	W _G	7	OH2	59	2.1	315
VAL	408	O	-	W ₄	214	H2	54	9.9	61
HEMA3	O1D	-	ARG	473	HH21	49	21.0	26	
TYR	280	HH	-	W _s	69	OH2	49	9.7	57
HIS	411	HE1	-	W _G	114	OH2	48	2.6	208
HIS	325	HE1	-	W _s	69	OH2	46	2.9	177
HIS	276	HN	-	TRP	272	O	44	2.9	171
TYR	328	HN	-	W ₄	172	OH2	43	3.9	124
GLY	319	O	-	W ₃	5248	H1	43	22.0	22
GLN	463	OE1	-	W _G	30	H2	43	3.1	159
TYR	280	HN	-	HIS	276	O	41	2.3	205
HIS	325	HD1	-	VAL	322	O	41	8.2	57
HIS	276	O	-	TYR	280	HN	41	2.3	205
ARG	473	HH22	-	W _s	13	OH2	41	3.3	140
W ₀	2000	H1	-	W ₃	5248	OH2	41	8.3	55
W _G	5	OH2	-	W ₄	214	H1	41	6.3	73
GLN	463	OE1	-	W _G	30	H1	40	3.0	151
VAL	408	O	-	W ₄	214	H1	39	10.5	42
THR	389	OG1	-	HIS	411	HD1	35	2.5	159
HEMA3	O2A	-	W ₃	1204	H1	34	26.9	14	
HEMA	O2D	-	ARG	474	HE	34	6.9	56	
HEMA3	O1D	-	ARG	473	HH12	32	7.8	46	
HEMA	OMA	-	GLN	463	HE21	32	2.0	184	
TYR	475	HN	-	W _G	173	OH2	31	3.7	95
ASP	399	OD1	-	W ₄	172	H1	31	6.5	53
W _G	130	OH2	-	W _G	114	H2	29	4.2	79
HEMA3	O1A	-	W ₄	172	H2	28	4.2	75	
TYR	475	HN	-	W _s	42	OH2	28	2.5	130
HEMA3	O1D	-	W ₃	1204	H2	27	20.0	15	
HIS	94	HN	-	MET	90	O	27	2.2	135
ARG	54	HE	-	W _G	31	OH2	26	11.2	26
HEMA3	O1A	-	W ₄	172	H1	25	4.8	60	
PHE	383	O	-	W _G	129	H2	23	4.0	65
GLY	387	O	-	W _G	114	H1	23	4.0	64
ARG	54	HE	-	W ₄	220	OH2	23	14.9	17
W _G	130	OH2	-	W _G	114	H1	23	3.7	69
THR	50	OG1	-	W ₄	220	H2	22	2.6	97
W _s	13	OH2	-	W _G	31	H1	22	2.6	94
W _G	130	H1	-	W _G	129	OH2	22	6.7	37
TRP	164	HE1	-	W _G	4	OH2	21	2.4	98
GLY	319	O	-	W _s	81	H2	21	79.2	3
W _G	99	H1	-	W _G	28	OH2	21	2.9	82
W _G	5	OH2	-	W ₄	214	H2	21	6.5	37
THR	50	OG1	-	W ₄	220	H1	20	2.5	91
W _s	81	OH2	-	W ₃	5248	H2	20	10.2	22

TABLE 7.B. Hydrogen bonds between internal water molecules and amino acids in the heme a_3 / Cu_B region during 1125ps of molecular dynamics production for the ‘W8’ set of water molecules binding sites (with occupancy $\geq 20\%$) (all listed protein residues belong to subunit I).

Hydrogen bonds				Occup.	Aver.	Events			
				(%)	lifetime				
OH⁻		OH2	-	W_s	69	H2	100	1125.0	1
HIS	413	HN	-	VAL	409	O	97	13.7	80
HIS	411	HD1	-	VAL	408	O	92	9.3	111
ASP	399	HN	-	GLN	395	O	91	5.2	198
HIS	325	HE1	-	OH⁻		OH2	82	20.5	45
HIS	94	HN	-	MET	90	O	81	3.6	253
HIS	276	HN	-	TRP	272	O	75	3.5	243
HIS	325	HD1	-	W_G	158	OH2	74	13.7	61
HIS	413	O	-	SER	417	HN	67	2.3	327
HIS	326	HN	-	W₀	1385	OH2	66	3.3	222
HEMA3		OMA	-	W_G	158	H2	65	9.1	81
HEMA3		O1D	-	ARG	473	HH21	63	7.1	100
W₀	1385	OH2	-	W_G	158	H1	60	5.9	115
HIS	94	O	-	MET	98	HN	55	2.3	269
SER	394	HG1	-	W_G	491	OH2	54	6.7	90
HIS	276	O	-	TYR	280	HD2	53	1.9	306
ARG	473	HE	-	W_G	51	OH2	53	7.2	83
HIS	413	HD1	-	W_s	42	OH2	50	6.2	91
HEMA3		O2A	-	ASP	399	HD2	49	91.6	6
W_s	69	OH2	-	W_G	453	H1	47	4.2	124
HIS	276	O	-	TYR	280	HN	43	2.1	225
ARG	473	HE	-	W_s	13	OH2	42	8.7	55
HIS	326	HE1	-	W_G	453	OH2	39	3.0	146
HEMA3		O1D	-	ARG	473	HH12	37	8.9	47
OH⁻		OH2	-	W_G	453	H2	34	385.8	1
TYR	280	HH	-	W_G	188	OH2	33	6.9	54
ARG	473	HH12	-	W₄	178	OH2	29	11.5	28
W_G	177	OH2	-	W_G	189	H2	29	6.6	49
TYR	406	HH	-	W_G	46	OH2	26	9.7	30
SER	394	HG1	-	W_G	188	OH2	26	10.0	29
ARG	474	HH11	-	W_G	31	OH2	26	9.0	33
HEMA3		O2D	-	ARG	473	HH21	25	4.3	66
HEMA3		O2D	-	W₄	182	H2	25	11.3	25
HIS	413	HD1	-	W₄	219	OH2	23	7.8	34
HIS	325	HN	-	W_G	326	OH2	23	3.0	87
TYR	406	HH	-	W₄	224	OH2	22	3.8	65
TYR	406	HH	-	W₁	3872	OH2	22	8.1	30
HIS	276	HE1	-	W_s	69	OH2	22	2.7	90
GLY	390	HN	-	W_G	395	OH2	22	2.5	101
GLY	352	HN	-	W₂	6384	OH2	22	2.4	106
W_G	515	OH2	-	W_G	178	H1	22	6.0	41
HIS	411	HN	-	TYR	407	O	21	1.9	125
HIS	326	O	-	W₃	799	H2	21	14.2	17
ARG	473	HH11	-	W_G	51	OH2	21	2.3	106
HEMA3		OMA	-	W_G	158	H1	20	5.3	43
HEMA3		O1D	-	W₄	191	H1	20	6.4	35
HEMA		O11	-	W_G	48	H1	20	2.7	83
W_s	81	H2	-	W₂	7182	OH2	20	9.7	23

TABLE 8. The average number of water molecules at instantaneous snapshots in the K- and D- pathways of cytochrome *c* oxidase from *Paracoccus denitrificans*.

Time, ps	Average number of water molecules			
	K-path 'W12'	K-path 'W8'	D-path 'W12'	D-path 'ALL'
0				
45	8	15	23	29
225	8	5	20	25
450	6	10	18	27
675	9	5	22	24
900	4	9	22	27
1125	3	13	24	26

CHAPTER 5.

Molecular dynamics study of cytochrome *c* oxidase from *Paracoccus denitrificans* in different stages of catalytic cycle.

This chapter describes an attempt towards a more detailed understanding of the properties of COX in different states of catalytic cycle. The molecular dynamics study of the two subunit cytochrome *c* oxidase from *Paracoccus denitrificans* was carried out using the molecular modeling program CHARMM (Brooks et al., 1983). We used as a starting points the crystal structures of the fully oxidized enzyme (Ostermeier et al., 1997). In our simulation we had a look at the structural changes of cytochrome *c* oxidase during the dynamics and the the high mobility of internal water molecules. To frame our discussion of the results presented here we pursue two themes in more detail: the importance of the hydrogen bonding network and the possible coupling of local structural changes to large scale changes in the cytochrome *c* oxidase as seen, for example, in the simulation of rhodopsin (Crozier et al., 2003).

5.1 Details of the simulation protocol.

The general protocol for molecular dynamics (MD) simulations (Berneche et al., 1998; Berneche and Roux, 2000) to construct the initial configuration of a protein-membrane-water system (see CHAPTER 4 for more details) was followed. The starting configuration of the fully-oxidized two-subunit COX from *Paracoccus denitrificans* has been described in CHAPTER 2. The final microscopic system consisted of cytochrome *c* oxidase (2 subunits of 549 and 252 amino acids, respectively), 181 DMPC lipids (90 in the top and 91 in the bottom layer), 88 internal crystal waters W_S , and 24,323 bulk water molecules (W_0, W_1, W_2, W_3, W_4). Additionally, 755 internal water molecules ('W8' water molecules W_G) constructed using the GRID method were included in the calculations, and 42 Na^+ and 30 Cl^- ions were randomly inserted to simulate a 100 mM aqueous salt solution (ions were positioned more than 6 Å

away from the bilayer and protein, and no ion pairs were allowed to form). After solvation the entire system consisted of 103,589 atoms ('W8' set of coordinates).

