

FREIE UNIVERSITÄT BERLIN

**Computing the pK_A Values of Functional
Residues in Proteins:
The Case of Proton Transfer in
Cytochrome *c* Oxidase**

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by

Jovan Dragelj

from Belgrade, Serbia

Year of submission 2018

Die vorliegende Arbeit wurde unter Anleitung von Prof. Dr. Ernst Walter Knapp im Zeitraum vom 01.04.2014 bis 13.12.2018 am Institut für Chemie der Freien Universität Berlin im Fachbereich Biologie, Chemie, Pharmazie durchgeführt.

1. Gutachter: Prof. Dr. Ernst Walter Knapp, Freie Universität Berlin
2. Gutachterin: Prof. Dr. Bettina Keller, Freie Universität Berlin

Disputation am 09.04.2019

Acknowledgements

First, I would like to thank my mentor Prof. Dr. Ernst Walter Knapp for his guidance and great support during my doctoral work at Freie Universität Berlin. It has been a privilege to be his student and get a share of his scientific knowledge and expertise.

I would also like to thank Prof. Dr. Bettina Keller for kindly accepting to review my dissertation.

A very special thanks to my dear colleague and a friend, Dr. Nadia Elghobashi-Meinhardt, who encouraged and advised me over the years and for her invaluable help with proofreading of this dissertation.

I am grateful to Dr. Milan Hodošček, who offered me continuous help with molecular modeling and instructed me with computer cluster administration.

To all my colleagues from AG Knapp for a very friendly and encouraging atmosphere in our group. Thanks to Dr. Tim Meyer and Dr. Anna Lena Wölke with whom I successfully carried out several projects.

I acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG) through the Sonderforschungsbereich (SFB) 1078 "Protonation Dynamics in Protein Function".

I would like to thank members of the SFB 1078 for helpful discussions, collaborations and very pleasant annual retreats. Many sincere thanks to Prof. Dr. Ulrike Alexiev and Alexander Wolf from Freie Universität Berlin, for fruitful collaboration over the past three years which greatly contributed to my doctoral work. I am grateful for discussions and successful projects with Prof. Dr. Maria Andrea Mroginski from Technische Universität Berlin and Prof. Dr. Petra Imhof from Freie Universität Berlin.

To my friends and family from Berlin, Belgrade and Vienna who had patience and supported me over the years.

Lastly but most importantly, I am profoundly grateful to my parents, Dušanka Dragelj and Ljubomir Dragelj who have always unconditionally supported and motivated me to pursue my goals. Therefore, I am dedicating this work to them. *Hvala vam!!!*

Table of Contents

Publications	7
Abstract	8
Zusammenfassung	10
1. Introduction	12
1.1 Electrostatic Theory	13
1.2 Computing pK _A values	17
1.3 Molecular dynamics simulations	24
1.4 pK _A computations in proteins with Karlsberg ²⁺	27
1.5 Cytochrome <i>c</i> Oxidase	31
2. Improving pK_A computations with Karlsberg²⁺	36
2.1 Introduction	36
2.2 Methods	38
2.3 Results and Discussion	44
2.4 Summary and Conclusions	47
3. Proton-Loading Site in Cytochrome <i>c</i> Oxidase	49
3.1 Introduction	49
3.2 Methods	52
3.3 Results and Discussion	61
3.4 Summary and Conclusions	71
4. Cytochrome <i>c</i> Oxidase near the K-channel entry	73
4.1 Introduction	73
4.2 Methods	76
4.3 Results and Discussion	83
4.4 Summary and Conclusions	96
5. Conclusions and Outlook	98
6. References	100
Appendix A	108
Appendix B	111
Appendix C	116
Curriculum Vitae	127

Publications

Publications connected to the results presented in chapter 4 (“*Cytochrome c Oxidase near the K-channel entry*”) are in preparation.

Publication of results from chapter 3 (“*Proton-Loading Site in Cytochrome c oxidase*”) is planned.

During the time period of my doctoral work, I participated in the following publications:

1. H. Batebi, **J. Dragelj**, P. Imhof, Role of AP-endonuclease (Ape1) active site residues in stabilization of the reactant enzyme-DNA complex, *Proteins* 86 (2018) 439-453.
2. A. Takiden, F. Velazquez-Escobar, **J. Dragelj**, A.L. Woelke, E.W. Knapp, P. Piwowarski, F. Bart, P. Hildebrandt, M.A. Mroginski, Structural and Vibrational Characterization of the Chromophore Binding Site of Bacterial Phytochrome Agp1, *Photochemistry and Photobiology* 93 (2017) 713-723.
3. **J.L. Dragelj**, I.M. Stanković, D.M. Božinovski, T. Meyer, D.Z. Veljković, V.B. Medaković, E.W. Knapp, S.D. Zarić, C-H/O Interactions of Aromatic CH Donors within Proteins: A Crystallographic Study, *Crystal Growth and Design* 16 (2016) 1948–1957.
4. D.P. Malenov, **J.L. Dragelj**, G. V. Janjić, S.D. Zarić, Coordinating benzenes stack stronger than noncoordinating benzenes, even at large horizontal displacements, *Crystal Growth and Design* 16 (2016) 4169–4172.

Abstract

The structure and function of proteins depend on the protonation states of amino acids; ergo, the pK_A values of titratable residues in proteins and the influence of protein environment on these pK_A values are of great interest in biochemistry. In this work, the method for pK_A computations is improved. More importantly, electrostatic energy and pK_A computations are used to shed light on the function of a prominent enzyme in energy metabolism, namely cytochrome *c* oxidase.

Karlsberg2⁺ is a modular software used for pK_A computations of proteins with known structure, using an implementation of the established protocol from its preceding version. The pK_A values are obtained by solving the linearized Poisson-Boltzmann equation. To adapt the protein conformation to different pH ranges, local conformational changes are modeled, i.e. salt-bridges are perturbed. The slow modeling procedure from the software predecessor is improved, accelerating the computation 16 times on the average. The software was tested on the extensive benchmark set of 194 experimental pK_A values, however, computation accuracy remains unchanged.

The important elements of cytochrome *c* oxidase (CcO) function are elucidated. CcO is a terminal enzyme of the electron transport chain in mitochondria and some aerobic bacteria. It is reducing the molecular oxygen and using the released energy to pump protons across the membrane against the electrochemical gradient. With each electron from cytochrome *c*, one proton is pumped and one is spent in the chemical reaction (chemical proton) at the binuclear center (BNC). In total, four pumped and two chemical protons are translocated via the D-channel and remaining two chemical protons via the K-channel.

The proton loading site (PLS) is a molecular group whose pK_A increases upon reduction of BNC and temporarily store a pumped proton. The arrival of the chemical proton in the BNC and possibly the next upcoming pumped proton lowers the pK_A of the PLS, pumping proton to the other side of the membrane. The required difference in the pK_A values should approximately be 5 pH units. In this study, potential PLS candidates are identified and studied. The propionates D- and A- of heme a_3 (PRDa₃ and PRAa₃, respectively) can act as the PLS, when the salt-bridge, blocking the protonation of PRDa₃, is open and when PRAa₃ is not involved in hydrogen bonding. Coupled with charge changes at the BNC, the pK_A change of PRDa₃ is 2.24 and of PRAa₃ is 1.36, suggesting that other changes in the environment of the

PLS may take place in order to fully explain its function. Propionates of heme a are excluded as PLS candidates. Loading and unloading schemes of the PLS involving propionates of heme a₃ and His ligand of Cu_B are postulated.

K-channel is active in the reductive part of the CcO redox cycle. In this study, pioneering work, which combines molecular dynamics (MD) simulations and pK_A computations with site-directed fluorescence labelling experiments, explains some aspects of the K-channel activity. His73 is proposed as the K-channel proton entry point, possibly serving as a proton shuttle. The direct electrostatic influence of redox changes at the BNC on pK_A values of residues at the CcO surface close to the K-channel entry is marginal. As a consequence, it is implied that in the reductive part of the CcO cycle, subtle conformational changes take place, namely rearrangements of hydrogen-bond networks, possibly affecting some secondary structural elements.

Zusammenfassung

Die Struktur und Funktion von Proteinen hängt von den Protonierungszuständen der Aminosäuren ab. Daher sind die pK_A Werte von titrierbaren Residuen in Proteinen und der Einfluss der Proteinumgebung auf diese pK_A Werte in der Biochemie von großem Interesse. In dieser Arbeit wird die Methode für pK_A Berechnungen verbessert. Außerdem werden elektrostatische Energie und pK_A -Berechnungen verwendet, um die Funktion eines wichtigen Enzyms im Energiestoffwechsel, nämlich der Cytochrom-*c*-Oxidase, zu untersuchen.

Karlsberg2⁺ ist eine modulare Software, die für die pK_A -Berechnung von Proteinen mit bekannter Struktur verwendet wird, wobei eine Implementierung aus der Vorgängerversion verwendet wird. Die pK_A -Werte werden durch Lösen der linearisierten Poisson-Boltzmann-Gleichung erhalten. Um die Proteinkonformation an verschiedene pH-Bereiche anzupassen, werden lokale Konformationsänderungen modelliert, d. h. Salzbrücken werden geöffnet. Die CPU-intensive Modellierungsmethode des Softwarevorgängers wurde verbessert, wodurch die Berechnung von pK_A Werten im Durchschnitt 16-fach beschleunigt wird. Die Software wurde mit dem umfangreichen Benchmark von 194 experimentellen pK_A -Werten getestet. Die veränderte Methode liefert pK_A -Werte mit derselben Genauigkeit.

Die wichtigen Elemente der Funktion der Cytochrom-*c*-Oxidase (CcO) werden erläutert. CcO ist ein terminales Enzym der Elektronentransportkette in Mitochondrien und einigen aeroben Bakterien. Es reduziert den molekularen Sauerstoff und nutzt die freigesetzte Energie, um Protonen gegen den elektrochemischen Gradienten durch die Membran zu pumpen. Mit jedem Elektron aus Cytochrom *c* wird ein Proton gepumpt und eines in der chemischen Reaktion (chemisches Proton) am aktiven Zentrum (BNC) verbraucht. Insgesamt werden vier gepumpte und zwei chemische Protonen über den D-Kanal und die verbleibenden zwei chemischen Protonen über den K-Kanal transportiert.

Die Protonenladungsstelle (PLS) ist eine molekulare Gruppe, deren pK_A bei der Reduktion von BNC ansteigt und vorübergehend ein gepumptes Proton speichert. Das Eintreffen des chemischen Protons im BNC und möglicherweise des nächsten gepumpten Protons senkt den pK_A Wert des PLS ab. Als Folge verlässt das Proton den PLS und wandert auf die andere Seite der Membran. Der dazu notwendige Unterschied in den pK_A -Werten beträgt etwa 5 pH Einheiten. In dieser Studie werden potenzielle PLS-Kandidaten identifiziert und untersucht.

Die Propionate D- und A- von Häm a₃ (PRDa₃ und PRAa₃) können als PLS fungieren, wenn die Salzbrücke, die die Protonierung von PRDa₃ blockiert, offen ist und PRAa₃ nicht an der Wasserstoffbrückenbindung beteiligt ist. Zusammen mit den Ladungsänderungen am BNC ist die pK_A-Änderung von PRDa₃ 2,24 und von PRAa₃ 1,36, was darauf hindeutet, dass weitere Änderungen in der Umgebung des PLS auftreten, um die Funktion der PLS vollständig zu erklären. Propionate von Häm a sind als PLS-Kandidaten ausgeschlossen. Das Laden und Entladen der PLS, an denen Propionate von Häm a₃ und His-Ligand von Cu_B beteiligt sind, wird postuliert.

Der K-Kanal ist im reduktiven Teil des CcO-Redox-Zyklus aktiv. In der vorliegenden Studie werden Molekulardynamik (MD) Simulationen und pK_A-Berechnungen mit Experimenten zur gezielten Fluoreszenzmarkierung kombiniert und einige Aspekte der K-Kanal-Aktivität erläutert. His73 wird als K-Kanal-Proton-Eintrittspunkt vorgeschlagen, der möglicherweise als Proton-Shuttle dient. Der direkte elektrostatische Einfluss von Redoxänderungen am BNC auf die pK_A-Werte von Residuen an der CcO-Oberfläche nahe dem K-Kanal-Eingang ist gering. Als Konsequenz wird impliziert, dass im reduktiven Teil des CcO-Zyklus subtile Konformationsänderungen stattfinden. Dies können Umlagerungen von Wasserstoffbrückenbindungsnetzwerken sein, die möglicherweise auch sekundäre Strukturelemente betreffen.

1. Introduction

Biochemical processes, e.g. enzymatic reactions, binding of ligands, allosteric control and other processes, are often governed by electrostatic interactions. In many of these processes, proteins, the working molecules in living cells, play the main role. Protein function, structure and many other properties depend on their net charge and charge distribution among their building blocks, the amino acids. The side-chains (residues) of six out of 20 standard amino acids can be ionized. Generally, protonation states of titratable groups in amino acids (named titratable residues) depend on the pH value of solution but also on their environment within the protein structure. Ionized states of residues are favored when they are solvent exposed (e.g. on the protein surface) and neutral states are stabilized when residues are in hydrophobic environment (non-polar internal cavities). Electrostatic interactions between residues, e.g. salt-bridges and hydrogen bonds, have a strong effect on their protonation states. Hydrogen bonds, in particular, are one of the dominant interactions which stabilize specific protein structures. Therefore, the information on protonation of titratable residues is relevant for protein folding. Moreover, since the relationship between structure and function of proteins is generally well established, in many cases the protonation changes are crucial for enzymatic function. [1–5]

The information about the protonation behavior of a certain titratable group is contained in its pK_A value. The pK_A values of titratable groups can be determined experimentally, often with NMR experiments [6], or computed. The electrostatic interactions in proteins can be treated with classical laws of physics [3], which can be inherently used for pK_A computations. The basic principles of electrostatic theory and the methodology used to compute pK_A values are introduced in chapters 1.1 and 1.2 of this dissertation.

The computation of pK_A values in proteins is not a simple task. The complexity of protein structures requires the employment of robust computational methods. One of these methods, used in this work, is incorporated in software Karlsberg2⁺ [7]. In short, Karlsberg2⁺ can be used to estimate pK_A shifts of titratable residues in protein environment relative to the value in solution. The first project described in this dissertation (chapter 2), was carried out with the aim to improve the efficiency of Karlsberg2⁺.

Recent developments allow the use of molecular dynamics simulations (chapter 1.3) for pK_A computations, yielding better conformational sampling, which ultimately improves the results

[7]. The most interesting application of Karlsberg²⁺ combined with molecular dynamics is to interpret protein function. The enzyme whose function is investigated in this study through electrostatic energy and pK_A computations is cytochrome *c* oxidase [8–12]. As a proton pump, cytochrome *c* oxidase transfers protons across the membrane and as such it is a prime example of the protein structure-function relationship which is based on protonation changes within the protein. In this work, two functional aspects of cytochrome *c* oxidase were investigated: (i) the proton loading site, a titratable group that mediates proton transfer against the electrochemical gradient by changing its pK_A value (chapter 3) and (ii) the protonation changes on the surface of the enzyme close to one of the proton transport channels (chapter 4), in both cases coupled with redox changes of metal cofactors.

1.1. Electrostatic theory

Many biological phenomena can be explained on the molecular level with the computation of electrostatic energies. In computational chemistry, more specifically in molecular mechanics, atoms are often presented as point charges that obey the classical laws of physics. The pK_A value is one of the molecular properties that can be obtained from the electrostatic energies. Before introducing the details of pK_A computations in protein environment, basic concepts of electrostatic theory are presented.

The force acting between two point charges q_1 and q_2 in vacuum at the distance r can be described with Coulomb's law:

$$\vec{F} = \frac{1}{4\pi\epsilon_0} \cdot \frac{q_1 q_2}{|\vec{r}|^2} \frac{\vec{r}}{|\vec{r}|} \quad (1)$$

where ϵ_0 presents the vacuum permittivity. The electrostatic potential of one charge (q_1) is given as the Coulomb's potential:

$$\phi(\vec{r}) = \frac{1}{4\pi\epsilon_0} \cdot \frac{q_1}{|\vec{r}|} \quad (2)$$

The electrostatic interaction energy between the two charges is:

$$E = q_2 \cdot \phi_1(\vec{r}) \quad (3)$$

In a simple charge model, equation (3) can be used to calculate electrostatic interactions between the charges. However, in the complex systems like proteins solvated in water with

present ions, the dielectric medium is rather inhomogeneous and it requires a more general description of the electrostatic potential than in equation (2). The solution to this problem comes by solving the Poisson-Boltzmann equation (PBE), where an inhomogeneous dielectric can be used with a spatially varying dielectric constant $\epsilon(\mathbf{r})$ to account for dielectric screening and presence of the solvent. The PBE can also account for the presence of ions expressed as the salt concentration. In the following, these concepts are discussed in more detail.

The Poisson-Boltzmann equation

Gauss's law is one of the fundamental laws of electrostatic theory and it is based on one of the four axiomatic Maxwell equations of electromagnetism. Gauss's law implies that the charge is the source of the electric field and the net flux through the closed surface and the net electric charge (q) enclosed within that surface are related as follows:

$$\oint \vec{E} \cdot d\vec{A} = \frac{q}{\epsilon_0} \quad (4)$$

where ϵ_0 is the vacuum permittivity (or permittivity of free space).

With the application of the divergence theorem, the behavior of the electric field can be expressed in a volume and furthermore its relation to the charge density is given as:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho(\vec{r})}{\epsilon_0} \quad (5)$$

The Poisson equation of electrostatics can be obtained by expressing the electric field in terms of negative gradient of the potential ϕ and combining it with equation (5):

$$\vec{\nabla} \cdot \vec{\nabla} \phi(\vec{r}) = -\frac{\rho(\vec{r})}{\epsilon_0} \quad (6)$$

Equation (6) can be applied only for the electric field in vacuum. In other dielectric media, the presence of induced dipoles reduces the effective electric field transmitted through it. This effect is called "dielectric screening". For systems like proteins solvated in water, mobile water molecules can easily change their orientation to adapt to the charge distribution in the protein, which causes this effect. The electric field from equation (5) is replaced by the electrostatic displacement $\vec{D} = \epsilon \vec{E}$ and it gives a new formulation of equation (6):

$$\vec{\nabla} \epsilon(\vec{r}) \cdot \vec{\nabla} \phi(\vec{r}) = -\frac{\rho(\vec{r})}{\epsilon_0} \quad (7)$$

where the scalar function $\epsilon(\vec{r})$ represents the dielectric constant. Dielectric constant is practically the measure of polarity of a dielectric medium, assuming uniform polarizability of all atoms in the medium, which is an approximation. In electrostatic computations of solvated proteins, two dielectric volumes are considered. Bulk water around the protein is considered without explicit charges as a dielectric medium characterized by $\epsilon_w = 80$. The protein moiety is represented as a dielectric medium with explicit charges and commonly characterized by $\epsilon_p = 4$. The appropriate value of the protein dielectric constant is still an open question. It has been suggested that a value of protein dielectric constant should be between 2 and 20 [13]. In proteins, many amino acids carry permanent dipoles and can reorient their charges side-chains according to the electric field. There are discussions that the inner, more hydrophobic, part of the protein structure is best described with the dielectric constant of 4 and the surface, more polar part, with the dielectric constant of 20 [14]. In this work, the dielectric constant used for the protein volume was 4. The energies calculated from the ϕ obtained from equation (7) is valid only for the system solvated in distilled (pure) water. However, in reality, water surrounding proteins contains metal ions, which can be expressed as the salt concentration. To account for mobile ions in water, the Debye-Hückel theory can be employed to ‘correct’ equation (7), giving the Poisson-Boltzmann Equation (PBE):

$$\vec{\nabla}\epsilon(\vec{r}) \cdot \vec{\nabla}\phi(\vec{r}) = -4\pi \left[\rho(\vec{r}) + k^2 \frac{kT}{e_c} v(\vec{r}) \sinh\left(\frac{e_c\phi(\vec{r})}{k_B T}\right) \right] \quad (8)$$

with

$$k = \sqrt{\frac{8\pi N_A e_c^2 I_s}{k_B T}} \quad (9)$$

where k represents the inverse Debye length, e_c the elementary charge, N_A the Avogadro’s number, T absolute temperature and the k_B the Boltzmann constant. The ionic strength is expressed through parameter $I_s = \frac{1}{2} \sum_i c_i \cdot z_i^2$, where c_i represents the concentration of the ion i with charge z_i . The part of equation (8) expressed as $\sinh\left(\frac{e_c\phi(\vec{r})}{k_B T}\right)$ represents the statistical Boltzmann distribution of ions. The $v(\vec{r})$ is the volume exclusion function and it depends on the protein structure with its value being one in the ion accessible region (solvent) and zero in all other regions (protein, other molecules(e.g. membrane)). The PBE given in (8) cannot be solved in case of complex molecular structures, which is usually the case for systems investigated in this work. At low charge densities, the PBE can be simplified by approximating

the sine hyperbolic function with the Taylor series. The linearized Poisson-Boltzmann equation (**LPBE**) is expressed as:

$$\vec{\nabla}\varepsilon(\vec{r})\cdot\vec{\nabla}\phi(\vec{r}) + k^2(\vec{r})\phi(\vec{r}) = -4\pi\rho(\vec{r}) \quad (10)$$

The advantage of using the LPBE is the additivity of the obtained electrostatic potentials and energies. Additivity is widely used in chapter 1.2. (*“Computation of pK_A values”*), where electrostatic energy terms are evaluated separately and summed up to give the total electrostatic energy.

Solving the LBPE, shown in equation (10), can be done by several methods. One of the most commonly used methods is Finite Difference (FD), which is implemented in the software APBS [15]. This method was utilized in this work. In the FD method, a molecule is placed on the cubic grid and the atomic charges are interpolated onto the grid points. In this way a molecules is represented by the charge distribution $\rho(\vec{r})$, the spatial distribution of the dielectric constant $\varepsilon(\vec{r})$ and the volume exclusion function $v(\vec{r})$. The Grid has a fixed grid constant and a defined resolution. The best possible grid resolution is desired, however it comes at the cost in computer memory. For example, for a grid constant of about 0.3-0.25Å, between 1 and 4 GB of RAM is needed.

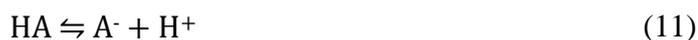
There are two issues when solving the LPBE. In order to solve the LPBE for every grid point, the electrostatic potential at the edges of the grid has to be defined, commonly named the “boundary value” issue. One solution to this problem is the usage of a very large grid, for which the potential at the edges vanishes. However, a large grid with high resolution is computationally too expansive. Another solution to the boundary problem is the application of the focusing method. The focusing method assumes the usage of a large grid with low resolution, where the electrostatic potential obtained from this grid is used as the boundary condition for the next computation where a grid with higher resolution is used and where the volumes contains the atomic charges. Usually three focusing steps are enough to obtain sufficiently precise results. The second issue when solving LBPE is called “the grid artifact”. The grid artifact occurs when point charges are used on discrete grids, because the resulting potential comes also from the interaction of grid points with each other. Therefore, the obtained electrostatic potential can safely be used only to compute the interaction energies or energy differences. The interaction energy can be evaluated for the two sets of charges that are not too close to one another. The energy difference is computed by subtracting the electrostatic

potential of the two computations, for the same charge distribution. In that way, the grid artifact is cancelled. The energy difference is used in this work to compute solvation energies, for which computation relies on a value of the dielectric constant used in the solvent volume.

1.2. Computing pK_A values

For protein systems it is very important to know the protonation states of amino-acid side-chains and the energy needed to change the protonation state, especially for a proton pump like Cytochrome *c* Oxidase. The pK_A value of a titratable group is the answer to those questions. In this work, the focus is on the computation of pK_A values of amino-acid side-chains but it can be applied to all titratable molecules.

The pK_A value of a molecular species in a sample is the pH value at which half of the population is in protonated and half in deprotonated state or simply, when the protonation probability is 0.5. The generic acid HA dissociates to the deprotonated A⁻ and a proton H⁺. The dissociation is described with the reaction:



The equilibrium constant of the dissociation is expressed via equilibrium concentrations:

$$K_A = \frac{[\text{A}^-][\text{H}^+]}{[\text{HA}]} \quad (12)$$

The pH is the negative decadic logarithm of [H⁺] and after taking the negative common logarithm of K_A, the Henderson-Hasselbalch equation is obtained:

$$\text{pK}_A = \text{pH} - \log\left(\frac{[\text{A}^-]}{[\text{HA}]}\right) \quad (13)$$

Equation (13) can be formulated in terms of protonation probabilities:

$$\text{pK}_A = \text{pH} - \log\left(\frac{1 - \langle\chi\rangle}{\langle\chi\rangle}\right) \quad (14)$$

From equation (14) the protonation probability can be expressed as:

$$\chi = \frac{e^{-\ln(10)[\text{pH}-\text{pK}_A]}}{1 + e^{-\ln(10)[\text{pH}-\text{pK}_A]}} \quad (15)$$

The Gibbs reaction free energy for the deprotonation of the acid can be derived with the help of equation (15):

$$\Delta G(\text{HA} \rightarrow \text{A}) = -\ln(10) \cdot RT \cdot [\text{pH} - \text{pK}_A] \quad (16)$$

where R is the universal gas constant and T the absolute temperature.

pK_A values in proteins

For a titratable molecule in water, the equations derived above can be used to compute pK_A values. However, in proteins the reaction free energy needs to be “corrected” for the presence of the protein environment to which the titratable molecule (often called titratable residue) belongs (e.g. amino acids) or is attached to (e.g. fluorescein label in Chapter 4, cofactors or ligands). The reaction free energy can be expressed as:

$$\Delta G(\text{HA} \rightarrow \text{A}) = -\ln(10) \cdot RT \cdot [\text{pH} - \text{pK}_A] + \Delta G_p \quad (17)$$

The added energy term ΔG_p represents exactly the change in the protonation free energy of the titratable residue due to the presence of the protein environment, which has to be accounted for in electrostatic energy computations. With pH changes, the protein structure may suffer some conformational changes but more likely protonation changes of titratable residues. Furthermore, the changes in protonation of one residue can affect the protonation of others. To account for these changes, ΔG_p is considered as a function of protonation states of all titratable residues in the protein (protonation pattern). The protonation pattern is represented as the protonation vector $p(\text{pH})$. Therefore, if assumed that at a given pH value all titratable residues change their protonation states from the reference protonation vector p_r into any protonation vector p_k , the free energy of this change is $\Delta G_p(p_r \rightarrow p_k, \text{pH})$. The reference protonation vector can be arbitrarily chosen, however in earlier work [16] it was defined as the one where all titratable residues are in their charge neutral state (protonated acids and deprotonated bases). In later method development, other reference vectors were chosen [7], as discussed later in text. The exact mathematical details on how ΔG_p is obtained in this work are described further below. The formulation $\Delta G_p(p_r \rightarrow p_k, \text{pH})$ allows a possibility to compare electrostatic energies of different protonation vectors and evaluate the probability of all protonation states at a chosen pH value, which was applied in this work, as described in chapters 2, 3 and 4. The probability $\chi(p_k)$ of a certain protonation vector p_k at the given pH value can be expressed as:

$$\chi(p_k, \text{pH}) = \frac{e^{-\frac{\Delta G_p(p_k, \text{pH})}{k_B T}}}{\sum_i^{N_p} e^{-\frac{\Delta G_p(p_i, \text{pH})}{k_B T}}} \quad (18)$$

where k_B is the Boltzmann factor, T is the absolute temperature and N_p the total number of protonation states. The relevant information is which protonation pattern is the most probable at the given pH (usually the physiological value of around 7). Ideally, all possible protonation states of all residues would need to be considered. However, even for a protein of the moderate size the number of possible protonation states becomes too large and computationally too demanding. To efficiently compute pK_A values, software Karlsberg [17] is used to evaluate the probabilities by utilizing a Metropolis-Monte-Carlo algorithm, which converges orders of magnitude faster than a demanding computation with all possible protonation states.

Software Karlsberg2⁺ [7] is the latest version of the in-house developed software [16] to perform the titrations of proteins. The protocol of Karlsberg2⁺ is described in the Chapter 1.3. By titrating a protein structure, it is possible to obtain pH dependent protonation probabilities for all titratable residues. In most cases, the protonation probability χ can be approximated by equation (15), although in some systems χ can be a complex function of pH. Therefore, the pK_A value of the titratable residue in the protein environment $\text{pK}_{A,\text{prot}}$ is essentially the shifted pK_A value in the aqueous solution.

The pK_A values of titratable residues in proteins can be obtained by utilizing equation (17). $\Delta G_p(p_r \rightarrow p_k, \text{pH})$, the change of free energy for changing from the reference protonation vector p_r (whose energy is taken to be zero) to any vector p_k , can be decomposed into individual energy terms [18–20]. The details on how ΔG_p is obtained is discussed in the following text.

The protonation energy is the first energy term to be computed (ΔG_{resi}) and it represents the energy needed to change the protonation state of a titratable residue in the protein environment with all other residues being in their reference protonation states. ΔG_{resi} can be computed by solving LPBE and utilizing the thermodynamic cycle presented in Figure 1.1. In the upper part of the cycle (steps 2→3), the energy needed to change the protonation state of a residue in aqueous solution ($\Delta G_{(\text{HA} \rightarrow \text{A})}$) is obtained from the experimental pK_A value for that residue. In this energy term, bond formation (or dissociation) and other quantum effects are taken into account. The energies ΔG_{HA} (steps 2→1) and ΔG_{A} (steps 3→4), called desolvation energies,

are the energies needed to take the residue from the aqueous solution to the protein environment, when the residue is in the protonated and deprotonated state, respectively.

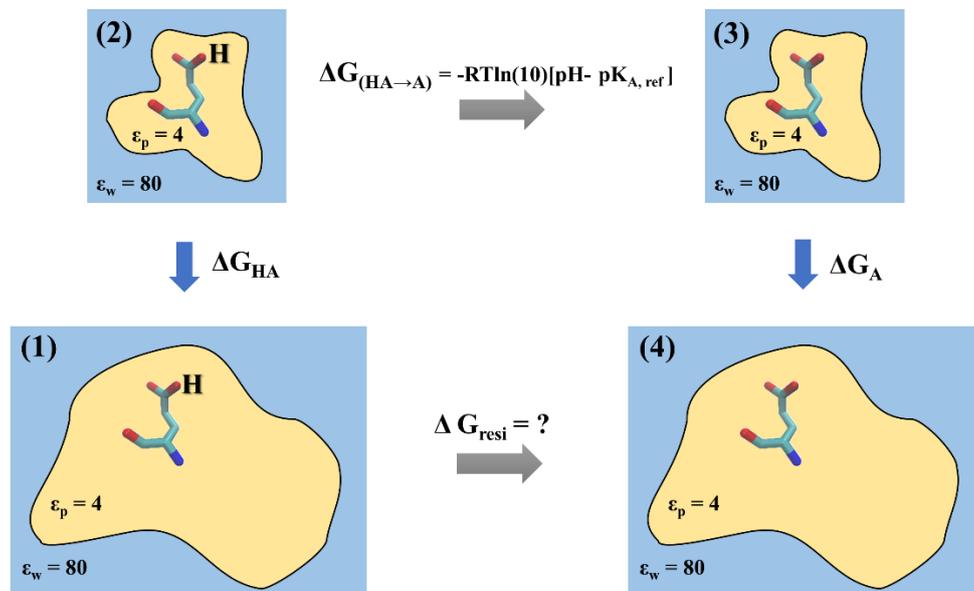


Figure 1.1. Schematic representation of the thermodynamic cycle used to compute the protonation energy (ΔG_{resi}) for deprotonation of glutamate. Blue represents the solvent volume, filled with the dielectric medium of $\epsilon_w = 80$, while the protein volume, filled with the dielectric of $\epsilon_p = 4$ is shown in yellow. The desolvation energies, ΔG_{HA} and ΔG_{A} are obtained by solving the LPBE for protonated and deprotonated glutamate, respectively. $\Delta G_{(\text{HA} \rightarrow \text{A})}$ is the free energy of deprotonation in the aqueous solution.

The desolvation energy of the titratable residue μ in the protonation states consists of two energy terms, given as:

$$\Delta G_{\mu,s} = \Delta G_{\text{Born}}^{\mu,s} + \Delta G_{\text{back}}^{\mu,s} \quad (19)$$

where $\Delta G_{\text{Born}}^{\mu,s}$ represents the self-energy of the residue, called Born energy and $\Delta G_{\text{back}}^{\mu,s}$ represents the background energy or the interaction energy of the residue with the rest of the protein environment. For the titratable residue, above defined as a set of N_q charges q_i , whose coordinates are defined with the vector r_i , the Born energy can be expressed as:

$$\Delta G_{\text{Born}} = \frac{1}{2} \sum_i^{N_q} q_i [\phi_{\text{prot}}(r_i) - \phi_{\text{solv}}(r_i)] \quad (20)$$

The interaction of the residue with the background charges is:

$$\Delta G_{\text{back}} = \sum_j^{N_{\text{prot}}} q_{\text{prot},j} \phi_{\text{prot}}(r_j) - \sum_j^{N_{\text{solv}}} q_{\text{solv},j} \phi_{\text{solv}}(r_j) \quad (21)$$

To compute ϕ_{prot} , only the set of N_{prot} charges of the protein that are not already in contained in the set of charges used for the computation of Born energy are considered ($q_{\text{prot},j}$). For the computations of ϕ_{solv} , only the remaining N_{solv} charges of the titratable residue that were not used for the computation of Born energy are considered ($q_{\text{solv},j}$). The coordinates of the background atoms are defined as vectors r_j . The grid artifact is cancelled in these computations. With the computed separate energy terms of the desolvation energy, the missing energy term for the ΔG_p is the difference in desolvation energy of the titratable residue μ in the protonation state s_μ and in the reference protonation $s_{r,\mu}$:

$$\Delta\Delta G_{\text{desolv},\mu}(s_\mu) = \Delta G(s_\mu) - \Delta G(s_{r,\mu}) \quad (22)$$

Since the changes of protonation of a titratable residue in the protein can affect the protonation of the other titratable residues, an energy term describing the interaction between all titratable residues needs to be added to ΔG_p . This is partially computed in the $\Delta\Delta G_{\text{desolv},\mu}(s_\mu)$, as it contains the interaction energies of the residue μ in protonation state s_μ with the other titratable residues when they are in their reference protonation states. Therefore, the pairwise interactions between the residues when they are in non-reference states are missing. This energy term is named the W-matrix, defined as:

$$W_{\mu,\eta}(s_\mu, s_\eta) = \sum_i^{N_\mu} [q_{i,\mu}(s_\mu) - q_{i,\mu}(s_\mu^r)] \cdot [\phi_{\text{prot},\eta}(s_\eta, r_{i,\mu}) - \phi_{\text{prot},\eta}(s_\eta^r, r_{i,\mu})] \quad (23)$$

where $q_{i,\mu}(s_\mu)$ is a charge from the set of N_μ charges of the residue μ in the state s_μ with coordinates given as $r_{i,\mu}$. The residue η in the state s_η generated the potential $\phi_{\text{prot},\eta}$ in the protein environment. The reference state of a residue is defined as s^r in this case. No additional computations are needed for this energy term, as the needed potentials are already obtained for the desolvation energies.