Charges on the protein atoms and ionizable groups in different protonation states were taken from the CHARMM22 force field (MacKerell et al., 1995; MacKerell et al., 1998). Lys354 was assigned as neutral because of its hydrophobic environment in accordance with the results of electrostatic calculations (Kannt et al., 1998). Partial atomic charges for neutral Lys354 as well as for protonated Asp399, and for protonated Glu278 were taken from the AMBER96 force field. Because all titrating protons have been positioned in the structure, we can describe the deprotonation / protonation of the particular residue as the removal/addition of an explicit hydrogen atom.

The atomic partial charges for the redox centers – the heme *a* and heme *a*₃ sites, the His276 - Tyr280 crosslink and for Cu_B and its three histidine ligands (His276, His325 and His326) for the fully oxidized enzyme were obtained from quantum chemical calculations (see CHAPTER 2). The atomic partial charges for the heme *a* and heme *a*₃ sites for the reduced state were taken from CHARMM 22 force fields. The partial charge for the reduced Cu_B atom was obtained on the basis of quantum chemical calculations (TABLE 9). We did not carry out *pK_a* calculations for all titratable residues.

TABLE 9 . Partial charges for the heme *a* and heme *a*₃ sites and Cu_B sites.

Atom (PDB convention)	Reduced form (FeII)	Oxidized form (FeIII)
Fe	0.24	0.36
NA	-0.22	-0.10
NB	-0.22	-0.10
NC	-0.22	-0.10
ND	-0.22	-0.10
CHA		-0.10
HA		0.10
O1A		
O2A		
CHB		-0.10
HB		0.10
CHC		-0.10
HC		0.10
CHD		-0.10
HD		0.10
CGD		
O1D		
O2D		
Cu _B	0.39	1.39
NE2 (His411)	-0.54	-0.42
NE2 (His413, His97)	-0.27	-0.14

The partial charges do not change from the oxidized to the reduced state where no numbers are provided.

5.2 Results of the simulations.

In the simulations presented here for different redox states of cytochrome *c* oxidase we observed the structural changes of the enzyme and the high mobility of internal water molecules. Water molecules move in and out of the enzyme during the catalytic cycle and this has been shown to limit electron transfer (Kornblatt, 1998). More recent theoretical work indicates that water molecules might indeed provide proton transfer pathway from Glu278 to both the heme propionate and the binuclear center (Hofacker and Schulten, 1998; Zheng et al., 2003; Olkhova et al., 2003), consistent with the role of the D-pathway in both translocation and consumption of protons. It has been shown that the hydrogen bonded network in COX is not uniformly distributed (Olkhova et al., 2003), and that the degree of water arrangement is variable. Networks of hydrogen bonded water molecules are found that extend in many directions above the heme *a* and heme *a*₃ for every simulated variant (FIGURE 20 - 23). Some of the water molecules in the vicinity of the binuclear center have long lifetimes (as long as the whole time of the simulations), but in general the majority of internal water molecules is much more mobile. It can be concluded from the results of the simulations that the variation in the oxidation state (and thus charge) of the active sites and protonatable residues of the protein leads to changes of the electrostatic interaction energy between the latter and the water molecules. This effect could lead to redox-state dependent rearrangements of the water molecules and even in the direction of the hydrogen bonded network in the enzyme (similarly as published by Wikström et al. recently, 2003). In particular, the electric field between the hemes, parallel to the membrane, may orient the water molecules (dipoles) in the region between them to form a proton transfer path from Glu278 to the heme *a*₃ propionates, or to the binuclear center (Wikström et al., 2003). Another interesting result of this study is that the hydrogen bonded network in the D- and K-pathways re-orientates and thus a possible connection of these proton transfer pathways may be established during the catalytic cycle. We analysed both the local hydrogen bonded network in a region of the active site of protein, and the re-orientation of some functionally important residues in the D- and K- proton transfer pathways during the catalytic cycle.

At each simulated state hydrogen bonded networks in the region above the hemes are observed that remain strongly coupled throughout the simulation (FIGURE 20 -- 23). The results show that the conformation of the hemes is not stable under reduced conditions, even though the simulations are quite short (150 – 200 ps). Our main hypothesis to explain these deformations is that the heme potential is not good enough in its present form. Given the instability observed in the simulations it was decided to perform some quantum chemical

calculations to determine more accurate heme charges (work in progress) for the reduced state, using the ESP module of NWChem 4.1 employing the Ahlrich VTZ basis set for iron, and 6-311+G** basis set for the porphyrin and imidazoles (NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.1. (2002)).

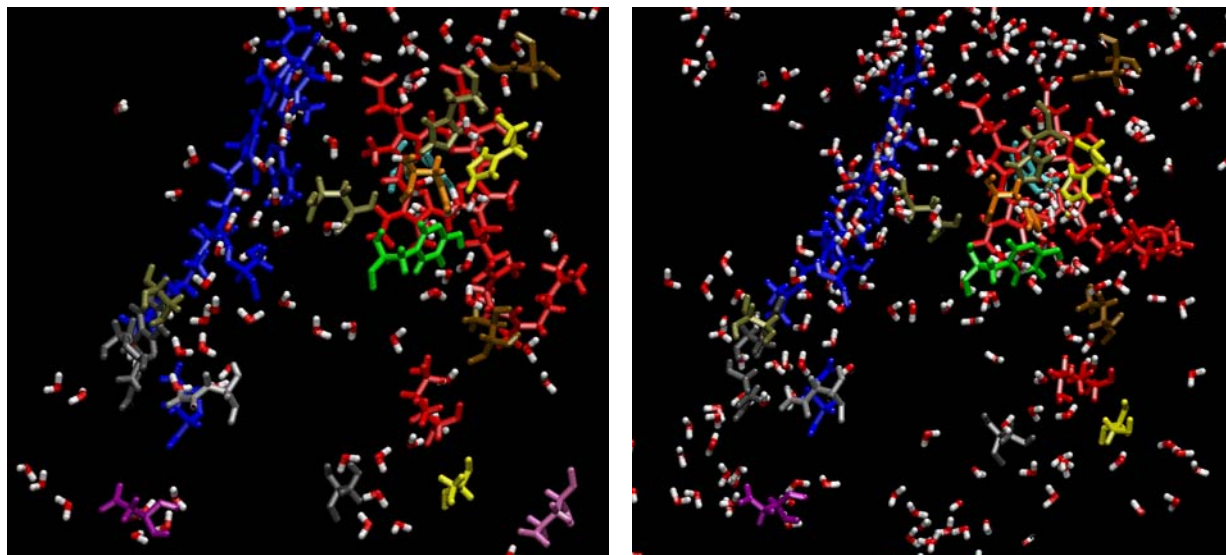


FIGURE 20 A and B . Cytochrome *c* oxidase – heme *a* oxidized, heme *a*₃ oxidized. For the D-pathway Asp124 is represented in *magenta*, Asn199 in *grey*, Asn113 in *blue*, Asn131 in *dark grey*, Tyr35 in *grey*, Ser134 in *gold*, Ser193 in *dark blue*, Glu278 in *gold*. For the K-pathway Lys354 is represented in *red*, Thr in *brown*, Tyr280 in *green*, and His276 in *orange*.

We visually analyzed the systematic mapping of the hydrogen bonded networks in the region close to the active site of the enzyme. While comparing these maps in the various structures, we notice the persistent residence of water molecules at the active site and dedicated to define their positions and assess their possible role in the enzyme's function.

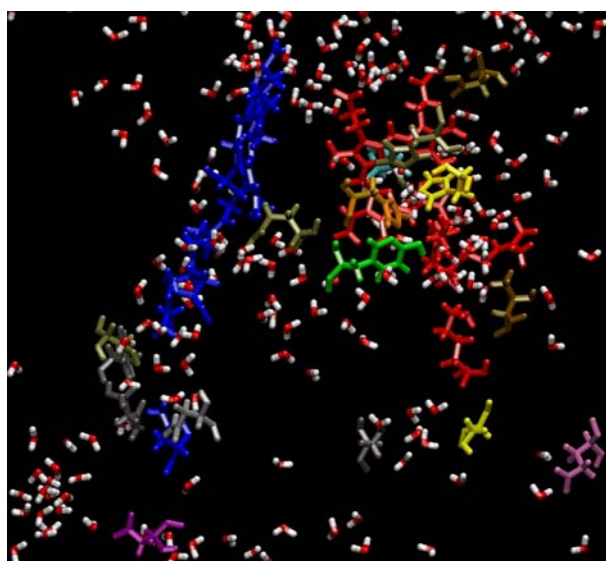


FIGURE 21. Cytochrome *c* oxidase – heme *a* reduced, heme *a*₃ oxidized. For the D-pathway Asp124 is represented in *magenta*, Asn199 in *grey*, Asn113 in *blue*, Asn131 in *dark grey*, Tyr35 in *grey*, Ser134 in *gold*, Ser193 in *dark blue*, Glu278 in *gold*. For the K-pathway Lys354 is represented in *red*, Thr in *brown*, Tyr280 in *green*, and His276 in *orange*.