Finally, the total protonation energy ΔG_p , needed to for the change from the protonation vector p_r to p_k , can be expressed by combining equations (17), (22) and (23), given as:

$$\begin{aligned} \Delta G_p(p_r \rightarrow p_k, \text{pH}) = & \sum_{\mu=0}^N (x_{\mu}^k - x_{\mu}^r) \cdot RT \cdot \ln(10) \cdot (\text{pH} - \text{pK}_{A, \text{ref}}^{\mu}) \\ & + \sum_{\mu=0}^N \Delta \Delta G_{\text{desolv}, \mu}(s_{\mu}) \\ & + \sum_{\mu}^N \sum_{\eta, \mu \neq \eta}^N W_{\mu, \eta}(s_{\mu}, s_{\eta}) \end{aligned} \quad (24)$$

where x_{μ}^k can be 0 or 1, depending if the residue μ is in the protonated or deprotonated state, respectively. x_{μ}^r is used to differentiate if the residue is an acid (value of 0) or a base (value of 1). With x_{μ}^k and x_{μ}^r , the correct sign of the free energy from the experimental pK_A value is assured.

Therefore, from equation (24) the ΔG_p can be used in equation (18) to compute pK_A values of all titratable residues for a single protein structure. The protonation changes are the most important effect when computing ΔG_p , however, it is possible that protonation changes cause local or even global conformational changes in the protein structure. In Karlsberg2⁺, the conformational variation is included and it is described in more detail in chapter 1.4. The total protonation energy is computed as a function of conformational energy, defined as:

$$\begin{aligned} \Delta G_p(p_r \rightarrow p_k, \lambda, \text{pH}) = & \sum_{\mu=0}^N (x_{\mu}^k - x_{\mu}^r) \cdot RT \cdot \ln(10) \cdot (\text{pH} - \text{pK}_{A, \text{ref}}^{\mu}) \\ & + \sum_{\mu=0}^N \Delta \Delta G_{\text{desolv}, \mu}^{\lambda}(s_{\mu}) \\ & + \sum_{\mu}^N \sum_{\eta, \mu \neq \eta}^N W_{\mu, \eta}^{\lambda}(s_{\mu}, s_{\eta}) \\ & + \Delta G_{\text{conf}}^{\lambda} \end{aligned} \quad (25)$$

where $\Delta G_{\text{conf}}^{\lambda}$ represents the conformational energy of the protein structure λ in which all titratable residues are in their reference states.

The conformational energies can also be computed by solving the LPBE. To compute the term $\Delta G_{\text{conf}}^\lambda$, a thermodynamic cycle (Figure 1.2) must be used due to the grid artifact problem. The free energy difference for the transition of conformation in (1) to the conformation in (4), shown in Figure 1.2, is obtained as follows:

$$\Delta G_{\text{conf}} = \Delta G_{\lambda_1} - \Delta G_{\lambda_2} + \Delta G_{\text{Coulomb}} \quad (26)$$

with

$$\Delta G_\lambda = \frac{1}{2} \sum_i^N q_i [\phi_{\text{prot},\lambda}(r_{\lambda,i}) - \phi_{\text{vac},\lambda}(r_{\lambda,i})] \quad (27)$$

where $\Delta G_{\text{Coulomb}}$ is the energy difference in the homogenous dielectric (vacuum) for the conformational change from λ_1 to λ_2 , ΔG_{λ_1} and ΔG_{λ_2} are the free energies of desolvation in the inhomogeneous dielectric for the conformation λ_1 and λ_2 , respectively. The grid artifact cancels as the same grid is used to compute ΔG_λ in both cases.

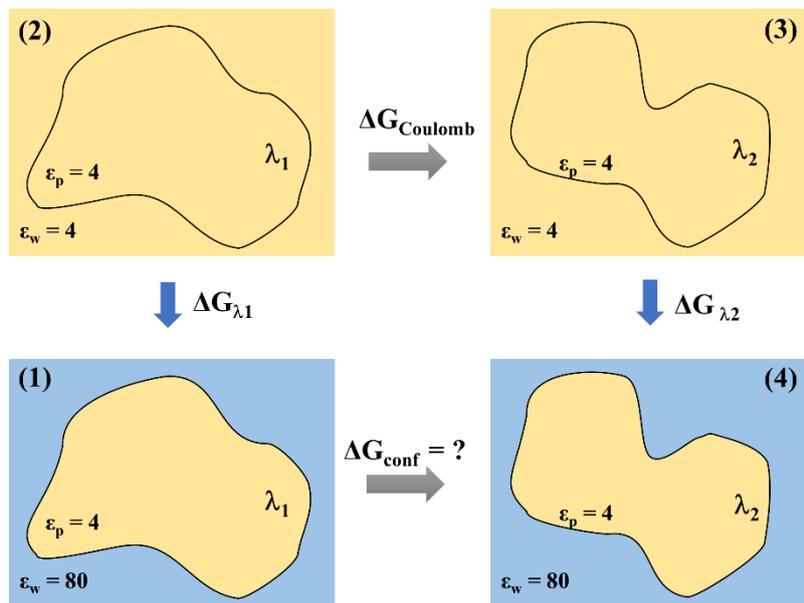


Figure 1.2. Schematic representation of the thermodynamic cycle used to compute the conformational energy (ΔG_{conf}) for two different conformations λ_1 and λ_2 . First, the energy of the conformational change is computed in the homogeneous dielectric environment (steps 2→3), by applying the Coulomb law. The energies ΔG_{λ_1} and ΔG_{λ_2} of the transition of each conformation from the homogenous into the inhomogeneous dielectric (step 1→2 and 3→4) are computed by solving the LPBE. ΔG_{conf} is obtained with equation (26).

Molecular dynamics and pK_A computations

Recent improvement of methodologies showed that Molecular Dynamics (MD) simulations can be used to improve pK_A computations. The procedure Karlsberg2⁺MD (KB2⁺MD)[7] is implemented in Karlsberg2⁺ software and its concept is explained in chapter 1.4. However, there are some conceptual differences on how some of the energy terms of ΔG_p are computed.

In this approach, the entire residue is considered for the computations in equations (20), (21), (23). The reference protonation vector, p_r is defined as a true reference state for which charges of all titratable residue are vanishing. Consequently, the W-matrix (equation (23)) contains pairwise electrostatic energies of all states of all titratable residues. These changes allow an easier interpretation of separate energy terms, which was impossible in the earlier approach (Karlsberg⁺ [16]). With the modification of the W-matrix, the conformation energy computation is computed with titratable residues excluded (charges set to zero), practically representing the conformational change of the background charges. An important detail of the computations when using MD trajectories is the exclusion of the electrostatic interactions of atoms that are separated by up to three chemical bonds (named 1-2, 1-3, 1-4 interactions), which can be achieved if $\Delta G_{\text{Coulomb}}$ is computed with CHARMM [21].

1.3. Molecular dynamics simulations

Research in computational chemistry, is mostly dedicated to elucidate chemical phenomena on the molecular level. While X-ray crystallography data provides structural and functional details of proteins, their behavior in solution cannot be predicted only by looking into the details of the crystal structures. Molecular dynamics (MD) simulations can be used to virtually observe the time dependent behavior not only of proteins but of many other systems (e.g. nucleic acids, material surfaces, etc). This method is used to complete the macroscopic information obtained from experiments with microscopic details on the atomistic level. The first MD simulation on proteins was performed in 1977 [22] on the bovine trypsin inhibitor, a protein with 58 amino acids, resulting only in a few picoseconds of dynamics. Today, MD simulations of large biological systems (over 500,000 atoms) are carried out for hundreds of nanoseconds and often go into the microsecond time range.

In molecular dynamics, molecules are described by the set of functions and parameters contained in a force field (FF). When these forces are derived from a classical treatment of the atoms as point charges, the method is referred to as classical molecular mechanics (MM). In this work the CHARMM [21] force field is used. Atoms are represented as soft spheres and bonds between them as simple harmonic oscillators (springs). In the force field, the total energy consists of two energy terms, for interactions between bonded atoms and for non-bonded, interactions that include electrostatic interactions and van der Waals interactions. The bonded energy terms include contributions from interactions of covalently connected atoms, two for bonds, three for angles and four for dihedral and improper dihedral angles. The electrostatic energy term is based on Coulomb's law and the van der Waals interactions on the Lennard-Jones potential (also known as 6-12 potential) [23]. The total (potential) energy can be expressed as the sum of all terms:

$$E_{\text{total}} = \sum E_{\text{bond}} + \sum E_{\text{angle}} + \sum E_{\text{dihe}} + \sum E_{\text{impr}} + \sum E_{\text{electrostatic}} + \sum E_{\text{vdW}} \quad (28)$$

Before performing dynamics, a relaxation of the system can be achieved by potential energy minimization. There are different minimization methods available in CHARMM [21], however, steepest descent (SD) and the adopted basis Newton Raphson method (ABNR) are commonly used [21]. Newton's law of motion can be applied to the system defined with the force field, to perform molecular dynamics simulations. For a chosen temperature, velocities from the Boltzmann distribution are assigned randomly to all atoms, which causes the motion of the system. Simplified concept of molecular dynamics is: after the movement of atoms, the velocities are recomputed from forces, which is repeated in an elementary time step, which usually is one or two femtoseconds, for a chosen period of time (total simulation time). The number of atoms (N) is constant in typical MD simulations and bonds cannot be broken or formed. Therefore, even if, for example, one titratable residue changes protonation, a new simulation is needed. The system is approximated with periodic boundary conditions (PBC), called the box that is translated infinitely in all directions. Therefore, with the usage of PBC, if atoms from one side leave the box, they reenter on the opposite side of the box. Electrostatic interactions can be computed with in a user-defined cutoff distance; for larger distances, the electrostatics are not computed, thus saving computation time. A more precise representation of the system is to compute all electrostatic interactions in the system. To mimic the realistic behavior of the simulated systems, several parameters can be controlled, e.g. temperature (T), volume of the system (V), or pressure (P). Depending on which statistical ensemble is used to

model the system, the MD simulations may be based on: a microcanonical ensemble (NVE), with volume and energy kept constant; or a canonical ensemble (NVT), with volume and temperature kept constant; or Gibbs' (isothermal-isobaric) ensemble (NPT), where temperature and pressure are kept constant. MD simulations carried out in this work were performed with the NPT-ensemble and Langevin dynamics, where the temperature is maintained by adjusting the friction applied to atoms and constant pressure managed by resizing of the box. In a standard procedure, MD simulation consists of three phases: (1) heating, during which temperature is slowly increased and adjusted to reach the chosen value, (2) equilibration, a step in which the system relaxes can be arbitrarily long. This phase depends on the observed quantities and (3) production, during which statistical properties can be measured from the equilibrated ensemble.

The quality of the MD simulation depends foremost on the force field parameters. A humorous expression, "garbage in – garbage out", is sometimes used to describe the importance of force field parameters for the quality of results. Parameterization is often a tedious task, in which experimental data and quantum-chemical computations are iteratively used to adjust FF parameters. In this work, most parameters were used from the CHARMM force field [21]. For the metal centers (Fe, Cu) in CcO, called cofactors, in-house parameters were developed in previous works [8,9]. They were used for simulations and computations in chapter 3 and chapter 4. In the chapter 4, a non-standard molecule, fluorescein, is added to the system for which charges had to be computed for different protonation states, as well as other parameters (force constants and equilibrium values for various interactions computed with equation (28)), which were carefully assembled from existing parameters from the CHARMM force field [21].

The speed of the MD simulation depends on the available computer power. Recent advances in parallel computing allow the usage of affordable graphical card units (GPUs) to perform very fast MD simulations, however, with a large number of powerful CPU cores, the simulations can be proportionally fast. In this work, GPUs were used to carry out all MD simulations.

1.4. pK_A computations in proteins with Karlsberg2⁺

The software Karlsberg2⁺ was written by Tim Meyer [7] and alternated and optimized by myself using the protocol originally implemented by Gernot Kieseritzky (previous version of the software named Karlsberg⁺ [16]). The new version is much faster and facilitates pK_A value computations of titratable groups in protein crystal structures. In this chapter the motivation for the development of such a software is addressed, as well as the philosophy and a short description of the protocol behind Karlsberg2⁺.

For the computation of pK_A values of titratable groups in proteins, crystal structures of proteins are usually used. The most common source of protein structures are crystal structures obtained with X-ray crystallography from the Protein Data Bank (PDB) [24]. Even though these structures contain the most relevant information about the protein structure and are reliable (given that the resolution is sufficiently high), some atomic coordinates may not be available. Those are usually coordinates of hydrogen atoms (unless the structure is of subatomic resolution), which are rarely present in X-ray crystal structures of proteins. However, in some cases coordinates of heavy atoms are also lacking. In order to proceed, all missing atoms have to be modeled prior to computations, which can be done with CHARMM [21]. Positions of hydrogen atoms define the hydrogen-bond network and are necessary for completing the electrostatic interactions in the protein, which inherently has a large influence on computed pK_A values of titratable residues. Finally, a properly prepared crystal structure can be used for electrostatic energy computations as described in chapter 1.2.

In a standard approach of protein pK_A computations the above mentioned protocol is applied on a single protein structure. In many cases this approach yields good results. But in some cases computed pK_A values can deviate considerably from the corresponding measured values [16]. Titrating a protein sample experimentally, a protein may not only adjust the protonation states of its titratable residues to the varying pH value, but may also change its conformation locally by rearrangement of hydrogen bonds or even globally. Considering only one conformation in a theoretical approach, the result may be biased because the protein conformation obtained from the PDB was appropriate for crystallization at a specific pH value. For evaluation of computational methods, an established benchmark set of 185 experimentally measured pK_A values from 13 different proteins was used for testing in all protocols during the development. Careful analysis of different protocols and results of pK_A computations on protein structures of the benchmark set gave insight why the single structure approach may fail for specific cases.

Two relevant and mutually connected aspects that changes in pH value have on the protein structure are the hydrogen-bond network and conformational changes. When some of the titratable groups change their protonation state, hydrogens in the environment can easily adjust their orientation as a reaction to this charge change which could lead to a rearrangement of the entire hydrogen-bond network. Therefore, an adequate modeling of the coordinates of hydrogen atoms is needed. Another important aspect of protein conformations is the presence of salt-bridges. Salt-bridges are strong ionic interactions between negatively charged (acids like Glu and Asp) and positively charged residues (bases like Lys, Arg and His) and they are usually stable in the pH range between 4 and 10. The analysis of pK_A values computed for the benchmark set showed that continuum electrostatics methods which are using a single structure yield results where ion pairs involved in salt-bridges have a strong tendency remaining ionized. Salt-bridges in a protein become energetically unfavorable at pH values lower than 4, where acids are more likely protonated (charge neutral) and at pH values higher than 10, where bases are more likely deprotonated (charge neutral). To overcome this problem, conformational sampling at different pH values was incorporated into the protocol, also known as pH adapted conformations.

The concept of pH adapted conformations

One of the main features of Karlsberg²⁺ is the consideration of different protein conformations by using pH adapted conformations (PACs). This was first done in the final version of software Karlsberg⁺ [16]. In simple words, a pH adapted conformation is a modeled protein structure with a protonation pattern corresponding to a chosen pH value. In order to determine the proper protonation pattern, an iterative procedure is used, as shown in Figure 1.3. As usual the first is the basic preparation of the protein structure from the PDB, which is then used for a specific PAC modeling. In the development of the Karlsberg⁺ software it was shown that modeling PACs for three pH values were enough to obtain reasonable agreement of computed pK_A values with the experiment. Thus, the same approach has been kept in Karlsberg²⁺. In Karlsberg²⁺ the three PACs represent the protein structure at pH values of -10, 7 and 20. At pH 7 all titratable residues are ionized. As previously explained, at pH -10, all titratable residues are considered to be protonated. Hence, acids are charge neutral and bases positively charged. At pH 20, all titratable residues are deprotonated. Hence, acids are negatively charged and bases are charge neutral. These values are chosen in order to ensure the change in protonation of the mentioned titratable species but the structures modeled according to these values are valid for a range of

pH values. However, the same structure is used at the beginning for all PACs. Following the initial modeling, the specific PAC modeling is done and the protonation pattern is computed as part of the self-consistent cycle (Figure 1.3). Finally, when the most suitable protonation pattern at a given pH value is found, the PAC is properly modeled and the structure can be used for the pK_A computations.