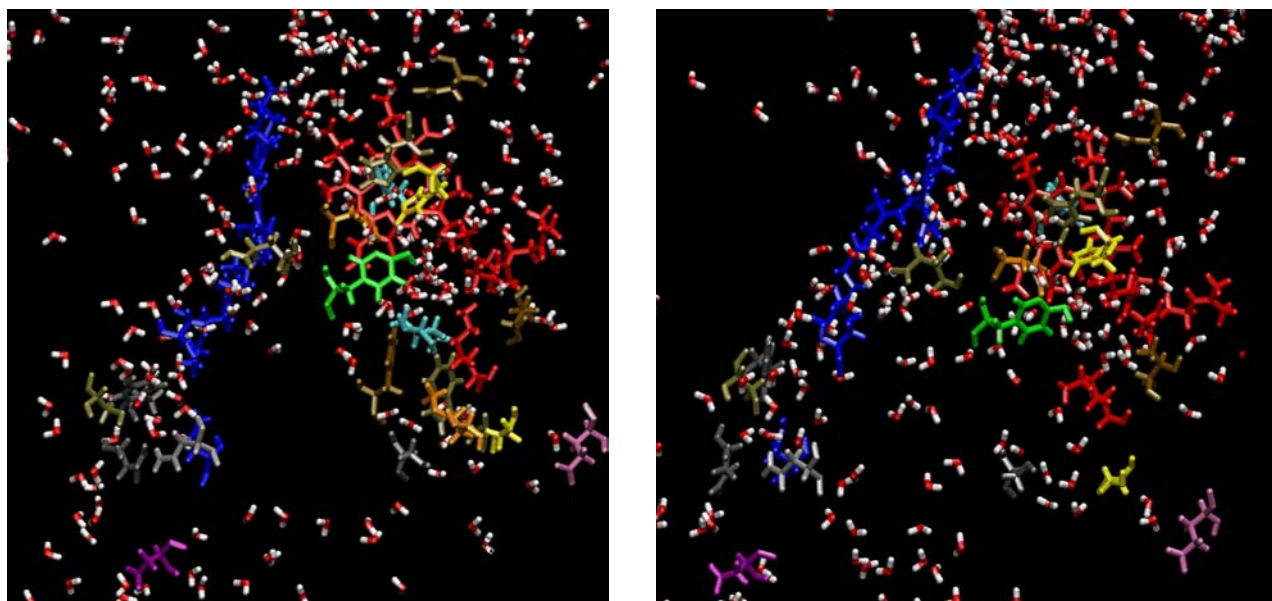


FIGURE 22 A and B. Cytochrome c oxidase – heme a oxidized, heme a_3 reduced. For the D-pathway Asp124 is represented in *magenta*, Asn199 in *grey*, Asn113 in *blue*, Asn131 in *dark grey*, Tyr35 in *grey*, Ser134 in *gold*, Ser193 in *dark blue*, Glu278 in *gold*. For the K-pathway Lys354 is represented in *red*, Thr in *brown*, Tyr280 in *green*, His276 in *orange*, Phe271 in *gold*, Met292 in *orange*, and Ile in *light blue*.

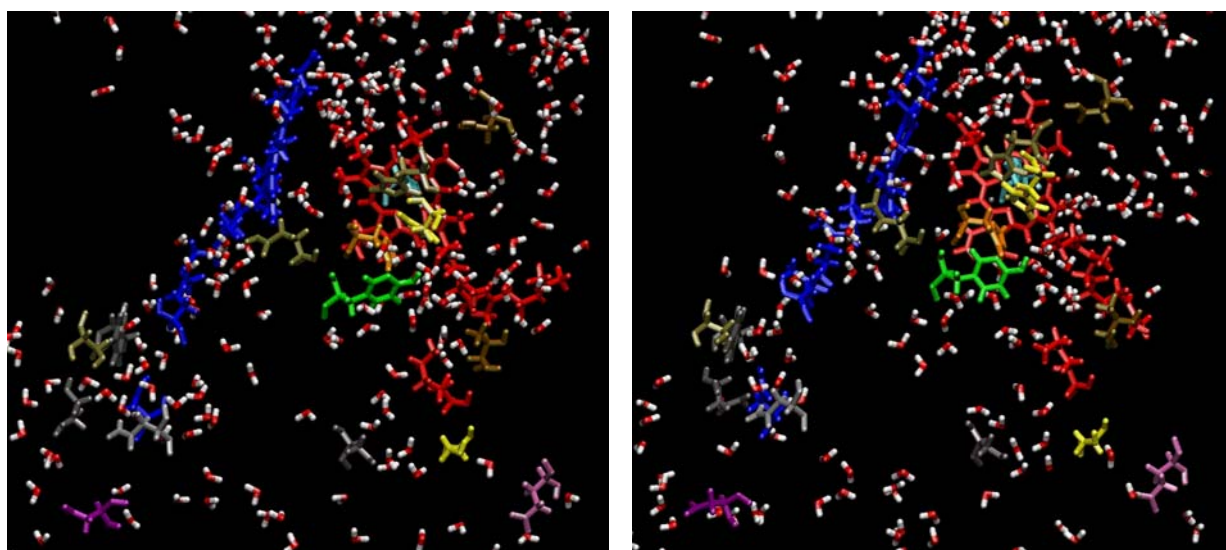


FIGURE 23 A and B. Cytochrome c oxidase – heme a reduced, heme a_3 reduced, Cu_B reduced. For the D-pathway Asp124 is represented in *magenta*, Asn199 in *grey*, Asn113 in *blue*, Asn131 in *dark grey*, Tyr35 in *grey*, Ser134 in *gold*, Ser193 in *dark blue*, Glu278 in *gold*. For the K-pathway Lys354 is represented in *red*, Thr in *brown*, Tyr280 in *green*, and His276 in *orange*.

The resulting analysis of all investigated redox states leads us to conclude that there are four water molecules in the region between Ser193 and Glu278 that remain in their positions during 150 ps of simulation time, which constitutes stable hydrogen bonding. For the fully oxidized enzyme we found short lived hydrogen bonding chain of water molecules which may connect the D- and K- proton transfer pathways, in addition to the results which are presented in CHAPTER 4. It will be very interesting to see whether similar possibilities

exist for other redox states of the enzyme, since this observation may be crucial for the understanding the mechanism of proton pumping in COX (work in progress).

Our results for the fully oxidized enzyme do not show a connection between the protonated Glu278 and the O1A propionate group of heme a_3 . Glu278 is one of the most conserved residues among the heme-copper oxidases, and mutations at this position have been shown to block the oxygen reaction and the uptake of protons, as well as proton translocation (Svensson-Ek et al., 1996), suggesting that this residue plays an important role in the catalytic mechanism and that it is a key residue for proton translocation (reviewed by Michel, 1998). FTIR data (Hellwig et al., 1998; Puustinen et al., 1997) has provided evidence for some kind of connection between Glu278 and Cu_B , possibly through a hydrogen-bonded water chain, although the crystal structures do not reveal any bound water molecules in this area. Proton transfer from Glu278 forward might involve conformational changes of the side chain of this amino acid. It could thus act as a switch (Kannt and Michel, 2001) transferring protons either to the binuclear center or towards a region surrounding the heme propionates (Behr et al., 1998) from which they could be expelled into the periplasmic phase. In agreement with this proposal, our additional short simulation of COX with deprotonated Glu278 shows a significant reorientation of the sidechain of Glu278 and the formation of a hydrogen bond chain between Glu278 via water molecules up to the O1A atom of the heme a_3 propionate. This is why we produced two types of simulations for all redox states – one with protonated Glu278, and other with deprotonated Glu278 (see FIGURE 21 – 23, A – protonated Glu278, B – deprotonated Glu278). For all simulated redox states with deprotonated Glu278 we observed the reorientation of the sidechain of Glu278 in the direction of heme a_3 . For the next oxidation state, when heme a is reduced and the binuclear center is oxidized, we observed a significant reorientation of the side chain of Lys354 and, further, a brief connection with Tyr280 via a chain of water molecules. The electron transfer from heme a to the binuclear center inverts the electric field between the redox sites. This switch orients several water molecules for proton transfer from Glu278 to the heme a_3 propionate O1D and provides an additional stable hydrogen bonded chain of water molecules in the vicinity of the K-pathway. The residues which are in the closest neighbourhood to this chain are Phe271, Met292, and Ile299. Possibly, this additional hydrogen bonding network could play an important role for the proton transfer in COX.

Certainly there is a great wealth of information in the simulations and we cannot yet analyze, or even present, all possible events. The MD simulations provide initial insights into the coupling between local changes in the active sites of the cytochrome c oxidase and

conformational changes in the whole protein, in particular, large conformational changes in the hydrogen bonded networks during the catalytic cycle.

Overall Conclusions and Future Work.

Conclusions

This thesis describes computer simulations of cytochrome *c* oxidase from *Paracoccus denitrificans*. The combination of different techniques presented here yielded some compelling findings, which may allow a better interpretation of the experimental observations, and provide some insights into the mechanism of proton translocation in such a complex system. This study can also function as a guide towards an understanding why simple theoretical approaches **do** or **do not** work, and lead to the development of some new and more accurate models.