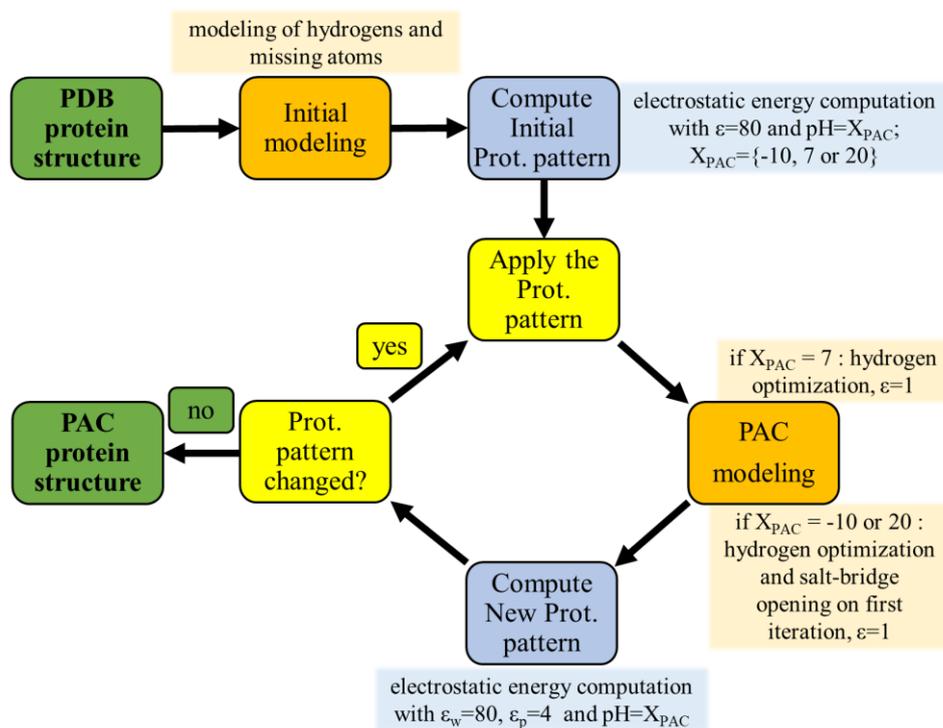


Figure 1.3. A scheme of the self-consistent cycle used for modeling pH adapted conformations (PACs). The initial information is usually the crystal structure and the different pH values for the PACs are -10, 7 and 20 in the standard procedure. The most suitable protonation pattern at a given pH value is obtained from the cycle, which is subsequently applied, as described in chapter 1.2, to obtain a final pH adapted structure.

In Karlsberg2⁺ two modeling procedures are used alternatively generating PACs in an optimization cycle. (i) For a given pH the protonation pattern is determined by electrostatic energy computation with $\epsilon_p = 4$ in the protein volume and $\epsilon_w = 4$ outside in the solvent volume. (ii) The hydrogen atom coordinates are optimized using CHARMM by energy minimization with $\epsilon = 4$ for this protonation pattern [21]. These two steps are repeated until the protonation pattern does not change anymore. In case of protein conformations at pH values of -10 or 20, the side-chains of residues involved in salt-bridges are modeled, before the cycle of iterations is performed, to obtain a chemically reasonable open conformation of the salt-bridge. The

details of salt-bridge modeling are discussed and explained in detail in chapter 2 (“*Improving pK_A computations with Karlsberg2⁺*”). After the PACs are properly modeled, they can be used in a combined titration algorithm according to the equation (25), where the different PACs are weighted with the Boltzmann factor. These titrations are performed for pH values in the range of -10 to 20, with a step of 0.5 pH units (equation (18)) yielding the pK_A values.

Software Karlsberg2⁺

The source code of software Karlsberg2⁺ (KB2⁺) is written in the programming language Python. It can be obtained on request from AG Knapp at Free University of Berlin. The previous version, Karlsberg⁺ was written in Pearl programming language and even though the code was organized well, it was simply too inflexible for changes and future updates, as for example for the implementation of the Karlsberg2⁺MD procedure.

The software KB2⁺ is organized in a modular format, where each module performs a specific task in pK_A computations. The flexibility of the software is that modules are independent from one another and therefore can also be used separately for those tasks, if deemed necessary. Here are the most important functionalities of the modules involved in the pK_A computations:

- Read and write protein structures in various formats (.pdb, .crd, .pqr, .psf, ...)
- Protein modeling in CHARMM[21], using CHARMM22 force field [25]
- Protonation energy computations by using a modified version of APBS [15], named TAPBS. This software is available on the webpage of AG Knapp
- Conformational energy computations by using software APBS [15]
- Titration or the evaluation of the Boltzmann sum (equation (18)) for a range of pH values, which is done by software Karlsberg [17], which utilizes the Metropolis Monte Carlo algorithm. This software is also available on the webpage of AG Knapp
- Karlsberg2⁺MD [7] procedure for evaluating pK_A values by using different MD trajectories
- pK_A computations module that connects all needed modules to compute pK_A values for proteins

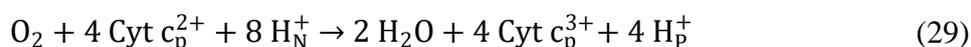
Karlsberg2⁺MD procedure

The new protocol (KB2⁺MD [7]) for computation of pK_A values uses data from MD simulations. It still follows the general philosophy for pK_A value computation introduced above, using PACs. However, here data from MD simulations replace the rather local conformational changes obtained with the standard procedure of PACs. Several steps are needed for the KB2⁺MD procedure: (1) initial modeling of the crystal structure; (2) a choice of protonation patterns; (3) preparation and performing MD simulations with the protonation patterns from (2); titration of the frames extracted from all trajectories; (3) averaging of the computed electrostatic energies for each trajectory and finally combining the averaged results to obtain pK_A values with equation (18). Computational details that differ from the standard Karlsberg2⁺ procedure are described in the chapter 1.2 (section “*Molecular dynamics and pK_A computations*”).

1.5. Cytochrome c oxidase

The chemical energy conversion from aerobic respiration is the fundament of life in many living creatures. The energy released from chemical decomposition of nutrients and the reduction of oxygen is stored by synthesizing ATP (adenosine triphosphate), which acts as the “currency” for energy transfer within the cell. The last steps of aerobic respiration are carried out at the respiratory chain, a system of protein complexes (I to V), located on the inner membrane of mitochondria of eukaryotes and in some bacteria. Electrons obtained from other metabolic processes (glycolysis, citric acid cycle, etc.) are used by complexes I to IV to pump protons from the negatively charged (the N-side) to the positively charged (the P-side) side of the membrane. ATP-synthase (complex V) utilizes the created electrochemical gradient to for ATP synthesis from phosphate and ADP (adenosine diphosphate). [26,27]

Cytochrome c oxidase (CcO) (Figure 1.4), or complex IV, is an enzyme that catalyzes the reduction of molecular oxygen (O₂) and utilizes the released energy for the proton pumping. The reaction can be summed as follows:



In steps, 4 electrons are donated from the soluble cytochrome *c* from the P-side to the CcO (Figure 1.4). The electrons tunnel via CcO cofactors, first to bimetallic Cu_A and subsequent via heme a to the binuclear center (BNC), which is a collective name for heme a₃ and Cu_B-complex where the reduction of oxygen takes place. Coupled with the electron transport, protons are taken up from the N-side. Four protons are involved in the chemical reaction (called chemical protons) and four are pumped across the membrane (pumped protons) to the P-side against the electrochemical gradient (Figure 1.4).

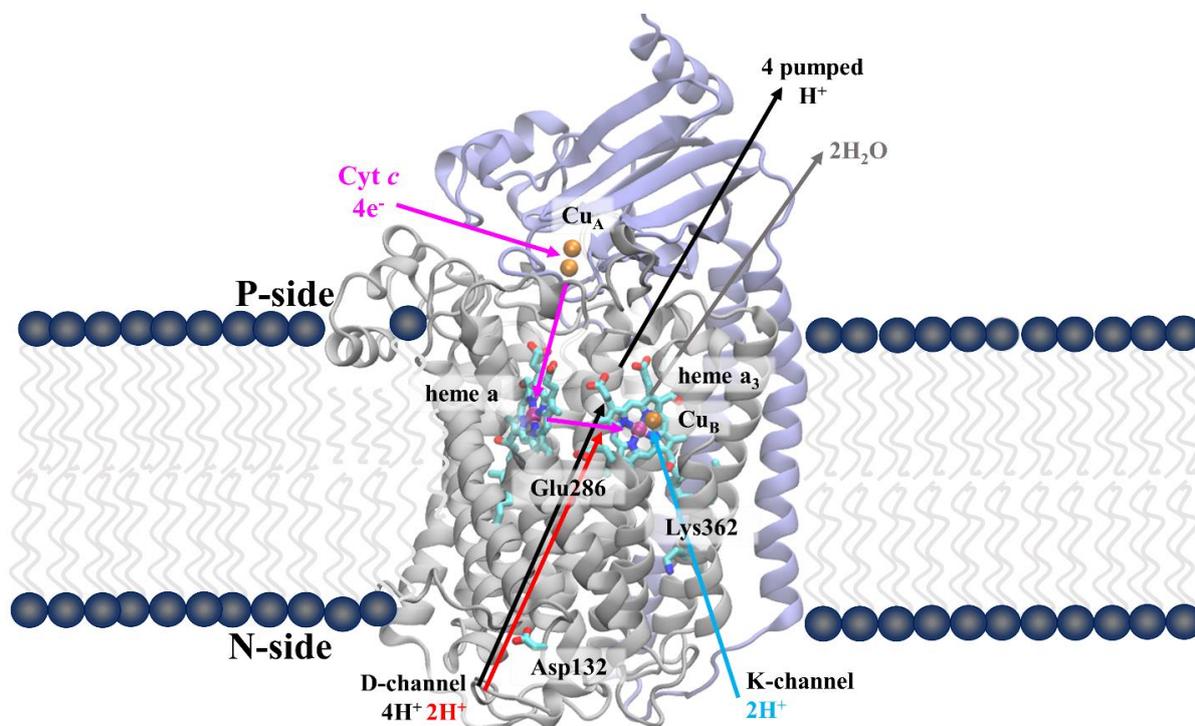


Figure 1.4. Crystal structure of cytochrome *c* oxidase (CcO) from *R. sphaeroides* (PDB: 2GSM[28]). Movement of electrons, protons and water is shown approximately, as indicated by arrows (chemical protons via K- (blue) and D- (red) channel, electron tunneling (magenta), pumped protons via D-channel (black), water exit (grey)).

CcO is a widespread enzyme among different life forms, however the CcO complexes from different species are not identical. There are three known evolutionary related groups of CcO: A-, B- and C-families. The CcO from these families have similar structures and thermodynamic properties but there are important structural and functional differences [29]. In this work, A-type CcO from two different species is investigated. In chapter 4, the X-ray crystal structure of CcO from *Paracoccus denitrificans* (PDB: 3HB3 [30,31]) with the resolution of 2.25 Å was investigated, since the experiments preceding computational work were done with the CcO

from this species. In chapter 3, the X-ray crystal structure of CcO from for *Rhodobacter Sphaeroides* (PDB: 2GSM[28]) (Figure 1.4), with a better resolution of 2.0Å was used. In their native form, both enzymes consist of four protein subunits, from which two are catalytically active [32]. Experiments discussed in chapter 4 were done with these two subunits. In this chapter, the numbering of residues is adopted from CcO from *R. sphaeroides*, although the two structures have the same conserved and functionally relevant structural elements. There are other A-type CcO X-ray crystal structures, solved with different number of subunits and redox states, some of which are: a large 13 subunit structure from bovine heart (PDB: 1V55 [33], (resolution 1.9Å)), from *R. Sphaeroides* (PDBs: 3FYE [34] (resolution 2.15Å), 1M56 [35] (resolution 2.3Å)) and from *P. denitrificans* (PDB:1AR1 [30], (resolution 2.7Å)).

CcO is a proton pump and there are structural elements that such an enzyme must contain: (1) the active site, where the chemical reaction takes place, (2) proton conducting pathways for uptake and exit, (3) the proton-loading site that changes pK_A values coupled with the redox states of cofactors and (4) gating mechanism to prevent backflow of protons. Finally, all four pump elements together enable CcO to pump protons across the membrane, in a process mostly governed by electrostatics [11,36]. The pump elements are briefly discussed in the following text.

The active center of CcO is the BNC, where the chemical reaction takes place. In the BNC, iron of the high-spin heme a₃ and the Cu_B-complex are involved in the oxygen splitting at the beginning of the redox-cycle (Figure 1.5). The Cu_B is ligating three histidines, His334A, His333A and His284A. The latter is covalently bound to Tyr288A (the catalytic tyrosine). In every step of the cycle, the BNC has a different chemical composition, since not all electrons arrive simultaneously, shown in Figure 1.5. In A-type CcO, for each electron, one proton is used for the chemical reaction and one proton is pumped across the membrane, i.e. with 1 H⁺/e⁻ pumping stoichiometry [11]. The oxidative part of the CcO cycle starts with the P_M state. It continues with the F state followed by the O_H state and from O_H to the E_H and R states in the reductive part of the reaction cycle. After the R state the cycle is completed, and two water molecules are created.

Protons are translocated (Grotthuss mechanism of transport [37]) via two input channels, the D- and K-channel in A-type CcO [8,38]. The D-channel is named after the conserved residue Asp132A, located at the channel entrance on the N-side. Two chemical protons are transported via the D-channel (oxidative part of the cycle) as well as all four pumped protons (Figure 1.4

and Figure 1.5). In the reductive part of the redox cycle, the K-channel, named after a conserved Lys362A, is translocating the other two chemical protons. In this work, protonation states of residues close to the K-channel entry and structural changes potentially related to activation of the K-channel were investigated (chapter 4, “*Cytochrome c Oxidase near the K-channel entry*”). The exit channels for protons and water are poorly investigated, however some suggestions have been made [39].

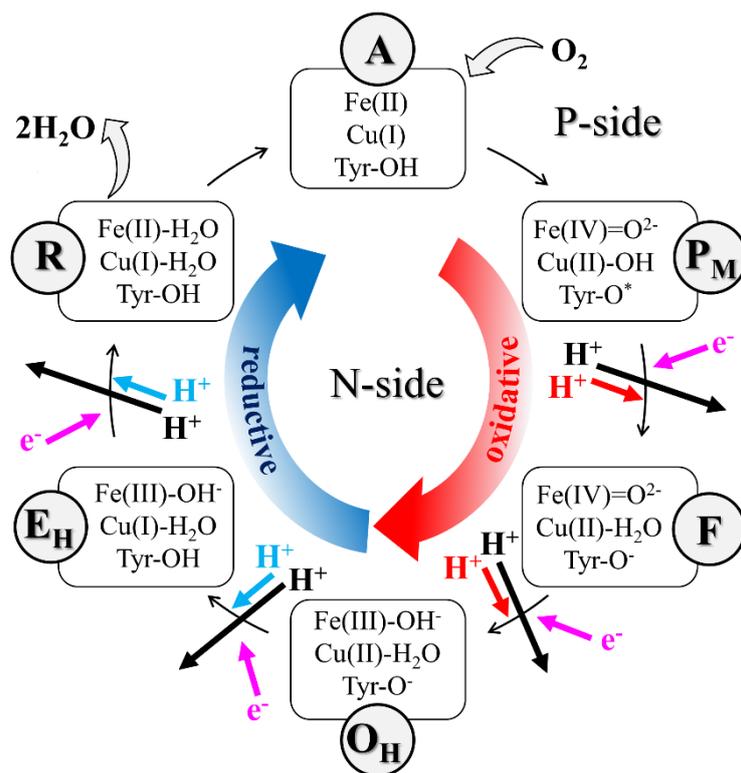


Figure 1.5. Catalytic cycle of CcO. The description of states is taken from [8,11]. Chemical protons are translocated via the D-channel in the oxidative half (red H⁺) and via the K-channel in the reductive half (blue H⁺) of the cycle. Pumped protons (H⁺) are all transported via the D-channel in the entire catalytic cycle (black H⁺). Electron (magenta) and proton uptake and proton pumping are indicated by the arrows. The figure is adapted from [8].

The proton-loading site (PLS) is the most elusive pump element. In short, CcO has to overcome the energy barrier to translocate the pumped proton against membrane electric potential of about 200mV [40] together with the proton gradient, resulting in the pH difference [41,42]. The PLS is a group or a residue, involved in proton translocation, and lowers the energy barrier of proton translocation by changing its pK_A value coupled with redox changes of CcO cofactors. Several potential candidates have been suggested to act as the PLS [43–46]. In this

work, propionates of heme a and heme a₃ were investigated as PLS candidates (chapter 3, “*Proton-Loading Site in Cytochrome c oxidase*”).

The gating mechanism is needed to ensure the unidirectional proton transfer across the membrane and it can be purely kinetic [41] and/or involve conformational changes of gating residues [8,9,47,48]. At the end of the D-channel, Glu286A (Figure 1.4) is identified as a part of the gate (or valve) by changing its conformation (“up” or “down”) and by fast reprotonation [9,48]. Gating mechanism in the D-channel and loading and unloading of the PLS must also be coupled in order to prevent proton backflow. In the K-channel, Lys362A is involved in an electrostatically activated gating, as well by changing its conformation coupled with redox states of the BNC, which practically activates the K-channel [8].

2. Improving pK_A computations with Karlsberg2⁺

Methods, improvements, computations and results presented in this section were performed and obtained by myself utilizing the software Karlsberg2⁺ [7]. The results in this part are unpublished. A major part of the software Karlsberg2⁺ was created by my colleague Tim Meyer [7]. However, modules involving pK_A computations were written by Tim Meyer and myself. In addition, all future software improvements and maintenance were under my responsibility until the end of my doctoral work. Prof. Dr. Ernst Walter Knapp was the supervisor of the project.

2.1. Introduction

The usual approach to pK_A computations in proteins is to use the crystal structure of the protein obtained from the Protein Data Bank (PDB) [49]. This structure represents only a protein conformation that corresponds to the pH value of crystallization. When pK_A values are determined experimentally, the protein population in the sample will be exposed to a range of pH values, which will allow individual protein molecules to adapt their conformation to the pH value throughout the titration. This effect is not properly taken into account in computations if only a single structure of the protein is used. The influence of pH on the protein conformation can range from subtle, like hydrogen bond network rearrangements as hydrogen atoms react easily to polarity changes in the environment, to larger changes like side-chain reorientations and even global conformational changes. Therefore, in order to obtain reliable results, multiple conformations of the protein have to be used when computing pK_A values. The software Karlsberg⁺ [16] (KB⁺) developed in 2008, by Gernot Kieseritzky takes into account multiple protein conformations, called pH adapted conformations or PACs. These PACs are protein conformations that are modeled to be appropriate for a specific pH value. In KB⁺ [16], these include changes in the hydrogen bond network, and side-chain orientation changes of residues involved in salt-bridges. Salt-bridges in proteins are expected to open at pH values below 4 or above 10. If the salt-bridges are not broken while modeling pH adapted conformations at low and high pH values, the resulting pK_A values will be shifted dramatically. For acids the pK_A value is down-shifted, while for bases it is up-shifted. Neither modeling of the hydrogen bond network nor opening of salt-bridges is an easy task. Generating appropriate conformations of residues involved in salt-bridges at pH values far below or above pH 7 required extensive work over the years [18,50,51]. In the procedure integrated in KB⁺ [16], many protein conformations

are modeled in order to find only few that could properly describe these changes. This is unfortunately costly in CPU time, especially if computing pK_A values of very large proteins. In 2014, a new version of the software Karlsberg2⁺ (KB2⁺) was created by Tim Meyer, which is the project I joined later on and where I developed several integral program modules. The development of KB2⁺ [7] was an opportunity to improve pK_A computations by developing a more efficient modeling procedure for salt-bridge opening.

Protonation changes, as a consequence of a pH change can lead to large conformational changes of the protein structure. It is possible that a protein structure suffers partial unfolding at low pH values, opening the interior of the protein towards the solvent. Unfortunately, neither KB2⁺ nor KB⁺ account for potential global conformational changes of the protein structure in the main workflow. This problem is solved by combining MD simulations with electrostatic energy computations in order to get better sampling. In 2015 a very efficient method, called Karlsberg2⁺MD, was developed by Tim Meyer. Details are explained in chapter 1.4. The pK_A values computed with Karlsberg2⁺MD agree better with the corresponding experimental values of the benchmark set of 194 pK_A values in comparison to the PAC procedure of KB2⁺ (and KB⁺). However, Karlsberg2⁺MD is much more CPU time demanding. Therefore, in order to obtain results that are slightly less precise but at much greater speed, the former approach is still a suitable option. In this chapter, modeling of all PACs, proper testing of the software and comparing it to the pK_A values of the benchmark set mentioned above is presented.

2.2. Methods

Benchmark set of experimentally determined pK_A values

The protocols for pK_A computations are validated using a benchmark set of 194 measured pK_A values of 13 proteins as reference [7]. This benchmark set differs slightly from the original benchmark set used for testing of KB⁺ [16]. Comparing results of KB⁺ and Karlsberg2⁺MD, (KB2⁺MD), pK_A values were taken from 11 proteins of the original benchmark set of 15 proteins [16] used for KB⁺. The original benchmark set used for KB⁺ evaluation contained 15 proteins. Testing KB2⁺MD the new benchmark set contains two additional proteins which were used in similar approaches involving conformational sampling by MD simulation [52]. The total set of 13 proteins is listed in Table 2.1. (PDB codes: 4PTI [53], 1PGA [54], 1A2P [55], 2LZT [56], 3RN3 [57], 2RN2 [58], 1HNG [59], 3ICB [60], 1PPF [61], 1ERT [62], 1XNB [63] and the two additional proteins 2ZTA [64] and 3BDC [65]). Most of the pK_A values in this set were collected from literature by Georiesky et al. [50], but some are from other works: xylanase (PDB code: 1XNB) from Joshi et al. [66], leucine zipper (PDB code: 2ZTA) from Matousek et al. [67] and for a staphylococcal nuclease (SNase) Δ+PHS variant (PDB code: 3BDC) from Castaneda et al. [65].

Some protein structures required specific preparations in addition to the standard protocol of initial modeling of protein structures in KB2⁺. Prior to the pK_A computation, crystal water, ligands and ions were removed from the protein crystal structures. In the crystal structure of Δ+PHS SNase variant used in this work (PDB code: 3BDC) the first six and last eight residues are disordered. Instead of remodeling them, the C-terminus was methylated and N-terminus was acetylated to avoid interference of charges in locations where they technically do not occur. In case of T-lymphocyte adhesion glycoprotein CD2 (PDB code: 1HNG), only part of the crystal structure has been used for pK_A measurements, therefore only that part was used in our computations. The third domain of the turkey ovomucoid inhibitor (PDB code: 1PPF) contains the protein PMN elastase, therefore only chain I was used for the pK_A computations. Since the structure of the leucine zipper (PDB code: 2ZTA) was modified prior to pK_A measurements, the structure needed adjustment adding a glycine to the N-terminus and a glutamate to the C-terminus. In the crystal structure of xylanase (PDB code: 1XNB) the coordinates of the atoms ND2 and OD1 of the residue Asn35 belonging to the amide group were interchanged, as it was shown that Asn35 has a large influence on the pK_A value of Glu172 in the neighborhood [7]. A special case is thioredoxin (PDB code: 1ERT) that contains a buried water molecule 136 in

chain A and in close contact with Asp26. This water molecule was kept in all computations. All side-chains in the structure 1ERT in were energy minimized by using CHARMM [21], first in vacuum and subsequently with an implicit water model using GBSW [68], as it was done in previous benchmark computations [7,16].

Benchmarking

In order to compare pK_A values to the experiment a simple but yet appropriate method should be used. In many cases, the computed shifts of pK_A values from the value in solution are not very large and a concept of root mean square deviation (RMSD) can be used for evaluation of accuracy. The root mean square deviation is defined as $RMSD = \sqrt{\sum_{i=0}^N (pK_{A,i}^{exp} - pK_{A,i}^{comp})^2}$, between the experimentally measured and the computed pK_A values. In some cases, computed pK_A values were outside of the considered pH interval [-10, 20]. These residues have been excluded from the RMSD calculation, as explained in Table 2.1. Individual RMSD values for each protein in the benchmark set were calculated as well as a total RMSD for the entire set of 194 pK_A values. These results were compared to the values obtained with KB^+ and $KB2^+MD$, which were not recomputed in this work but rather taken from the work of Tim Meyer [7].

pK_A computations

The details of the protocol for pK_A computations with $KB2^+$ was introduced in the chapter 1.4. Just like the protocol for the previous version, KB^+ , it involves several steps. Also here the above mentioned necessary individual modeling steps were applied to the crystal structures, before the conventional preparation of the protein structures was performed. The latter involves the automatic creation of PACs at pH values of -10, 7 and 20. After generating PACs the results of electrostatic energy computations are combined to obtain pK_A values of titratable groups [16].

There is a crucial difference in PAC generation between $KB2^+$ and KB^+ , which was one of the main goals of this project. In KB^+ , 11 PACs are created, while in $KB2^+$ three PACs were enough for obtaining good results. The PAC at pH 7 is generated in the same manner in both versions of the software, simply optimizing the hydrogens by energy minimization, as previously explained. The details and comparison of creation of PACs at pH values of -10 in KB^+ and 20 are described in the following paragraphs.

In the published work on Karlsberg⁺[16] the PAC at pH -8 is used, instead of the PAC at pH -10. The final available version of the software Karlsberg⁺ was updated to use the PAC at pH -10 as well. Modeling of the PAC at pH -8 in KB⁺ and the PAC at pH -10 in KB²⁺ is identical; however, a more comprehensive pH range has been used for titration in the final version of KB⁺ and was adopted in KB²⁺.

Salt-bridge opening in Karlsberg⁺

For generating PACs in KB⁺ at pH values of -10 and 20, salt-bridges are perturbed as mentioned in the chapter 1.4. For a residue pair involved in a salt-bridge the side-chain conformations are randomized one by one. Next, all atoms belonging to the randomized side-chains are optimized by energy minimization with the remaining part of the protein being fixed. Since the resulting open salt-bridge conformation may energetically be unfavorable, 30 different conformations are generated. The total CHARMM energy obtained in a homogenous dielectric medium with $\epsilon=1$ is used to select the "best" open salt-bridge conformation for this residue pair. This procedure is performed for all residues involved in a salt-bridge except for the ones which are completely buried such that the salt-bridge cannot be opened alone by local conformational changes. For the resulting preliminary PAC all hydrogen atoms are optimized in a homogenous dielectric medium with $\epsilon=1$ and the total conformational energy is calculated using CHARMM [21]. The whole procedure is repeated five times for the pH values of -10 and 20, respectively. Finally the PAC with the lowest energy is selected as the appropriate PAC for the corresponding pH value [16]. This procedure is very CPU time demanding. As described in the chapter 1.4, with inclusion of a PAC at pH 7, 11 PACs are needed for the pK_A computations.

Modeling of the salt-bridge opening in Karlsberg²⁺

In the following it is shown how one can avoid the CPU time demanding procedure of KB⁺ where for the pH values of -10 in KB²⁺) and 20 several PACs need to be generated. The aim is a modeling procedure where only a single PAC per pH value is necessary for yielding good agreement with experimental pK_A values for the benchmark set.

The stability of a salt-bridge is due to strong electrostatic interaction between the side-chains of negatively charged acids and positively charged bases, whose distance of heavy atoms is 4Å and shorter. The first attempts of salt-bridge opening in this work were performed by energy minimization of side-chains of the residue pair involved in the salt-bridge, where one of the residues is in the charge neutral state. At pH -10 this is the acidic residue and at pH 20 the basic

residue. However, the resulting conformations did not yield good results because the salt-bridges often did not properly open. This led to the idea of putting an artificial charge on one of the residues that equals the charge of the other such that they repel each other.

Therefore, a modeling procedure named “repelling charges” was introduced. It involves two steps: (1) opening of the salt-bridge by performing geometry optimization of involved side-chains using “repelling charges” followed by (2) geometry optimization with charges corresponding to the pH of a PAC to obtain a chemically reasonable structure (Figure 2.1) Both steps are done by using CHARMM [21] with $\epsilon=1$.

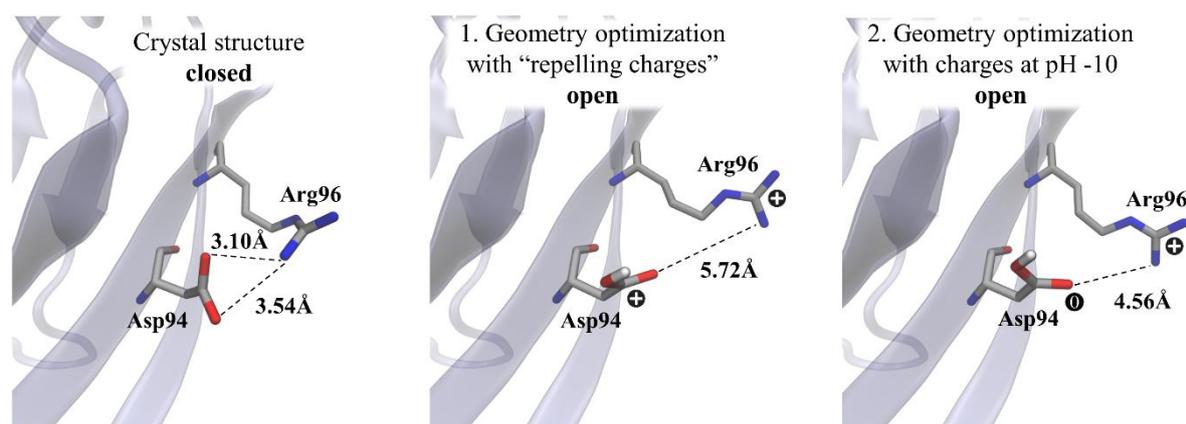


Figure 2.1. “Repelling charges” modeling the opening of salt-bridges for the ion-pair Asp94 and Arg96 in the crystal structure of the protein rat T-lymphocyte adhesion glycoprotein (PDB code: 1HNG [59]), for the PAC at pH -10. **Left:** The initial conformation of the side-chains found in the crystal structure. **Middle:** Optimized side-chains geometry after the charge modeling of Asp94 (charge pattern in Figure 2.2, right). **Right:** Conformations of side-chain after the geometry optimization with chemically appropriate charges according to pH -10, where Asp94 is neutral (charge pattern in Figure 2.2, left), in order to obtain a correct structure.

In case of a PAC for the pH value of -10, an artificial positive charge is placed on Asp, Glu side-chains and N-terminus involved in salt-bridges to achieve repelling with bases that are in this case positively charged. In case of a PAC at pH value of 20, an artificial negative charged was placed on Lys, Arg and His side-chains involved in salt-bridges that are repelled by already negatively charged acidic salt-bridge partners. In the CHARMM force field [21], the charge of the side-chain of amino acids is not localized on one atom but rather distributed over several atoms of the side-chain. In order to introduce a new charge state for this modeling procedure, specific atoms had to be chosen, as shown in Figure 2.2 and Figure 2.3. In all versions of the Karlsberg⁺ software, the titratable hydrogens of residues are present also in the deprotonated

state as shown in Figure 2.2 and Figure 2.3. The deprotonated state is modeled by placing a vanishing charge on the corresponding titratable hydrogen. Glu and Asp are for instance modeled with hydrogen atoms attached to both acidic oxygens, but only one carries a non-vanishing charge in the protonated state. Here, the hydrogen atoms carrying no charge are in reality not present. Nevertheless, they are still subject to Lennard-Jones interactions. This introduces an artefact, which is considered to be negligible.

Here, the neutral charge states were used as a template as one of the salt-bridge partners is in the neutral state at these two extreme pH values. In the case of acidic side-chains, one excess positive charge is imposed on the carboxylic carbon atoms, while keeping the atomic partial charges of other atoms of the residue intact, as seen for glutamate and aspartate in Figure 2.2. The exception here is the C-terminus (Appendix A, Figure A.1), where there are no hydrogen atoms present, as it is protonated implicitly.

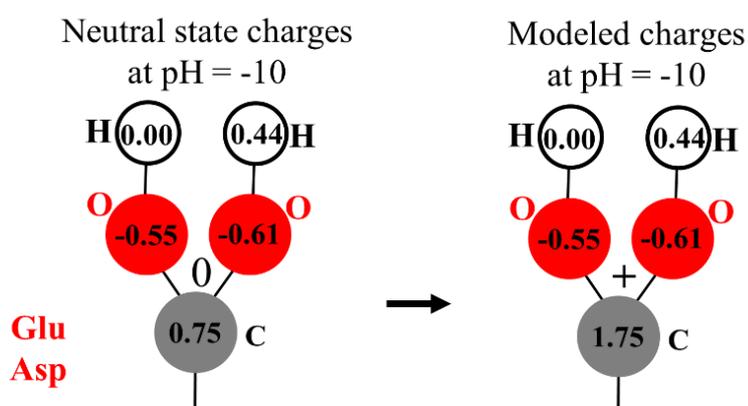


Figure 2.2. Atomic partial charges of titratable residues for PAC at pH -10 using “repelling charges” for the acidic residues, Asp and Glu. **Left:** atomic partial charges of the neutral state and; **Right:** charges with an additional unit charge at the carboxylic carbon, resulting in a total positive charge. The full set of atomic partial charges of titratable residues can be found in the Appendix A.

For basic residues, a unit negative charge is spread over the basic nitrogen atoms while the other charges are not changed, as it is shown in case of arginine (Figure 2.3). Similarly, the atomic partial charges of the lysine are remodeled (Appendix A, Figure A.2). The exception to this is histidine because it has two tautomeric neutral states, where a non-vanishing charge is placed on the hydrogen atom at the δ -N atom or ϵ -N atom in KB^{2+} , leading to an asymmetric charge pattern in the imidazole ring. To simplify the charge modeling procedure, the protonated state was used for the “repelling charge” model, but two additional negative unit charges were

placed at the δ - and ϵ -nitrogen atoms of the imidazole ring, resulting in a negative total charge of histidine (Appendix A, Figure A.3). After placing the excess charges at the residues involved in salt-bridges, geometries of these residues are optimized in vacuum ($\epsilon=1$), which opens the salt-bridges due to electrostatic repulsion (Figure 2.1, middle).