The main results of this thesis can be summarized as follows:

Methods based on a single protein structure (usually determined by X-ray crystallography) characterize the protonation / deprotonation equilibrium correctly only in the limited pH interval for which the corresponding structure is relevant. As far as the change in average protonation of the protein depends primarily on the orientation of the side chain of one single amino acid (Lys-II-191), local flexibility can be considered as the main origin of pK changes. In another point of view, upon changes of the charge state (change of pH), conformational changes may occur.

Since the presence of internal water molecules and local and global fluctuations of the residues may be associated with redox changes in the protein, a number of molecular dynamic simulations (CHAPTER 4 and CHAPTER 5) and internal water molecule predictions (CHAPTER 2) have been performed. A much higher average number of internal water molecules than observed in crystal structures (Iwata et al., 1995; Ostermeier et al., 1997; Tsukihara et al., 1996; Svensson-Ek et al., 2002) was found in our simulations.

The dynamic model discussed here led to a stable system. The MD simulation required about 1 ns to reach a dynamic equilibrium, due to the effect of protein – membrane and protein-water interactions. Also, the water-protein interactions took a long time to reach

equilibrium. Similar observations were reported for other systems (Garcia and Hummer, 2000; Bret et al., 2002). The hydrogen-bonded network in cytochrome *c* oxidase is not uniformly distributed, and the water arrangement is variable. The protein – membrane – water model provides a detailed description of the structure and dynamics of a hydrogen bonded network and identified a number of permanent water molecules in the K- and D- pathways. Networks of hydrogen bonded water molecules are found that extend in many directions above the heme *a* and heme *a*₃. Some of the water molecules in the vicinity of the binuclear center have lifetimes around 1 ns, but in general the majority of internal water molecules are much more mobile. The presence of mobile water is indicated in regions of the protein where no water has been found by X-ray crystallography. The explanation for the large discrepancy in the number of internal water molecules as found during simulation and as compared to the crystallographically identified internal water molecules is the high mobility and exchange rate of water molecules between internal and external locations.

Although our simulations agree with previous theoretical studies in general (Riistama et al., 1997; Hofacker and Schulten, 1998; Backgren et al., 2000; Zheng et al., 2003), there are some differences. The main one concerns the significant diffusion of individual water molecules in the D- and K-pathways as well as in the region connecting them to the binuclear center (Olkhova et al., 2003). Proton transfer along these pathways may be different in character than that along narrow water files as described in gramicidin A (Pomès and Roux, 1996) and carbon nanotubes (Hummer et al., 2001). We initially planned to identify a small number of unique hydrogen bond networks in the D- and K-pathways as described, for example, for crystals of vitamin B₁₂ (Savage, 1986). However, the hydrogen-bonded network in cytochrome *c* oxidase is so dynamic and of such a high dimensionality that it cannot be accurately characterized only by pictures and tables. Each hydrogen-bonded pathway is transient and exists only for a certain period of time. Proton flow should be thought of as using of many fluctuating alternative transport pathways, which are the result of a reorientation and exchange of the participating water molecules instead of a single hydrogen bonded connection. Nevertheless, a driving force has to provide directionality. The MD simulations provide initial insights into the coupling between local changes in the active sites of the cytochrome *c* oxidase and large conformational changes in the whole protein, in particular, conformational changes in the hydrogen bonded network, during the catalytic cycle.

Outlook

A few issues could not be completely solved in the framework of this study and should be the topic of future research:

First, from a technical point of view, different simulation protocols and membrane systems could be tested. Regarding the results of this work, it must be said that the stability of the membrane bilayer has large errors. However, the origin of the unstable membrane is currently unclear. Our system (~ 100.000 atoms) is much larger than other systems used for simulations of membrane proteins like bacteriorhodopsin, the photosynthetic reaction center or the KcsA channel. Therefore, longer equilibration times for lipid phase and added ions etc. may be required before allowing the protein to move. Our system is the first simulation of an odd-shaped membrane protein that substantially sticks out of the membrane. Therefore, we may be lacking the typical stabilization of parallel lipid bilayers by our use of somewhat artificial periodic boundary conditions. We also feel that the combination of CHARMM force field and Ewald long-range electrostatics needs to be evaluated more thoroughly for systems of this size. However, due to the large computational costs involved, such studies are just beginning to become possible with the increasing power of new computers.

Second, the dependence of the protonation / deprotonation equilibria of titratable groups on their conformational states, which is a well-documented effect, appears to be one of the major problems for the accurate prediction of ionisation equilibria in proteins. Modern macroscopic approaches for pK calculations differ mostly in how the conformational heterogeneity is treated. In different approaches, structural changes rank from involving only hydrogens (Bashford et al., 1993; Alexov and Gunner, 1997), through side-chain fluctuations (Alexov and Gunner, 1999; You and Bashford, 1995; Beroza and Case, 1996), to accounting for global flexibility by applying MD (van Vlijmen et al., 1998; Bashford and Gerwert, 1992) or MC dynamics.

Third, on the basis of the MD simulations results it has been found that the heme potential in its present form is not sufficient for the stability of the atomic system. Given the instability observed in the simulations it was decided to perform some quantum chemical calculations to obtain more accurate heme charges for the reduced state (Smith et al., 2003), using the ESP module of NWChem 4.5 (NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.5. (2002)).

Finally, the complex coupling of several electron and proton transfer steps must be elucidated in atomic and electronic detail. In general, molecular dynamics simulations of

biomolecular systems are performed for a fixed protonation state. In the future we plan to include flexible protonation to allow proton hopping between titratable states. These calculations will be done using the new software Q-HOP (Lill and Helms, 2000) which includes stochastic proton hopping events. Hopping probabilities between typical biochemical donor-acceptor systems are pre-calculated and parameterized against quantum-mechanical calculations. The method is ideally suited for studying complex dynamic proton transfer situations in large biomolecular systems. Applications of the method include the proton shuttle between chromophore and surrounding amino acids in green fluorescent protein (Lill and Helms, 2001), and proton diffusion through aquaporin (De Groot et al. 2003).

Appendix A. PROGRAMS.

A.1 Molecular Discovery Programmes.

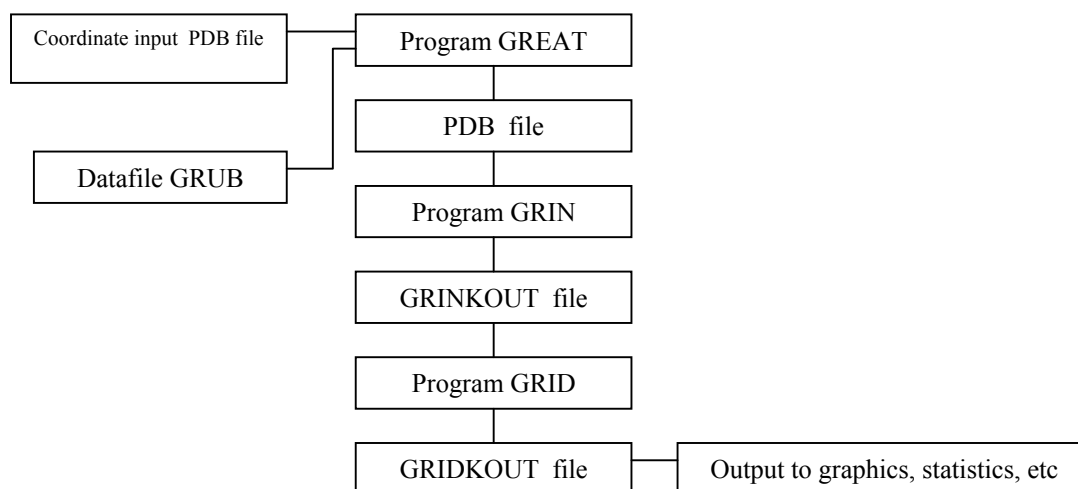
The Molecular Discovery Programmes are used in order to predict specific noncovalent interactions between a molecule of known three dimensional structure (the ‘target’), and a small chemical group (the ‘probe’) whose properties are defined by user.

A wide range of Targets may be studied including enzymes, nucleic acids, poly-saccharides, glycoproteins, peptides, membranes, crystals, drug molecules and many others.

A.1.1 GRIN: Program GRIN is used in order to prepare the input for the main program GRID. The primary input for GRIN is the xyz positions of the atoms in the target molecule. These coordinates must be provided in the recognised format specified for the PDB data bank. The secondary input for GRIN is a set of ‘energy variables’ which are contained in a datafile called GRUB. These energy variables must be correctly specified for every atom in the target molecule, in order to define how the target will interact with the probe.

A.1.2 GRID: GRID is the main program, and the overall GRID method was developed by P. Goodford (1985). The GRID user takes GRINKOUT as an input file which defines the properties of the target. However, it is necessary to decide on the chemical properties of the probe. This information is supplied to programm GRID which then computes the places on the target at which the probe would interact favourable.