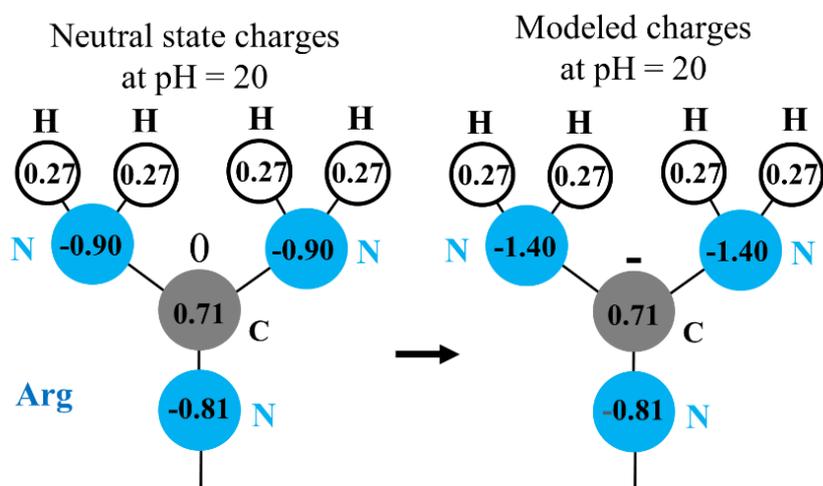


Figure 2.3. Atomic partial charges of titratable residues for PAC at pH 20 using “repelling charges” for basic residues, e.g. Arg. **Left:** atomic partial charges of the neutral state; **Right:** charges with an additional unit charge spread over two nitrogen atoms of the guanidinium group of arginine, resulting in a total negative charge. The full set of atomic partial charges of titratable residues can be found in the Appendix A.

Subsequently, all previous charge changes are disabled using the neutral charge state at pH -10 for acidic residues and at pH 20 for basic residues. Then another geometry optimization is performed at $\epsilon=1$ for all side-chain atoms of residues involved in salt-bridges to obtain a chemically valid structure with open salt-bridges (Figure 2.1, right). This is done in order to correct the conformations of side-chains as they could form irregular interactions with the environment in the previous step while having imposed artificial charges.

Tyrosine is a residue that normally does not form salt-bridges at physiological pH values as it is not ionized. However, if tyrosine forms a strong hydrogen bond with the side-chains of basic or acidic residues it was included in the modeling procedure, since it may deprotonate at high pH values. This was done before also in KB^+ [16]. Charge modeling of the tyrosine side-chain is done for the PAC at pH -10, where it is expected to be in the neutral state. In this case, an excess positive charge is placed on the oxygen atom of the phenyl group of Tyr, resulting the

total positive charge of the residue (Appendix A, Figure A.1). By performing an energy minimization, the hydrogen bond of Tyr with the other residues will be broken.

Arginine is a special case for benchmarking pK_A computations because there are practically no experimental pK_A values available for this amino acid. Therefore, it is normally excluded from pK_A computations. On the other hand, arginine is a frequent salt-bridge partner of acidic residues, which is the reason why it is included in salt-bridge opening procedures. Preliminary testing showed that for PACs at pH 20, the pK_A values of acidic residues involved in salt-bridges with an arginine may deviate strongly from the measured values, if these salt-bridges are not opened.

2.3. Results and Discussion

pK_A computations

A complete overview of pK_A values computed for the benchmark set using $KB2^+$ is shown in Table 2.1. Here, the “repelling charges” approach was used to open salt-bridges for PACs at the pH value of -10 and 20. The total pK_A -RMSD is 1.17 (Appendix A, Table A.1), which perfectly matches the previously obtained result with the older version of the software KB^+ . $KB2^+MD$ yields a pK_A -RMSD of 0.79. But, the latter approach is much more CPU time intensive as will be shown later. Overall, the methods used for pK_A computations are evaluated by observing the pK_A -RMSD for the benchmark set, but when comparing individual protein pK_A -RMSD values, no clear trend can be observed.

The good agreement in terms of the pK_A -RMSD value for the entire benchmark set indicates that the open salt-bridge geometries describe the conformations of proteins at PACs at pH values of -10 and 20 well enough. The approach of “repelling charges” was developed to open salt-bridges, as described in the methods section. The atoms whose charges are changed in this approach, do not interact directly with atoms of other residues in the surrounding. Hence, the interaction scenario is only moderately changed by these artificial charges. Thus, avoiding unrealistic side-chain conformations, which may not be repaired in subsequent modeling steps.

One of the main goals of this work was to generate a flexible software with an approach similar to KB^+ and if possible with improving the accuracy of the computation. In the former version (KB^+), many different conformations of side-chains involved salt-bridges were randomized and

subsequently optimized, which is similar to the approach in this work, in which after applying “repelling charges” a subsequent geometry optimization by performing an energy minimization with chemically reasonable charges follows. In both cases, the protein environment around the salt-bridge is kept rigid. Proper opening of a salt-bridge may require more than just changes of hydrogen bonds and side-chain conformations of the residues forming the salt-bridge. Such changes could be local conformational adjustments of the protein backbone and of the other side-chains in the vicinity of the salt-bridge. This is effectively solved in KB²⁺MD by MD sampling, where the salt-bridges are opened and the environment is adjusted. The time frames are then used to compute pK_A values, as described in the chapter 1.4.

Table 2.1. Root-mean square deviations (RMSDs) in pH units between measured and computed pK_A values of titratable residues from a benchmark set containing 13 proteins and 194 pK_A values. The pK_A-RMSD values are obtained with KB⁺, a new version KB²⁺ and the protocol that involves the MD approach called KB²⁺MD. The pK_A values of KB⁺ and KB²⁺MD were taken from previous work [7].

Proteins					RMSD between measured and computed pK _A values		
No.	Protein name	PDB code	No. of pK _A ^a	No. of atoms ^b	KB ⁺	KB ²⁺	KB ²⁺ MD
1.	lysozyme	2LZT	20	1960	1.08	0.93	0.92
2.	pancrea trypsin inhibitor	4PTI	14	892	0.95	0.92	0.65
3.	bovine ribonuclease A	3RN3	14 ^d	1856	0.77	0.80	0.61
4.	ribonuclease H	2RN2	20 ^e	2455	1.67	1.32	0.84
5.	barnase	1A2P	12 ^c	1727	1.03	1.73	0.68
6.	rat T-lymphocyte adhesion glycoprotein	1HNG	14	1576	1.09	0.59	0.81
7.	Ca binding protein	3ICB	19	1202	0.95	1.08	0.70
8.	ovomucoid inhib OMTKY3	1PPF	11	418	0.84	1.01	0.65
9.	thioredoxin	1ERT	17	821	1.22	1.59	1.16
10.	xylanase ^f	1XNB	7	1448	1.47	0.90	1.13
11.	streptococcal protein G	1PGA	15	855	0.94	0.84	0.58
12.	GCN4 leucine zipper ^g	2ZTA	16/16	1120	1.22/1.13	1.08/1.22	0.80/0.85
13.	staphylococ. nucl. (Δ+PHS)	3BDC	15	2101	1.65	1.64	0.50
all residues			194		1.17	1.17	0.79

^a number of pK_A values computed per protein

^b number of protein atoms

^c Asp75 was excluded from the computation [16]. The obtained pK_A value is very low and the measured value (3.1) is obtained under unfolding conditions [69].

^d His48 was excluded from the computation as it suffers a greater conformational change coupled with Gln101 [16]

^e Asp10 was excluded, since it is a part of the Mg²⁺ binding site [16].

^f in case of xylanase the pK_A values were computed after interchanging O and N atoms of Asn35, as described above.

^g The structure of Leucine zipper is dimeric, therefore the pK_A values were computed for two monomers, as presented with “/”, however the overall RMSD values is averaged over the results from the two chains. The number of atoms refers to the entire dimeric structure.

The CPU time

The invested CPU time in the pK_A computations is an important factor, as it can range from minutes for a quick computation to several days for complex, but usually more accurate procedures. In this case, the main issue with the old procedure in KB^+ is the CPU time for PAC generation and the opening of salt-bridges. In Table 2.2 the CPU times are shown for pK_A computations of the proteins from the benchmark set for the procedure KB^+ and $KB2^+$, when performing computations on CPU cores of the same hardware quality (Intel Xeon 5670, 2.93GHz).

Table 2.2. CPU time (in minutes) needed to compute pK_A values of titratable residues from the benchmark set containing 13 proteins and 194 pK_A values for two versions of the Karlsberg⁺ software, $KB2^+$ and KB^+ . Generation of PACs and electrostatic energy computations were done on the CPU core of the same hardware quality (Intel Xeon 5670, 2.93GHz).

No.	Protein name	Proteins				CPU time (min)		
		PDB code	No. of pK_A^a	No. of atoms ^b	No. of salt-bridge residues ^c	KB^+ (t_1)	$KB2^+$ (t_2)	factor ^d t_1/t_2
1.	lysozyme	2LZT	20	1960	7	638	78	8.2
2.	pancrea trypsin inhibitor	4PTI	14	892	4	517	36	14.4
3.	bovine ribonuclease A	3RN3	14	1856	11	1419	84	16.9
4.	ribonuclease H	2RN2	20	2455	19	2398	132	18.2
5.	barnase	1A2P	12	1727	16	1012	84	12.0
6.	rat T-lymphocyte adhesion glycoprotein	1HNG	14	1576	19	1661	63	26.4
7.	Ca binding protein	3ICB	19	1202	4	429	57	7.5
8.	ovomucoid inhib OMTKY3	1PPF	11	418	5	319	27	11.8
9.	thioredoxin	1ERT	17	821	7	737	81	9.1
10.	xylanase	1XNB	7	1448	13	2211	105	21.1
11.	streptococcal protein G	1PGA	15	855	8	286	27	10.6
12.	GCN4 leucine zipper	2ZTA	16/16	1120	12	913	78	11.7
13.	staphylococ. nucl. (Δ +PHS)	3BDC	15	2101	23	4741	93	51.0

^a number of pK_A values computed per protein

^b number of protein atoms

^c total number of amino-acid side-chains involved in salt-bridges. In some cases, instead of an ion pair, several residues fulfill geometric criteria for salt-bridges by forming what could be interpreted as a small cluster. Tyrosine residues forming hydrogen bonds with acidic or basic residues are also included.

^d factor by which is the pK_A computation decreased in terms of CPU time when using $KB2^+$ compared to KB^+ .

For most proteins, the CPU times needed for pK_A computation is reduced by the factor of 16 on the average when $KB2^+$ is used instead of KB^+ , which is a considerable improvement. This is mainly due to the more efficient procedure for the salt-bridge opening when modeling PACs in $KB2^+$. In KB^+ several steps are needed to model an open salt-bridge: generation of 30

randomized side-chain conformations of the residue-pairs involved in salt-bridges, followed by optimization of hydrogens and the affected side-chains and conformational energy computations, as described above. For proteins with many residues in salt-bridges, this procedure becomes much slower, i.e. more modeling of open salt-bridges has to be performed. This is best demonstrated for staphylococcal nuclease (SNase) Δ +PHS variant (PDB code: 3BDC) where KB^+ is about a factor of 51 slower than KB2^+ . On the other hand, for the Ca binding protein (PDB code: 3ICB) this factor decreases 7.5 (see Table 2.2).

Procedure $\text{KB2}^+\text{MD}$ is by far the most precise method developed in our group to compute pK_A values, as observed in a very small pK_A -RMSD value. But, it requires a large amount of CPU time. As described in the chapter 1.4, it requires multiple MD simulations of at least 10ns, followed by titrations of at least 90 extracted structures from the MD time frames. In this case, it is a compensation between the CPU time and accuracy, which can be important depending on the task and the system.

2.4. Summary and Conclusions

In the study presented in this chapter, a protocol to compute pK_A values has been improved as a part of the new software version Karlsberg2⁺ (KB2^+). In KB2^+ , as in its predecessor Karlsberg⁺ (KB^+), pK_A values are computed from electrostatic energies by solving the Poisson Boltzmann Equation. The improved procedure follows the same concept established in KB^+ , using pH adapted conformations, starting from the protein crystal structures (details in chapter 1.4). This is needed in order to obtain good agreement with measured pK_A values, especially for residues whose side-chains are involved in salt-bridges. In this work, modeling of PACs at pH values of -10 and 20 was improved, optimizing the stepwise procedure of salt-bridge opening, which was implemented in a the new version of the software, KB2^+ . The open salt-bridge conformation is obtained by using “repelling charges”, as described in the methods section, where in a first step the ion pair is broken by electrostatic repulsion and then brought into a chemically reasonable open salt-bridge conformation. The performance of KB2^+ was evaluated for a set of 13 standard proteins with 194 measured pK_A values, also called standard benchmark set, which was used in previous developments of pK_A computations performed in our group [7,16]. The results demonstrate no improvement of computed pK_A values, since the total pK_A -RMSD is with 1.17 the same as with the preceding KB^+ software (Table 2.1). The

lack of improvement in terms of total pK_A RMSDs is to be expected, since conceptually $KB2^+$ and KB^+ are equivalent. However, technically they differ, since $KB2^+$ is much more efficient than KB^+ . In KB^+ , several modeling steps including geometry optimization and conformational energy evaluations are needed to model PACs with pH of -10 in $KB2^+$ and 20, which is reduced to only two geometry optimizations in $KB2^+$, making it about 16 times faster on the average. This is also a great advantage in comparison to the $KB2^+MD$ procedure, which requires a much larger amount of CPU time. Overall, $KB2^+$ is a good choice for fast pK_A computations in proteins, with reasonable accuracy. However, for better agreement with measured pK_A values the $KB2^+MD$ procedure should be used, where MD simulations of different protonation pattern provide the necessary conformational sampling yielding a pK_A -RMSD of 0.79.

Further improvements could be attempted in the future. Some of the possibilities are: geometry optimization of all titratable residues with the dielectric constant of 4; using a generalized Born model for the solvent in the subsequent optimization of hydrogens; remodeling of the charges of titratable residues used in electrostatic energy computations, as suggested in previous works [7]; adding an additional PAC that would serve to better define histidine protonation.

Karlsberg2⁺ was designed as an improved version of KB^+ . This involved combining the existing modules that perform different tasks into a main workflow that can compute pK_A values. In this way, it became a very flexible software that can separately utilize modules for electrostatic energy computations, molecular modeling or computation of pK_A values of proteins. These features and independent modules have been used in this doctoral work for different projects.

3. Proton-Loading Site in Cytochrome *c* Oxidase

Computational work and analysis of data were done by the author of this thesis. The results from this work are unpublished. Preparation of the manuscript and the publication of results are planned. Prof. Dr. Ernst Walter Knapp was the supervisor of the project.

3.1. Introduction

Cytochrome *c* oxidase (CcO) is an important member of the respiratory chain and the primary topic of this dissertation, as described in more detail in chapter 1.5. Here, the focus is one the proton-loading site in CcO. In this project, CcO from *Rhodobacter sphaeroides* was investigated.

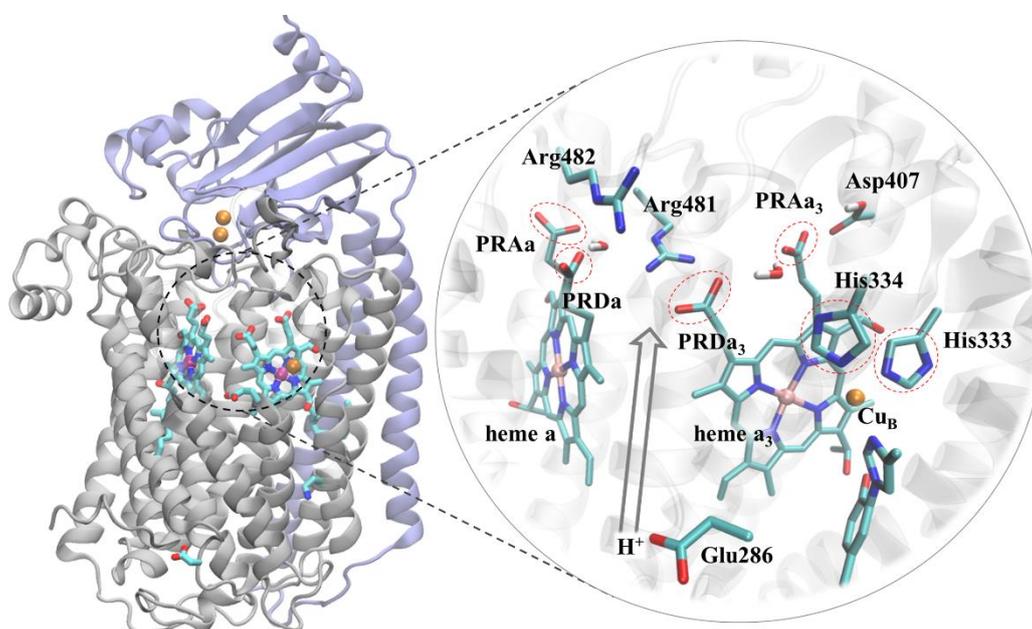


Figure 3.1. **Left:** Crystal structure of CcO with two subunits from *R. sphaeroides* (PDB code: 2GSM[28]). **Right:** The overview of the central part of CcO with the focus on some of the proposed proton-loading site (PLS) candidates (circled with red dotted lines). The gray arrow indicates the direction of the next proton to be taken up into the PLS.

Despite extensive research efforts, proton-pumping in CcO is not fully understood [11,40]. The requirements of a proton-pump are discussed in more detail in chapter 1.5. In CcO, as a stepwise process, four electrons are donated from cytochrome *c* at the positively charged side of the membrane (P-side) and tunnel through metallic cofactors, heme *a* and Cu_A to the BNC

(a collective name for heme a_3 and Cu_B -complex) (Figure 1.4 in chapter 1.5, Figure 3.1). Coupled with the electron transport, protons are translocated from the negatively charged side of the membrane (N-side) via two input channels, the D- and the K-channel. In A-type CcO [8,38], for each electron received, one proton is consumed in the chemical reaction at the BNC (named chemical proton) and one proton is pumped across the membrane (named pumped proton). Experiments suggest that a group, commonly called the proton-loading site (PLS) can hold a pumped proton in an intermediate step and release it to the P-side coupled with the redox change of CcO cofactors [47]. The PLS is needed for the proton-pump process to transfer protons against the gradient and maintain the rate of pumped protons [36,70]. The PLS helps to overcome the energy barrier of transferring a proton from the N- to the P-side by the change in its pK_A value. The required change is estimated to be ~ 5 pH units [41,42]. The exact identity of the PLS is not clear, but several candidates (Figure 3.1) have been proposed: mainly propionate-D and -A of heme a_3 [12,42,43,45,71,72] and/or histidine ligand of Cu_B -complex [39,73] or a water cluster close to the propionates [46].

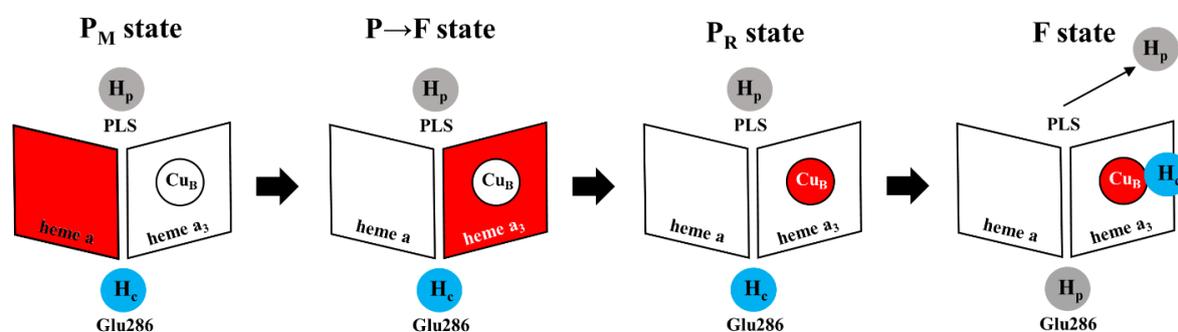


Figure 3.2. A schematic representation of the electron-proton coupled transport from P_M to F state, where one proton (H_p) is pumped (gray circle). The chemical proton is labelled as H_c (blue circle); the two parallelograms represent heme a and heme a_3 together with the Cu_B -complex (circle in right parallelogram) forming the BNC. The PLS is an unspecified group above heme a and heme B. The red color shows the location of the electron (e^-). The chemical composition of cofactors is shown in Table 3.1. Experimentally evaluated rates and equilibrium parameters can be found in [74,75].

The function of the PLS must be coupled with the sequence of events of electron-proton transfer. In Figure 3.2, the basic cycle when one proton is pumped is shown (“one proton cycle”), for the transition from state P_M to F. The one proton cycle is repeated four times, (Figure 1.5, chapter 1.5). The order of proton and electron transfer events is commonly accepted, as suggested in many works [11,43,76,77]. The P_M state is formed in the following

steps: (i) the electron (e^-) tunnels from Cu_A to heme a; (ii) The pumped proton (H_p) is transferred from Glu286 to the PLS, (iii) Glu286 is reprotonated, holding now the chemical proton (H_c); the P \rightarrow F intermediate state is formed upon e^- transfer from heme a to heme a_3 ; the P_R state is formed under highly reduced conditions [78] upon e^- transfer from heme a_3 to Cu_B complex (to form Tyr-O $^-$ from Tyr-O *). The F state is formed in the following steps: (i) The H_c is translocated from Glu286 to the BNC, (ii) Glu286 is reprotonated with the next H_p , (iii) The PLS is ready to release the proton H_p to the P-side only after the electron is neutralized with the chemical proton arriving to the BNC. After Glu286 delivered the chemical proton, there is a risk for proton backflow from the PLS. While not impossible to occur, the backflow is prevented by Glu286 itself as it is simultaneously involved in gating and delivering of the proton via the D-channel by changing its conformation and presumably quickly receiving the next proton [9,48]. Since the PLS is located above the BNC along the unidirectional proton pathway to the P-side, backflow is prevented except in the scenario described above. The arrival of the next pumped proton is considered to finally initiate the release of the proton loaded at the PLS to the P-side.

To elucidate the suitability of some of the proposed PLS candidates (Figure 3.1), propionates-A and -D of the heme a (PRAa, PRDa, respectively) and propionates-A and -D of the heme a_3 (PRA a_3 , PRDa $_3$, respectively) were investigated with molecular dynamics (MD) simulations and pK_A computations. In the crystal structure all four propionates are involved in non-covalent interactions with nearby polar groups (salt-bridges and hydrogen bonds) rendering protonation unlikely. In order to overcome this problem, alternative conformations of the local environment of propionates were modeled to enable protonation of PRAa, PRDa $_3$ and PRA a_3 . The differences of the equilibrium pK_A values between different states of CcO (Figure 3.2) indicated that PRDa $_3$ and PRA a_3 are coupled to the redox change of the cofactors and may be involved in PLS function.

3.2. Methods

Relevant residues will be denoted without the subunit to which they belong to (shown here in brackets): Glu286(A); Lys362(A); Trp95(A); Tyr414(A); Asp407(A); Arg408(A); Arg481(A); Arg482(A); His411;(A); His334(A); His333(A); His284(A); Tyr288(A); Asp412(A); Trp172 (A); Glu254(B).

Preparation of CcO structures and MD simulations

Wild type CcO from *R. sphaeroides* is a four subunit protein (PDB code: 1M56 [35]), but the two subunit crystal structure (PDB code: 2GSM [28]) can be used for investigation, as explained in chapter 1.5. Therefore, the crystal structure with the two central catalytically active subunits I (chain A) and II (chain B) of CcO from *R. sphaeroides* (PDB code: 2GSM [28]), taken from the Protein Data Bank [24] was used for modeling and MD simulations. MD simulations considering only the two catalytically active subunits yield lower backbone RMSD values as shown in previous works on this system [8,9].

Structure modeling and energy minimizations were done with CHARMM [21], using the CHARMM22 force field [21] combined with the CHARMM36 extension for lipids [79] and in-house developed parameters for cofactors [9] (heme a, heme a₃, the Cu_B-complex and Cu_a), as used in previous application on this system [8–10]. MD simulations were performed with software NAMD [80] with 2 fs time step using SHAKE [81] fixing hydrogen atom bond lengths and using Langevin dynamics at 300 K with a small friction constant of $\beta = 1 \text{ ps}^{-1}$ avoiding slowing down of dynamics [82]. In all models, a membrane bilayer is present in the x-y plane. To stabilize the system and keep the membrane in place, a flexible cell size was used for periodic boundary conditions, with a constant ratio of 1:1 for the x- and y-dimensions. Before performing any MD simulations without constraints, an energy minimization of the system was performed. Heating was done by rescaling velocities from 0 K to 300 K over 50 ps time, by increasing the temperature by 6 K every 1 ps. For all MD simulations, the formal equilibration time was considered to be 1 ns after heating.

In the first step modeling of hydrogen atoms with CHARMM [21] was done, followed by the embedding of the two-subunit CcO structure in a phosphatidylcholine lipid bilayer by using a VMD plugin [83] and solvating in a TIP3P [84] water box with periodic boundary conditions. Crystal waters were kept in the system. The system was relaxed by energy minimizing only water molecules, followed by the energy minimization of lipids. As described above, after

heating the molecular system the next step was the equilibration of water and membrane, which was done with an 80ns MD simulation while keeping the protein atoms fixed, as used in previous works [8,9]. The prepared protein-membrane complex was used as a starting point for MD simulations without constraints and as a template for additional modeling of the PLS.

Following the preparation of CcO structure, MD simulations were set up (in steps described above) and performed for four sets of simulations. In the first set, the CcO structure was prepared as described with no additional modeling of CcO structure, referred to as crystal structure MD simulations. Heme propionates, PRAa, PRDa, PRAa₃ and PRDa₃ were deprotonated. These simulations were 25ns long. The other three sets are MD simulations where additional modeling of starting structures was done, i.e. heme a or heme a₃ propionates and their environment were modeled to investigate their capabilities to act as the PLS. More details on MD simulations and additional PLS modeling are given later in the text. In MD simulations, the root-mean-square-deviations (RMSD) of CcO backbone atoms relative to the crystal structure (PDB code: 2GSM[28]), are below 1.45Å (Appendix B, Figure B.1), which is in an agreement with previous works on the same system [8,9].

Redox states of CcO and atomic partial charges of CcO cofactors

In each set of MD simulations, four redox states of CcO were simulated (Figure 3.2). The chemical compositions of CcO cofactors Cu_A, heme a and the BNC (heme a₃ and Cu_B-complex) are given in Table 3.1. The extended notation of states is as follows: for cofactors Cu_A and heme a, O is used for oxidized state or R for the reduced state; for BNC the states are labelled as in Figure 3.2. During the catalytic cycle of CcO, Cu_A and heme a will be in either oxidized or reduced state independent of the BNC. Further in the text, only the unique state label referring to the BNC will be used (bolded letters in the first column of Table 3.1), since only one combination of redox states of Cu_A and heme a is investigated for each redox state of the BNC. For CcO cofactors (Cu_A, heme a, heme a₃ and Cu_B-complex), atomic partial charges were taken from the previously published work [8,9], as mentioned in chapter 1.3.

Table 3.1. Redox states of the cofactors (Cu_A, heme a, heme a₃ and Cu_B-complex) in CcO as obtained from [8]. In this work, only the bolded part of the state notation from the first column will be used for simplicity.

CcO state ^a	Cu _A	heme a	BNC ^b	
			heme a ₃	Cu _B -complex
(OR) P _M ^c	Cu _A ^{II}	Fe ^{II}	Fe ^{IV} =O ²⁻	Cu _B ^{II} -OH ⁻ /Tyr280-O* ^c
(OO) P → F	Cu _A ^{II}	Fe ^{III}	Fe ^{III} =O ²⁻	Cu _B ^{II} -OH ⁻ /Tyr280-O* ^c
(OO) P _R	Cu _A ^{II}	Fe ^{III}	Fe ^{IV} =O ²⁻	Cu _B ^{II} -OH ⁻ /Tyr280-O ⁻
(OO) F ^d	Cu _A ^{II}	Fe ^{III}	Fe ^{IV} =O ²⁻	Cu _B ^{II} -H ₂ O/Tyr280-O ⁻

^a The first two letters in the brackets represent the redox state of Cu_A and heme a, respectively. O is used for the oxidized state of the cofactor and R for the reduced state. The bolded letters show the state of the BNC.

^b BNC is an abbreviation for the binuclear center, a collective name for heme a₃ and Cu_B-complex.

^c In the P_M and P→F state Tyr280 is a tyrosyl radical, marked with an asterisk.

pK_A computations

The initial protonation pattern was determined based on the crystal structure (PDB code: 2GSM[28]), by computing pK_A values with Karlsberg2⁺ (KB2⁺) [7] as in previous applications [8–10,85,86]. The protocol for pK_A computations is introduced in chapter 1.4, namely the electrostatic energies were computed by solving the linearized Poisson-Boltzmann (LPB) equation with the program "Adaptive Poisson-Boltzmann Solver" (APBS) [15]; Protonation energies were computed with TAPBS [16], an in-house modification of APBS. To obtain reliable results with APBS, a three-fold focusing procedure was used, with the resolution of 0.3 Å for the finest grid [7,15]. Additionally, the implicit ion concentration was set to 100 mM. This approach was also used for the titration of time frames of MD trajectories.

The resulting initial protonation pattern is the same for the four investigated states of CcO (Table 3.1) and it was applied to all MD simulations, unless otherwise stated. The same protonation pattern was obtained in previous work [8], including the precise determination of histidine tautomeric states. Most residues are in their standard protonation (charged state) with the exception of neutral Lys362 in the K-channel, neutral Glu286 in the D-channel and protonated Asp407, which shares a proton with PRAa₃ (Figure 3.5). Asp407 is involved in the PLS modeling and its protonation state was varied in some MD simulation, as described later in the text.

The pK_A values were computed for the modeled structures, to evaluate the outcome of the modeling procedure. The results of these computations are shown in the section related to the modeling of PRAa, PRDa₃ and PRAa₃ functioning as potential PLS, later in the text.

The pK_A computations and electrostatic energy evaluations were considered not only for the crystal structure and modeled structures but also for the time frames of generated MD trajectories. To save CPU time needed for pK_A computations, a group of 18 titratable residues within 10 Å of heme a and heme a₃ propionates were titrated (PRDa₃, PRAa₃, PRDa, PRAa, Glu286(A), Tyr414(A), Asp407(A), Arg408(A), Arg481(A), Arg482(A), His411(A), Asp412(A), Arg52(A), Asp299(B), Lys227(B), Tyr336(B), Tyr415(B), Tyr175(B)), contrary to the determination of the initial protonation pattern where all residues were titrated. The remaining residues were kept in protonation states in the initial protonation pattern obtained from the crystal structure. pK_A values were computed with KB2⁺, by utilizing an adaptation of the KB2⁺MD procedure [18], described in the next paragraph. In this procedure, the electrostatic energies are computed by solving the LPB equation with the software APBS [15] and averaged over time frames of the trajectory, taken every 50 ps as proposed in previous work [8–10,52]. The averaged energy terms can be used to generate Boltzmann factors for each trajectory and combined to compute equilibrium pK_A values. The Boltzmann factors obtained for each trajectory can be used to weight trajectories [7] for computing precise pK_A values and for revealing which trajectories are energetically more relevant. In these computations, all water molecules were eliminated, except for the three water molecules coordinating Mg²⁺ (crystal waters W5042, W5048 and W5055) (Figure 3.4 and Figure 3.5). The volume formerly occupied by eliminated water molecules was treated as a dielectric continuum with $\epsilon = 80$ and the volume of the protein-membrane complex was treated as a dielectric continuum with $\epsilon = 4$, as previously described in chapter 1.2. In the central part of CcO, where titrated residues are located, there is a considerable number of internal water molecules. To properly treat the presence of water in internal cavities, an advanced cavity search was performed, using a feature of KB2⁺ [87]. The cavity parameter was $c=0.9$ [87] and the resulting cavity volumes were appended to the solvent volume and treated as a dielectric continuum with $\epsilon = 80$. The cavity parameter of 0.9 was used instead of 0.7, as suggested in [87] for crystal structures, since the protein structure relaxes throughout MD simulation allowing better defined cavity volumes. The atomic partial charges of the membrane were not considered explicitly, since the membrane is at least 12 Å away from heme propionates.

Modification of the KB2⁺MD procedure

In this work, a subtle deviation from the standard KB2⁺MD protocol was applied. In general, the total electrostatic energy was averaged over time frames for each trajectory (chapter 1.4,

section “*Karlsberg2⁺MD procedure*”). Here, it includes all contributions except for the interactions among background charges (described in chapter 1.2), which belong to non-titratable residues of protein-membrane complex. This part of the total energy is practically unaffected by the protonation of the titrated residues. Moreover, preliminary work during the development of the KB2⁺MD procedure [7] showed that this energy term has only a marginal influence on the final weighting of trajectories. To compute these background interactions with high precision, large amounts of CPU power and RAM are needed due to the large size of the molecular system. Therefore, these energy terms were not computed.

Conformational energy computations

To evaluate the stability of conformations of modeled structures, the conformational energy of models and time frames of corresponding trajectories were computed. To compute conformational energies, software APBS [15] was used. Treatment of water molecules was identical as for the pK_A computations (described above); cavity algorithm was utilized with the $c=0.9$; volume formerly filled with water molecules was treated as a dielectric continuum with $\epsilon = 80$; protein volume was treated as a dielectric continuum with $\epsilon = 4$. Computing conformational energy requires large amounts of RAM and CPU time. To save computer resources, membrane was removed, grid resolution for this computation was 0.5 Å and the energies were computed only for frames taken every 500 ps.

Modeling the PLS and MD simulations of modeled structures

For each of PRAa, PRDa₃, PRAa₃ a local conformational change was modeled in order to explore their ability to act as PLS. For PRDa no suitable alternative conformation was found after careful inspection of the crystal structure (discussed later in the text). In all modeling protocols described below, CcO was in the P_R state.

Modeling and MD simulations of PRAa

Propionate A on heme a, PRAa forms hydrogen bonds with the environment in the crystal structure. PRAa has a low pK_A value because of the hydrogen bonds with Tyr414, Trp95 and the backbone N atom of the Arg482 (Figure 3.3), and it is unable to store a proton to play the role of PLS.