The general relationships between files and programmes are shown in diagram below:

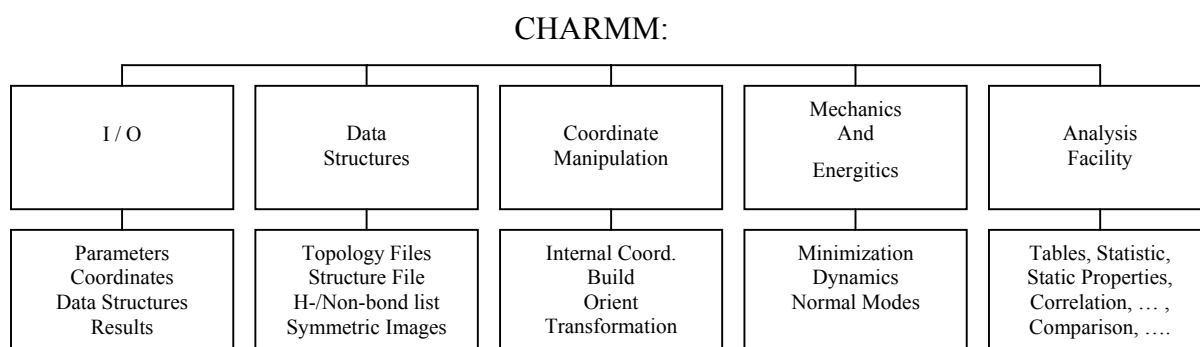


Online information about Molecular Discovery Programmes is available from:

http://www.moldiscovery.com/soft_grid.php

A.2 CHARMM

CHARMM is a program for simulating biologically relevant macromolecules (proteins, DNA, RNA) and complexes thereof (2). It allows to investigate the structure and dynamics of large molecules (solutes) in the condensed phase (solvent or crystal). CHARMM can be used to calculate free energy differences upon mutations or ligand binding (12). Moreover, it has been used to simulate the reversible folding of structured peptides (5,7) and determine folding free energy surfaces (1,6). It uses classical mechanical methods to investigate potential energy surfaces derived from experimental and “ab initio” quantum chemical calculations (14). Furthermore, mixed quantum mechanical/classical systems can be defined to investigate chemical processes such as enzyme catalysis. One of the most common application of CHARMM is molecular dynamics (MD), in which the Newton equation of motion is discretized and solved by an integration procedure (Verlet algorithm). The force on the atoms is the negative gradient of the CHARMM potential energy (14). The organization of CHARMM is given in the scheme below:



For the execution of the CHARMM code we used the Regatta SP4 IBM RS6000 supercomputer (RZG, MPG). One cluster is formed by 32 nodes. On every node there are 32 processors with 400 MHz clock. The cache memory for the CPUs is 512 kB. The random access memory of each node is 256 MB. Every node has a SCSI hard disk drive with a capacity of ca. 9 GB.

Mainly, two source files were modified in order to parallelized CHARMM c28b2 on IBM64 platform. The file machio.src has now entries for the record length in the open statements for IBMSPP. This was necessary since there does not seem to be a sensible install value (which was 9223372036854775806).

Another problem was due to the use of definition of the preprocessor symbol RCFFT, which enables sequential Fortran code, which then screws up the communication. The file install.com does not contain anymore the RCFFT option. The programs believe after initialisation, that only 1 CPU is available, regardless of how many have been found

previously. One can either use the option/proprocessor symbol RCFFT of PARALLEL/PARAFULL, but must not combine them. It is necessary in `parall.src` file to introduce the function PSYNC. The Makefile must be changed to `AR = ar -X64`.

A proper understanding of the performance of CHARMM requires a systematic workload characterization of the code. The next scheme shows the scalability test (elapsed time and CPU time) of the classical MD simulations for 100 steps:

ELAPSED TIME	Number of processors (PEs)	Time (min)	CPU TIME	Number of processors (PEs)	Time (min)
	1	42.07		1	40.30
	2	23.53		2	21.36
	4	16.18		4	11.48
	8	11.77		8	8.09
	16	8.23		16	6.43

Online information about CHARMM (Harvard) is available from: <http://yuri.harvard.edu>.

A.3 DelPhi

DelPhi is an electrostatics program which can be used to generate maps of electrostatic potentials on PDB files.

Online information about DelPhi is available from: <http://bcr.musc.edu/manuals/delphi.htm>

A.4 GRASP

GRASP is an electrostatics program related to DELPHI which can be used to generate images of electrostatic potentials from PDB files. GRASP calculates a less exact electrostatic potential than does DELPHI, but it can be induced to use DELPHI so that a more exact representation results.

Online information about GRASP is available from: <http://trantor.bioc.columbia.edu/grasp>

A.4 InsightII

Insight II is a 3D graphical environment for molecular modeling. Its powerful user interface enables the seamless flow of data between a wide range of scientific applications. The Insight II environment integrates builder modules, development tools, force fields, simulation and visualization tools with tools specifically developed for applications in the life

and materials sciences. Insight II runs on Silicon Graphics workstations and servers. Online information about InsightII (Accelrys) is available from: <http://www.accelrys.com/insight/>

A.5 VMD

Visual Molecular Dynamic Software VMD 1.8 (Humphrey et al., 1998) is a molecular graphics program designed for the interactive visualization and analysis of the biopolymers such as proteins, nucleic acids, lipids, and membranes. Online information about VMD is available from: <http://www.ks.uiuc.edu/Research/vmd>.

List of key VMD features:

- general molecular visualization
- visualization of dynamic molecular data
- interactive molecular dynamics simulations
- molecular analysis commands
- Tcl and Python scripting languages
- support for multimodal input and various display systems

REFERENCES

- Aagaard, A., G. Gilderson, D. A. Mills, S. Ferguson-Miller, P. Brzezinski. 2000. Redesign of the proton-pumping machinery of cytochrome *c* oxidase: proton pumping does not require Glu(I-286). *Biochemistry* 39(51):15847-15850
- Alard, P., S. J. Wodak. 1991. Detection of cavities in a set of interpenetrating spheres. *J. Comp. Chem.* 12:918-922
- Alexov, E., and M. Gunner. 1997. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys J.* 72:2075-2093
- Alexov, E., and M. Gunner. 1999. Calculated protein and proton motion coupled to electron transfer: Electron transfer from Q_A-Q_B to Q_B in bacterial photosynthetic reaction centers. *Biochemistry* 38:8253-8270
- Allen, M. P., D. J. Tildesley. 1987. *Computer Simulation of Liquids*. Clarendon Press, Oxford
- Allen, T. W., A. Bilznyk, A. P. Rendell, S. Kuyucak, S. H. Ching. 2000. The potassium channel: structure, selectivity and diffusion. *J. Chem. Phys.*, 112:8191-8204
- Andersen, H. C., C. B. Berde, B. S. Hudson. 1980. A theory of the effects of head-group structure and chain unsaturation on the chain melting transition of phospholipid dispersions. *Biochemistry* 19(18):4279-93
- Åqvist, J., and V. Luzkov. 2000. Ion permeation mechanism of the potassium channel. *Nature* 404:881-884
- Babcock, G. T., M. Wikström. 1992. Oxygen activation and the conservation of energy in cell respiration. *Nature* 356(6367):301-309
- Backgren, C., G. Hummer, M. Wikström, A. Puustinen. 2000. Proton translocation by cytochrome *c* oxidase can take place without the conserved glutamic acid in subunit I. *Biochemistry* 39(27):7863-7867
- Bakowies, D., W. F. van Gunsteren. 2002. Water in protein cavities: A procedure to identify internal water and exchange pathways and application to fatty acid-binding protein. *Proteins: Structure, Function, and Genetics* 47:534-545
- Banci, L., I. Bestini, C. Luchinat, P. Turano. 2000. Solution Structures of Hemoproteins. in *The porphyrin Handbook* (Kadish K.M., Smith K.M. and Guillard R., Eds) Vol. 5, pp. 323-350, Academic Press, San Diego, CA

- Bashford, D., K. Gerwert. 1992. Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. *Biophys J.* 66:1341-1352
- Bashford, D., D. A. Case, C. Dalvit, L. Tennant, and P. E. Wright. 1993. Electrostatic calculations of side-chain pK_a values in myoglobin and comparison with NMR data for histidine. *Biochemistry* 32:8045-8056
- Battistuzzi, G., M. Corsari, M. Sola, F. Francia. 1997. Redox thermodynamics of the native and alkaline forms of eukaryotic and bacterial class I cytochromes *c*. *Biochemistry* 36:16247-16258
- Baudry, J., E. Tajkhorshid, F. Molnar, J. Phillips, K. Schulten. 2001. Molecular dynamics study of bacteriorhodopsin and the purple membrane. *J. Phys. Chem. B*, 105:905-918
- Ben-Naim, A., Y. Marcus. 1984. Solvation thermodynamics of nonionic solutes. *J Chem Phys.* 81:2016-2027
- Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684-3690
- Berneche, S., M. Nina, and B. Roux. 1998. Molecular dynamics simulation of melittin in a dimyristoylphosphatidylcholine bilayer membrane. *Biophys. J.* 75:1603-1618
- Berneche, S., and B. Roux. 2000. Molecular dynamics of the KcsA K⁺ channel in a bilayer membrane. *Biophys. J.* 78:2900-2917
- Berneche, S., and B. Roux. 2001. Energetics of ion conduction through the K⁺ channel. *Nature* 414:73-77
- Berosa, P., D. A. Case. 1996. Includine side chain flexibility in continuum electrostatic calculations of protein titration. *J Phys Chem* 100:20156-20163
- Bizzarri, A. R., and Cannistrato S. 2002. Molecular dynamics of water at the protein-solvent interface. *J. Phys. Chem. B*, 106:6617-6633
- Boobbyer, D. N., P. J. Goodford, P. M. McWhinnie, R. C. Wade. 1989. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.* 32:1083-1094
- Bret, S., M. Roth, S. Nørager, E. C. Hatchikian, and M. J. Field. 2002. Molecular dynamics study of *Desulfovibrio africanus* cytochrome *c*₃ in oxidized and reduced forms. *Biophys. J.* 83:3049-3065
- Brooks, C. L. I., M. Kaplun, B. M. Pettitt. 1988. Proteins: A theoretical perspective of dynamics, structure, and thermodynamics. John Wiley and Sons, New York

- Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* 4:187-217
- Brünger, A., M. Karplus. 1988. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins* 4(2):148-156
- Brzezinski, P., P. Adelroth. 1998. Proton-controlled electron transfer in cytochrome c oxidase: functional role of the pathways through Glu 286 and Lys 362. *Acta Physiol. Scand. Suppl.* 643:7-16
- Brzezinski, P., and G. Larsson. 2003. Redox-driven proton pumping by heme-copper oxidases. *Biochim. Biophys. Acta* 1605(1-3):1-13
- Bushnell, G. W., G. V. Louie, G. D. Brayer. 1990. High-resolution 3-dimensional structure of horse heart cytochrome c. *J. Mol. Biol.* 214:585-595
- Cai, W., M. Zhang, B. Maigret. 1998. New approach for representation of molecular surface. *J. Comp. Chem.* 19:1805-1815
- Carugo, O., and D. Bordo. 1999. How many water molecules can be detected by protein crystallography? *Acta Crystallogr.* D55:479-483
- Connolly, M. L. 1983. Analytical molecular surface calculation. *J. Appl. Crystallogr.* 16:548-558
- Crozier, P. S., M. J. Stevens, L. R. Forrest, T. B. Woolf. 2003. Molecular dynamics simulation of dark-adapted rhodopsin in an explicit membrane bilayer: coupling between local retinal and larger scale conformational change. *J Mol Biol*, 333(3):493-514
- De Groot, B. L., and H. Grubmüller. 2001. Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science* 294:2354-2357
- Delaney, J. S. 1992. Finding and filling protein cavities using cellular login operations. *J. Mol. Graphics* 10:174-177
- Denisov, V. P., B. Halle. 1995. Hydrogen exchange and protein hydration: the deuteron spin relaxation dispersions of bovine pancreatic trypsin inhibitor and ubiquitin. *J Mol Biol.* 245(5):698-709
- Denisov V. P., B. Halle. 1996. Protein hydration dynamics in aqueous solution. *Faraday discuss.* 103:227-244
- Denisov V. P., B. H. Jonsson, B. Halle. 1999. Hydration of denatured and molten globule proteins. *Nat. Struct. Biol.* 6(3):253-60

- Douliez, J. P., A. Leonard, E. J. Dufourc. 1996. Conformational order of DMPC sn-1 versus sn-2 chains and membrane thickness: An approach to molecular protrusion by solid state H²-NMR and neutron diffraction. *J. Phys. Chem.* 100:18450-18457
- Dunitz, J. D. 1994. The entropic cost of bound water in crystals and biomolecules. *Science* 264:670
- Elmore, D. E., and D. A. Dougherty. 2001. Molecular dynamics simulations of wild type and mutant forms of the Mycobacterium tuberculosis MscL channel. *Biophys. J.* 81:1345-1359
- Eriksson, M., T. Härd, and L. Nilsson. 1995a. Molecular dynamics simulations of the glucocorticoid receptor DNA-binding domain in complex with DNA and free in solution. *Biophys. J.* 68:402-426
- Eriksson, M., and L. Nilsson. 1995b. Structure, thermodynamics and cooperativity of glucocorticoid receptor DNA-DNA-binding domain in complex with different response elements: molecular dynamics and free energy perturbation study. *J. Mol. Biol.* 253:453-472
- Ernst, J. A., R. T. Clubb, H. X. Zhou, A. M. Gronenborn, G. M. Clore. 1995. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science* 267:1813-1817
- Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577-8593
- Feller, S. E., R. M. Venable, and R. W. Pastor. 1997. Computer simulation of a DPPC phospholipid bilayer. Structural changes as a function of molecular surface area. *Langmuir.* 13:6555-6561
- Feller, S. E., Y. H. Zhang, R. W. Pastor, and B. R. Brooks. 1995. Constant pressure molecular dynamics simulation – the Langevin piston method. *J. Chem. Phys.* 103:4613-4621
- Ferguson-Miller, S., G. T. Babcock. 1996. Heme/copper terminal oxidases. *Chem Rev.* 96(7):2889-2908
- Fetter, J. R., J. Quian, J. Shapleigh, J. W. Thomas, J. A. Garcia-Horsman, E. Schmidt, J. Hosler, G. T. Babcock, R. B. Gennis, and S. Ferguson-Miller. 1995. Possible proton relay pathways in cytochrome *c* oxidase. *Proc. Natl. Acad. Sci. USA* 92:1604-1608
- Finney, J. L. 1977. The organization and function of water in protein crystals. *Philos Trans R Soc Lond B Biol Sci.* 278(959):3-32
- Gao, X., and T. C. Wong. 2001. Molecular dynamics simulation of adrenocorticotropin (1-10) peptide in a solvated dodecylphosphocholine micelle. *Biopolymers* 58:643-659

- Garcia-Horsman, J. A., A. Puustinen, R. B. Gennis, and M. Wikström. 1995. Proton transfer in cytochrome *bo*₃ ubiquinol oxidase of *Escherichia coli*: second-site mutations in subunit I that restore proton pumping in the mutant Asp-135 → Asn. *Biochemistry* 34:4428-4433
- Garcia, A. E., and G. Hummer. 2000. Water penetration and escape in proteins. *Proteins* 38:261-272
- Gennis, R. B. 1998. How does cytochrome oxidase pump protons? *Proc Natl Acad Sci U S A* 95(22):12747-9
- Gibson, Q. H., and C. Greenwood. 1963. Reactions of cytochrome oxidase with oxygen and carbon monoxide. *Biochem J.* 86:541-54
- Goodford, P. J. 1985. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28:849-857
- Greenwood, C., and Q. H. Gibson. 1967. The reaction of reduced cytochrome C oxidase with oxygen. *J Biol Chem* 242(8):1782-7
- Halle, B., H. Johannesson, K. Venu. 1998. Model-free analysis of stretched relaxation dispersions. *J Magn Reson.* 135(1):1-13
- Hayashi, S., E. Tajkhorshid, K. Schulten. 2002. Structural changes during the formation of early intermediates in the bacteriorhodopsin photocycle. *Biophys. J.* 83(3):1281-97
- Helms, V., and R. C. Wade. 1995. Thermodynamics of water mediating protein-ligand interactions in cytochrome P450cam: a molecular dynamics study. *Biophys. J.* 69(3):810-824
- Helms, V., and R. C. Wade. 1998. Hydration energy landscape of the active site cavity in cytochrome P450cam. *Proteins* 32:381-396
- Henchman, R. H., and J. A. McCammon. 2002. Structural and dynamic properties of water around acetylcholinesterase. *Protein Science* 11:2080-2090
- Hermans, J., S. Shankar. 1986. The free energy of xenon binding to myoglobin from molecular dynamics simulations. *Isr. J. Chem.* 27:225-227
- Hockney, R. W. 1970. The potential calculation and some applications. *Methods in Computat. Physics*, 9:135-211
- Hofacker, I., and K. Schulten. 1998. Oxygen and proton pathways in cytochrome *c* oxidase. *Proteins* 30(1):100-107
- Hoover, W.G. 1985. Canonical dynamics-equilibrium phase-space distribution. *Phys. Rev. A*, 31:1695-1697
- Garcia, A. E., G. Hummer, D. M. Soumpasis. 1996. Theoretical description of biomolecular hydration. Application to A-DNA. *Basic Life Sci.* 64:299-308

- Hummer, G., J. C. Rasaiah, and J. P. Noworyta. 2001. Water conduction through the hydrophobic channel of a carbon nanotube. *Nature* 414:188-190
- Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33-38
- Iwata, S., C. Ostermeier, B. Ludwig, H. Michel. 1995. Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature* 376:660-669
- Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926-935
- Kannt, A., C. R. D. Lancaster, and H. Michel. 1998. The coupling of electron transfer and proton translocation: electrostatic calculations on *Paracoccus denitrificans* cytochrome *c* oxidase. *Biophys. J.* 74:708-721
- Kitagawa, T., and T. Ogura. 1998. Time-resolved resonance Raman investigation of oxygen reduction mechanism of bovine cytochrome *c* oxidase. *J Bioenerg Biomembr.* 30(1):71-9
- Konstantinov, A. A., S. Siletsky, D. Mitchell, A. Kaulen, R. B. Gennis. 1997. *Proc. Natl. Acad. Sci. USA* 1997. 94(17):9085-90
- Kornblatt, J. A. 1998. The water channel of cytochrome *c* oxidase: interference from inhibitor studies. *Biophys. J.* 75:3127-3134
- Ladbury, J. E. 1996. Just add water! The effect of water on the specificity of protein–ligand binding sites and its potential application to drug design. *Chem Biol.* 3(12):973-80
- Lancaster, C. R. D., H. Michel. 1997. The coupling of light-induced electron transfer and proton uptake as derived from crystal structures of reaction centres from *Rhodospseudomonas viridis* modified at the binding site of the secondary quinone, Q_B. *Structure* 5:1339-1359
- Langen, R., G. D. Brayer, A. M. Berghuis, G. L. McLendon, F. Sherman, A. Warshel. 1992. Effect of the Asn52----Ile mutation on the redox potential of yeast cytochrome *c*. Theory and experiment. *J.Mol.Biol.* 224:589-600
- Leach, A. R. 1996. Molecular modeling: principles and applications. Pearson Education Limited, Harlow.
- Le Coutre, J., J. Tittor, D. Oesterhelt, K. Gerwert. 1995. Experimental evidence for hydrogen-bonded network proton transfer in bacteriorhodopsin shown by Fourier-transform infrared spectroscopy using azide as catalyst. *Proc. Natl. Acad. Sci. USA* 92(11):4962-6
- Lee, B., F. M. Richards. 1971. The interpretation of protein structure: estimation of static accessibility. *J. Mol. Biol.* 55:379-400

- Levitt, D. G., L. J. Banaszak. 1992. Pocket: A computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J. Mol. Graphics* 10:229-234
- Liang, J., H. Edelsbrunner, P. Fu, P. V. Sudhakar, S. Subramaniam. 1998. Analytical shape computation of macromolecules. I. Molecular area and volume through alpha shape. *Proteins* 33:1-17
- Liang, J., H. Edelsbrunner, C. Woodward. 1998. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* 7:184-1897
- Luecke, H. 2000. Atomic resolution structures of bacteriorhodopsin photocycle intermediates: the role of discrete water molecules in the function of this light-driven ion pump. *Biochim. Biophys. Acta* 1460:133-156
- McCammon, J. A., B. R. Gelin, M. Karplus. 1977. Dynamics of folded proteins. *Nature* 267:585-590
- MacKerell, Jr., A., J. Wiorkiewicz-Kuczera, and M. Karplus. 1995. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* 117:11946-11975
- MacKerell, A. D., Jr., D. Bashford, M. Bellot, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, D. Joseph-McCarthy, S. Ha, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586-3616
- Marrink, S. J., D. P. Tieleman, A. E. Mark. 2000. Molecular dynamics simulation of the kinetics of spontaneous micelle formation. *J. Phys. Chem. B.* 104:12165-12173
- Michel, H. 1998. The mechanism of proton pumping by cytochrome *c* oxidase. *Proc. Natl. Acad. Sci. USA* 95:12819-12824
- Michel, H. 1999. Cytochrome *c* oxidase: catalytic cycle and mechanisms of proton pumping--a discussion. *Biochemistry* 38(46): 15129-15140
- Michel, H., J. Behr, A. Harrenga, A. Kannt. 1998. Cytochrome *c* oxidase: structure and spectroscopy. *Annu Rev Biophys Biomol Struct.* 27:329-356
- Mills, D. A., and S. Ferguson-Miller. 1998. Proton uptake and release in cytochrome *c* oxidase: separate pathways in time and space? *Biochim. Biophys. Acta* 1365(1-2):46-52
- Mills, D. A., and S. Ferguson-Miller. 2003. Understanding the mechanism of proton movement linked to oxygen reduction in cytochrome *c* oxidase: lessons from other proteins. *FEBS Letters* 545:47-51

- Mills, R. 1973. Self-diffusion in normal and heavy water in the range 1-45°. *J. Phys. Chem.* 77:685-688
- Mitchell, P. 1966. Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. Glynn Research Ltd., Bodmin
- Morgan, J. E., M. I. Verkhovskiy, M. Wikstrom. 1994. The histidine cycle: a new model for proton translocation in the respiratory heme-copper oxidases. *J Bioenerg Biomembr* 26(6):599-608
- Morgan, J. E., M. I. Verkhovskiy, G. Palmer, and M. Wikstrom. Role of the PR intermediate in the reaction of cytochrome c oxidase with O₂. *Biochemistry* 40(23):6882-92
- Nagle, J. F., H. J. Morowitz. 1978. Molecular mechanism for proton transport in membranes. *Proc. Natl. Acad. Sci. USA* 75(1):298-302
- Norberg, J., L. Nilsson. 1994. Stacking-unstacking of the dinucleoside monophosphate guanylyl-3',5'-uridine studied with molecular dynamics. *Biophys. J.* 67(2):812-24
- Nosé, S. 1984. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 52:255-268
- Olkhova, E., M. C. Hutter, M. A. Lill, V. Helms, and H. Michel. Dynamic water networks in cytochrome c oxidase from *Paracoccus denitrificans* investigated by molecular dynamics simulations. *Biophys J.*, accepted
- Oprea, T. L., G. Hummer, A. E. Garcia. 1997. Identification of a functional water channel in cytochrome P450 enzymes. *Proc. Natl. Acad. Sci. USA* 94:2133-2138
- O'Rourke, J. 1995. Computational geometry. In C. 1st ed. Cambridge: Cambridge University Press
- Ostermeier, C., A. Harrenga, U. Ermler, and H. Michel. 1997. Structure at 2.7 Å resolution of the *Paracoccus denitrificans* two-subunit cytochrome c oxidase complexed with an antibody F_V fragment. *Proc. Natl. Acad. Sci. USA* 94:10547-10553
- Otting, G., E. Liepinsh, B. Halle, U. Frey. NMR identification of hydrophobic cavities with low water occupancies in protein structures using small gas molecules. 1997. *Nat Struct Biol.* 4(5):396-404
- Pastor, R. W., R. M. Venable, and M. Karplus. 1991. Model for the structure of the lipid bilayers. *Proc. Natl. Acad. Sci. USA* 88:892-896
- Petrache, H. I., A. Grossfield, K. R. MacKenzie, D. M. Engelman, T. B. Woolf. 2000. Modulation of glycophorin A transmembrane helix interactions by lipid bilayers: molecular dynamics calculations. *J. Mol. Biol.* 302(3):727-46

- Pfützner, U., K. Hoffmeier, A. Harrenga, A. Kannt, H. Michel, E. Bamberg, O. M. Richter, B. Ludwig. 2000. Tracing the D-pathway in reconstituted site-directed mutants of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Biochemistry* 39(23):6756-62
- Pomès, R., and B. Roux. 1996. Structure and dynamics of a proton wire: a theoretical study of H⁺ translocation along the single-file water chain in the gramicidin A channel. *Biophys. J.* 71:19-39
- Pomès, R., G. Hummer and M. Wikström. 1998. Structure and dynamics of a proton shuttle in cytochrome *c* oxidase. *Biochim. Biophys. Acta* 1365:255-260
- Poulos, T. L., B. C. Finzel, A. J. Howard. 1986. Crystal structure of substrate-free *Pseudomonas putida* cytochrome P450. *Biochemistry* 25:5314-5322
- Puustinen, A., M. Wikström. 1999. Proton exit from the heme-copper oxidase of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 96(1):35-7
- Qi, P. X., J. L. Urbauer, E. J. Fuentes, M. F. Leopold, A. J. Wand. 1994. Structural water in oxidized and reduced horse heart cytochrome *c*. *Nature Struct. Biol.* 1:378-382
- Resat, H., M. Mezei. 1994. Grand canonical Monte Carlo simulation of water positions in crystal hydrates. *J. Am. Chem. Soc.* 116:7451-7452
- Richards, F. M. 1977. Areas, volumes, packing, and protein structure. *Ann. Rev. Biophys. Bioeng.* 6:151-176
- Riistama, S., G. Hummer, A. Puustinen, R. B. Dyer, W. H. Woodruff, M. Wikström. 1997. Bound water in the protein translocation mechanism of the heme-copper oxidases. *FEBS Letters* 414:275-280
- Roux, B., M. Nina, R. Pomès, J. C. Smith. 1996. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: A molecular dynamics free energy perturbation study. *Biophys. J.* 71:670– 681
- Ruitenber, M., A. Kannt, E. Bamberg, K. Fendler, H. Michel. 2002. Reduction of cytochrome *c* oxidase by a second electron leads to proton translocation. *Nature* 417:99-102
- Rupley, J. A., G. Careri. 1991. Protein hydration and function. *Adv. Protein Chem.* 41:37-172
- Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equation of motions of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Chem.* 23:327-341
- Sastry, S., D. S. Corti, P. G. Debenedetti, F. H. Stillinger. 1997. Statistical geometry of particle packings. I. Algorithm for exact determination of connectivity, volume, and surface areas of void space in monodisperse and polydisperse sphere packing. *Phys. Rev. E* 56:5524-5532

- Savage, H. 1986. Water structure in vitamin B12 coenzyme crystals. I. Analysis of the neutron and x-ray solvent densities. *Biophys. J.* 50(5):947-956
- Shen, L., D. Bassolino, T. Stouch. 1997. Transmembrane helix structure, dynamics, and interactions: multi-nanosecond molecular dynamics simulations. *Biophys. J.* 73(1):3-20
- Schlenkrich, M. J., J. Brickmann, Jr., A. D. MacKerell, and M. Karplus. 1996. An empirical potential energy function for phospholipids: criteria for parameter optimization and applications. In *Biological Membranes. A Molecular Perspective from Computation and Experiment*. K. M. Merz, and B. Roux, editors. Birkhäuser, Boston. 31-81
- Schoenborn, B. P., A. Garcia, and R. Knott. 1995. Hydration in protein crystallography. *Progr. Biophys. Mol. Biol.* 64:105-119
- Siegbahn, P. E. M., M. R. A. Blomberg, M. L. Blomberg. 2003. Theoretical study of the energetics of proton pumping and oxygen reduction in cytochrome oxidase. *J. Phys. Chem. B*, in press
- Smondjyrev, A. M., and M. L. Berkowitz. 2001. Molecular dynamics simulation of the structure of dimyristoylphosphatidylcholine bilayers with cholesterol, ergosterol and lanosterol. *Biophys. J.* 80:1649-1658
- Soulimane, T., G. Buse, G. P. Bourenkov, H. D. Bartunik, R. Huber, M. E. Than. 2000. Structure and mechanism of the aberrant *ba₃* – cytochrome *c* oxidase from *Thermus thermophilus*. *EMBO Journal*, 8:1766-1776
- Springs, S. L., S. E. Bass, G. Bowman, I. Nodelman, C. E. Schutt, G. L. McLendon. 2002. A multigeneration analysis of cytochrome b(562) redox variants: evolutionary strategies for modulating redox potential revealed using a library approach. *Biochemistry* 41:4321-4328
- Svensson-Ek, M., J. Abramson, G. Larsson, S. Tornroth, P. Brzezinski, and S. Iwata. 2002. The X-ray crystal structures of wild-type and EQ(I-286) mutant cytochrome *c* oxidases from *Rhodobacter sphaeroides*. *J. Mol. Biol.* 321(2):329-39
- Tang, Y., and L. Nilsson. 1999. Molecular dynamics simulations of the complex between human U1A protein and hairpin II of U1 small nuclear RNA and of free RNA in solution. *Biophys. J.* 77:1284-1305
- Thomas, J. W., A. Puustinen, J. O. Alben, R. B. Gennis, and M. Wikström. 1993. Substitution of asparagine for aspartate-135 in subunit I of the cytochrome bo ubiquinol oxidase of *Escherichia coli* eliminates proton-pumping activity. *Biochemistry* 32:10923-10928
- Tieleman, D. P., and H. J. C. Berendsen. 1998. A molecular dynamics study of the pores formed by *Escherichia coli* OmpF porin in a fully hydrated palmitoylcholine bilayer. *Biophys. J.* 74:2786-2801

- Tieleman, D. P., and M. S. P. Sansom. 2001. Molecular dynamics simulations of antimicrobial peptides: from membrane binding to transmembrane channels. *Int. J. Quant. Chem.* 83 :166-179
- Tsukihara, T., H. Aoyama, E. Yamashita, T. Tomizaki, H. Yamaguchi, K. Shinzawa-Itoh, R. Nakashima, R. Yaono, and S. Yoshikawa. 1996. The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å. *Science* 272:1136-1144
- Tuckerman M.E., Liu Y., Ciccotti G., and Martyna G.J. (2001). Non-Hamiltonian molecular dynamics: generalizing Hamiltonian phase space principles to non-Hamiltonian systems. *J. Chem. Phys.* 115:1678-1702
- Van Vlijmen, H. W. T., M. Schaefer, M. Karplus. 1998. Improving the accuracy of protein pK_a calculations – conformational averaging versus the average structure. *Proteins* 33:145-158
- Venable, R. M., Y. Zhang, B. J. Hardy, and R. W. Pastor. 1993. Molecular dynamics simulations of a lipid bilayer and of hexadecane: an investigation of membrane fluidity. *Science* 262:223-226
- Verkhovskiy, M. I., A. Jasaitis, M. L. Verkhovskaya, J. E. Morgan, M. Wikström. 1999. Proton translocation by cytochrome *c* oxidase. *Nature* 400:480-483
- Voorintholt, R., M. T. Kusters, G. Vegter, G. Vriend, W. G. J. Hol. 1989. A very fast program for visualizing protein surfaces, channels and cavities. *J. Mol. Graphics* 7:243-245
- Wade, R. C. 1990. Solvation of the active site of cytochrome P450 cam. *J. Comp. Aid. Mol. Des.* 4:199-204
- Wade, R. C., M. H. Mazar, J. A. McCammon, F. A. Quiñocho. 1991. A molecular dynamics study of thermodynamic and structural aspects of the hydration of cavities in proteins. *Biopolymers* 31:919-931
- Wade, R.C., and P. J. Goodford. 1993. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.* 36(1):148-156
- Wikström, M. K. F. 1977. Proton pump coupled to cytochrome *c* oxidase in mitochondria. *Nature* 266:271-273
- Wikström, M. 1989. Identification of the electron transfers in cytochrome oxidase that are coupled to proton-pumping. *Nature* 338:776-778
- Wikström, M., M. I. Verkhovskiy, G. Hummer. 2003. Water-gated mechanism of proton translocation by cytochrome *c* oxidase. *Biochim. Biophys. Acta* 1604(2):61-5

- Woolf, T. B., B. Roux. 1994. Molecular dynamics simulation of the gramicidin channel in a phospholipid bilayer. *Proc. Natl. Acad. Sci. USA*. 91(24):11631-11635
- Woolf, T. B., B. Roux. 1994b. The conformational flexibility of o-phosphorylcholine and o-phosphorylethanolamine: a molecular dynamics study of solvation effects. *J. Am. Chem. Soc.*, 116, 5916-5926
- Woolf, T. B., B. Roux. 1996. Structure, energetics, and dynamics of lipid-protein interactions: A molecular dynamics study of the gramicidin A channel in a DMPC bilayer. *Proteins* 24:92-114
- You, T. J., D. Bashford. 1995. Conformation and hydrogen ion titration of proteins: A continuum electrostatic model with conformational flexibility. *Biophys J*. 69:1721-1733
- Yu, B., M. Blaber, A. M. Gronenborn, G. M. Clore, D. L. D. Caspar. 1999. Disordered water within a hydrophobic protein cavity visualized by X-ray crystallography. *Proc. Natl. Acad. Sci. USA* 96:103-108
- Zaslavsky, D., R. B. Gennis. 2000. Proton pumping by cytochrome oxidase: progress, problems and postulates. *Biochim. Biophys. Acta* 1458:164-179
- Zhang, L., J. Hermans. 1996. Hydrophilicity of cavities in proteins. *Proteins* 24:433-438
- Zheng, X., D. M. Medvedev, J. Swanson, A. A. Stuchebrukhov. 2003. Computer simulation of water in cytochrome *c* oxidase. *Biochim. Biophys. Acta* 1557:99-107
- Zhang, L., J. Hermans. 1996. Hydrophilicity of cavities in proteins. *Proteins* 24:433-438

ACKNOWLEDGEMENTS

I am grateful to Prof. Hartmut Michel for his supervision, his helpful discussions and his support throughout these years.

This entire work and every single piece of it are largely due to the expert guidance of Prof. Volkhard Helms who introduced me to the field of theoretical biophysics and to whom I am profoundly grateful.

I would also like to thank Prof. Werner Mäntele of the University of Frankfurt for the joint supervision of this work.

Dr. Roy Lancaster was always ready for discussions of any kind and helped me solving many problems, therefore I am very much grateful to him.

I am also grateful to Dagmar Flöck, Dr. Michael Hutter, Marcus Lill, and to all the members of the formal Theoretical Biophysics group for all the advice, help and company they gave me these years.

All the present and former members of the Department of Molecular Membrane Biology for being kind and helpful in many ways. A special thought goes to Alex Haas, Elena Herzog, Kristina Kirchberg and Kerstin Budiman for many discussions about this project and to Hildur Palsdottir and Dr. Stephen Marino for the critical reading of the thesis.

I thank Barbara Schiller, Dr. Lutz Kampmann and Dr. Johan Postma for their efforts to keep the network and all computers in our institute properly running.

I also thank Dr. Thomas Soddemann (Rechenzentrum Garching Max-Planck-Gesellschaft & IPP) for help with parallelization of the CHARMM version c28b2 program for the IBM Power4 machine. Computing time was generally provided by the Rechenzentrum Garching Max-Planck-Gesellschaft & IPP (Dr. I. Weidl and Dr. H. Lederer).

I would especially like to thank my parents for their encouragement and help.

This work was financially supported by the Deutsche Forschungsgemeinschaft (SFB 472), the Fonds der Chemischen Industrie, and the Max-Planck-Gesellschaft.

Lebenslauf.



Name Elena V. Olkhova

geboren am 10 Februar 1976 in Chelkovo, Moskau

Ausbildung

Schule 1983 – 1993 Gymnasium, Chelkovo, Moskau

Studium 1993 – 1999 Studium der Physik an der
Moskauer Lomonossow-Universität (Fachbereich Physik)

1996 – 1999 Studium der Biophysik an der
Moskauer Lomonossow-Universität (Fachbereich Biophysik)
Thema der Abschlussarbeit: “The role of adenylate metabolism
in the regulation of ATP level in human erythrocytes.”

Promotion May 2000 – December 2003 in der Abteilung von
Prof. Dr. Michel am Max-Planck-Institut für Biophysik
in Frankfurt am Main

Thema der Dissertation: “Molecular Dynamics Simulations and
Hydrogen-Bonded Network Dynamics of Cytochrome *c* Oxidase
from *Paracoccus denitrificans*”