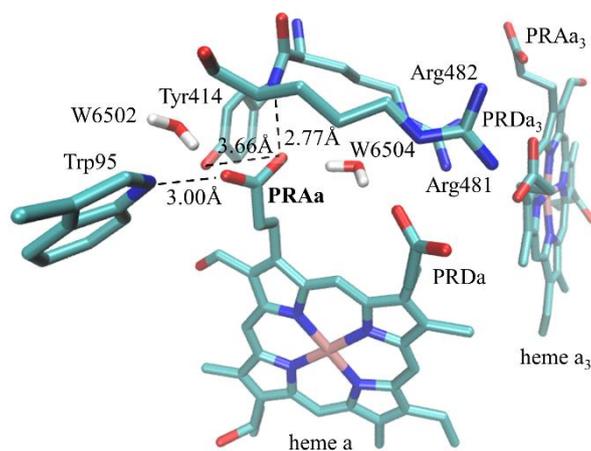


Figure 3.3. The overview of the environment around the PRA on heme a from the crystal structure (PDB code: 2GSM[28]). PRAa forms hydrogen bonds with Trp95, Tyr414 and the N atom of the backbone of Arg482 instead of salt-bridges. Nearby water molecules, except for the two crystal waters (W6502 and W6504), were removed for image clarity.

In an attempt to relax the immediate environment around PRAa, modeling with software CHARMM[21] was done. PRAa may be in one of the three protonation states with the following notation: deprotonated state PRA^- and two protonated states PRAH^1 and PRAH^2 . Formerly prepared crystal structure of CcO, with the membrane in the water box (described above), was used for modeling. One structure was prepared for each of the three protonation states of PRAa. In the next step, PRAa and the environment encompassing all atoms within 10\AA were energy minimized. The resulting structures were used for MD simulations, using the preparation steps described above. Therefore, three MD simulations with CcO in the P_R state of 15 ns length were performed for each of the protonation states of PRAa.

Modeling and MD simulations of PRDa₃

Propionate D on the heme a₃, PRDa₃ forms a strong salt-bridge with Arg481 in the crystal structure. In this salt-bridge, two arginine nitrogen atoms and both propionate oxygen atoms are involved, as seen in Figure 3.4 (left), downshifting the pK_A value of PRDa₃ dramatically. Hence, in the closed salt-bridge conformation, PRDa₃ is unable to act as the PLS. Adjacent to PRDa₃, PRDa is located and forms a salt-bridge with Arg482 (Figure 3.4, left). To model an open salt-bridge conformation, alternative salt-bridge partners for Arg481 and Arg482 were identified (Figure 3.4, right).

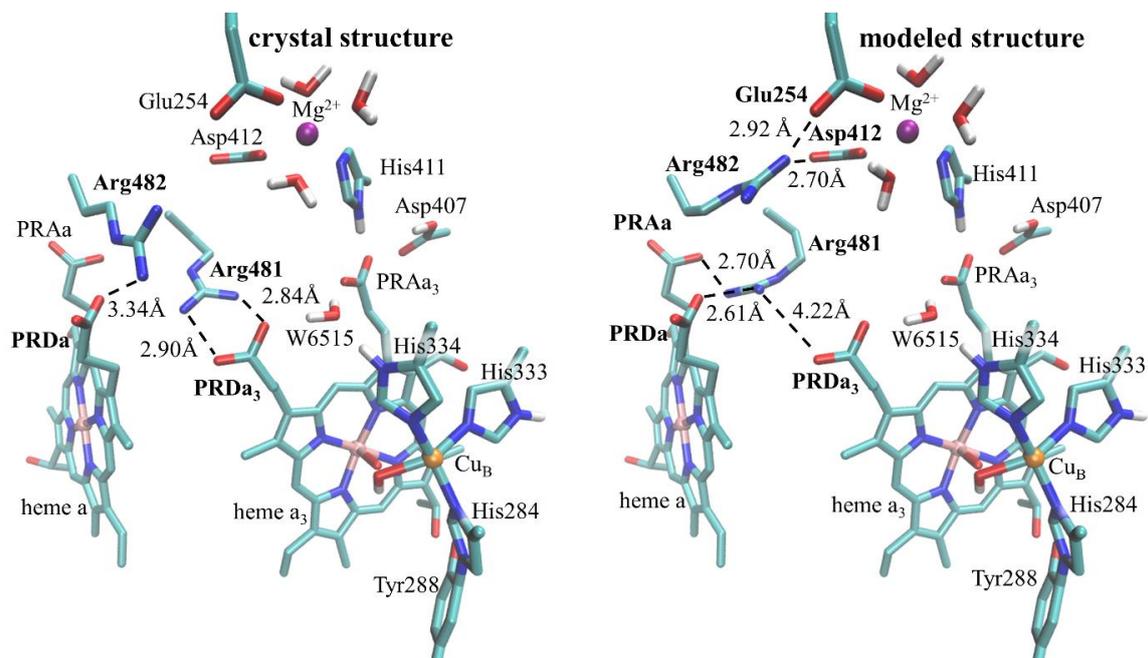


Figure 3.4. Left: The overview of the environment of PRDa₃ in the crystal structure of CcO (PDB code: 2GSM[28]) with salt-bridges PRDa₃-Arg481 and PRDa-Arg482. **Right:** The overview of the environment of PRDa₃ in the modeled structure after the formation of new salt-bridges PRAa-Arg481-PRDa and Glu254-Arg482-Asp412. Water molecules in the cavity above heme a and heme a₃ up to the Mg²⁺-complex are omitted for better visibility. PRDa₃, PRAa₃ and His334 are bridged by a crystal water molecule (W6515).

The rearrangement of salt-bridges of Arg481 and Arg482 was a multi-step task. The prepared crystal structure of CcO, embedded in the membrane and the water box (described above) was used as a starting structure for modeling. To model new salt-bridges, the RESD command of CHARMM [21] was used; it restrains the distances between chosen atom pairs. Appropriate atoms were paired; atoms of Arg481 with atoms of PRDa and PRAa and atoms of Arg482 with atoms of Glu254 and Asp412, which are coordinated to Mg²⁺. Following the pairing, side-chains of Arg481 and Arg482 together with water molecules within 10 Å were energy minimized, followed by another energy minimization of the same atoms without any distance restraints. The obtained structure was locally relaxed by performing a 100 ps MD simulation with CHARMM [21] while keeping all atoms in the system fixed except those that belong to the side-chains of Arg481 and Arg482 and previously selected water molecules. The procedure was done for each state of PRDa₃, which are denoted as: deprotonated PRD⁻ and two protonated PRDH¹ and PRDH² states. The resulting modeled structure for the deprotonated PRDa₃ is shown in Figure 3.4 (right).

After modeling the open salt-bridge conformation between Arg481 and PRDa₃, there was some unoccupied space in the cavity between heme a, heme a₃ and Mg²⁺-complex. Before starting MD simulations, six additional water molecules were added to the system to fill out the cavity. Two water molecules were added in the cavity between heme a, heme a₃ and Mg²⁺-complex, while four water molecules were modeled in the hydrophobic cavity between Glu286 and PRDa₃ (Appendix B, Figure B.2). MD simulations showed that water molecules leave the hydrophobic cavity between Glu286 and PRDa₃ after 1-2ns, and together with the rest of the modeled water reside in the cavity between heme a, heme a₃ and Mg²⁺-complex (Appendix B, Figure B.2).

The obtained modeled structures with different PRDa₃ protonation states were used for MD simulations. The structures were prepared with CcO cofactors in each of the four states shown in Table 3.1. Each of these 12 MD simulations of PRDa₃ were 25ns long.

Glu286 is an important residue located in the D-channel. It is involved in the gating mechanism of the proton transfer [9,48,88], as described in chapter 1.5. The protonated Glu286 (Glu286H) is in the "up" conformation (pointing to the P-side). If this proton is transferred to the PLS or BNC, Glu286 changes to the down conformation (pointing to the D-channel entry at the N-side) where it may quickly receive another proton [9]. To test the influence of the state of Glu286 on the behavior of the potential PLS, it was modeled in both up and down conformations in the modeled structure of CcO from Figure 3.4 (right). Following the modeling of Glu286, a transient "action" pK_A value of PRDa₃ was computed for various combinations of conformation and protonation states of Glu286 (see Appendix B, Table B.1). The pK_A values of PRDa₃ are slightly higher (see Appendix B, Table B.1) when Glu286H is in the down conformation compared to the up conformation, stabilizing the proton in the PLS and when Glu286⁻ is transiently in the up conformation. More details on the relationship between Glu286 and PRDa₃ as the PLS can be found in other works [9,88]. In the MD simulations described above, Glu286 was protonated (neutral) in the down conformation, since all investigated CcO redox states (Figure 3.2) represent the scenarios after the proton transfer from Glu286 and its reprotonation.

Modeling and MD simulations of PRAa₃

Propionate A on heme a₃, PRAa₃ forms a hydrogen bond with Asp407, or possibly shares a proton, in the crystal structure (Figure 3.5, left). In this conformation, PRAa₃ cannot act as PLS, mainly due to the hydrogen bond with Asp407, which is reflected in a very low pK_A

value. For another proton to arrive and stay on PRAa₃, Asp407 is modeled to form a salt-bridge with Arg408 (which had no salt-bridge partners in the crystal structure), as described below.

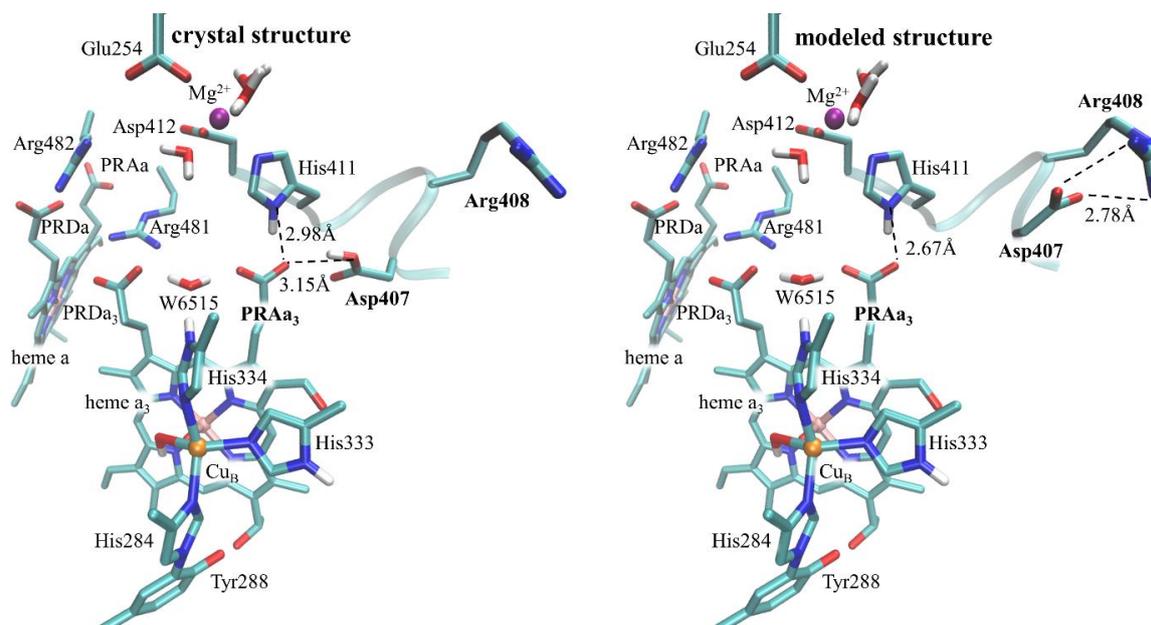


Figure 3.5. Left: The overview of the environment of PRAa₃ in the crystal structure of CcO (PDB code: 2GSM[28]) with the hydrogen bond between PRAa₃ and Asp407. **Right:** The overview of the environment of PRAa₃ in the modeled structure after the formation of the salt-bridge Asp407-Arg408, where hydrogen bond with PRAa₃ is broken. Water molecules in the cavity above heme a and heme a₃ up to the Mg²⁺-complex and those selected in the modeling procedure are omitted for better visibility. PRDa₃, PRAa₃ and His334 bridged by a crystal water molecule (W6515).

The modeling of a salt-bridge between Asp407 and Arg408 was done in several steps. The functionality RESD of software CHARMM[21] was used, similarly to the modeling of PRDa₃. The prepared crystal structure of CcO, embedded in the membrane in the water box (described above) was used for modeling. First, Asp407 was deprotonated. After constraining the distances of side-chains atoms of Asp407 and Arg408, the side-chains of the two residues were energy minimized with water molecules within 10 Å of Asp407. The resulting structure was additionally relaxed by another energy minimization of entire residues Asp407 and Arg408, atoms of the backbone belonging to residues at positions 406 and 409 and the same water molecules selected in the previous step. The procedure was done for each protonation state of PRAa₃, which are noted as: deprotonated PRA⁻ and two protonated PRAH¹ and PRAH² states. The resulting modeled structure for the deprotonated PRAa₃ is shown in Figure 3.5 (right).

The obtained modeled structures with different PRAa₃ protonation states were used for MD simulations. The structures were prepared with CcO cofactors in each of the four states described in Table 3.1. Each of these 12 MD simulations of PRAa₃ were 25ns long.

3.3. Results and Discussion

PLS candidates in the crystal structure

pK_A values of heme propionates

The pK_A values of heme propionates obtained from the computation of the initial protonation pattern of the crystal structure are very low: pK_A(PRAa) = -7.45, pK_A(PRDa) < -10.00, pK_A(PRDa₃) < -10.00, pK_A(PRAa₃) < -10.00, making these groups unlikely candidates for the PLS. The pK_A values of less -10.00 are not determined, since the titration range of Karlsberg2⁺ is from -10.00 to 20.00. All four propionates are forming salt-bridges or hydrogen bonds with the environment that are downshifting their pK_A values, as seen in Figures 3.3, 3.4 and 3.5. To obtain more realistic estimates of pK_A values, MD simulations of 25ns length were performed, with CcO in four oxidation states as shown in Table 3.1. The results of these computations are shown in Table 3.2.

Table 3.2. Computed pK_A values of heme propionic acids, PRAa, PRDa, PRDa₃, PRAa₃ obtained from MD simulations starting from the crystal structure conformation. The software Karlsberg2⁺[7] was used for these computations (KB2⁺MD procedure).

	CcO state			
	P _M	P→F	P _R	F
PRAa	-2.62	-5.56	-5.88	-5.69
PRDa	< -10.00	< -10.00	< -10.00	< -10.00
PRDa₃	< -10.00	< -10.00	< -10.00	< -10.00
PRAa₃	-5.48	-0.83	-1.25	-2.43

MD simulations starting from the crystal structure conformation showed that formerly mentioned interactions preventing heme propionates to act as PLS stay present during the entire time of trajectories. PRAa is still involved in hydrogen bonds with its environment (Figure 3.3), salt-bridges PRDa-Arg482 and PRDa₃-Arg481 are stable (Figure 3.4) and the hydrogen bond between PRAa₃ and Asp407 (Figure 3.5). The pK_A values of the heme propionates obtained from MD trajectories are all below zero, for all states of the CcO. While some of the

pK_A values are higher than in the initial computation on the crystal structure, the values are still too low for any of PRAa, PRDa, PRDa₃, PRAa₃ to accept a proton and act as the PLS.

To investigate PRAa, PRDa, PRDa₃, PRAa₃ in more detail, an attempt was made to find alternative conformations of propionates and their local environment, in which they could act as PLS. Modeling of structures was done, followed by MD simulations. The stability of generated conformations and pK_A values of the heme propionates were described in the following text.

PRAa as PLS candidate

Propionate A on heme a, PRAa is involved in several hydrogen bonds with Tyr414, Trp95 and the backbone N atom of the Arg482 (Figure 3.3) in the crystal structure. It is also connected to PRDa via one crystal water molecule (W6502). The modeling steps described in Methods were applied in order to relax the environment and enable PRAa to be protonated and act as the PLS. The pK_A value computed from modeled structures is -0.69. The MD simulations of modeled CcO structures in the P_R state yielded a PRAa pK_A value of -0.58. The pK_A values of PRAa over time and Boltzmann factors of generated trajectories are given in the Appendix B (Figure B.3). Even though PLS should be closer to a proton release in P_R state, a pK_A value obtained for PRAa is still too low to act as PLS.

PRDa₃ as PLS candidate

Propionic acid D on heme a₃, PRDa₃ forms a salt-bridge with Arg481 in the crystal structure, which prevents it from being protonated (Figure 3.4, left). Other interactions of PRDa₃ are hydrogen bonds with a crystal water molecule (W6515) connecting it to His334 (C_{UB} ligand) and PRAa₃ and with Trp172. To enable PRDa₃ storing a proton transiently, modeling of an open salt-bridge conformation was done, as described in Methods. The modeled structure with rearranged salt-bridges (newly formed PRAa-Arg481-PRDa and Glu254-Arg482-Asp412) is shown in Figure 3.4 (right). The pK_A value of PRDa₃ obtained by combining the computations done on generated models is 2.26. This is a significant improvement of the pK_A value (for the PLS) which was below -10.00 before. The next step was to investigate the conformational stability of the generated models, discussed below.

Conformational stability of the PRDa₃ model

The conformational energy (ΔG_{conf}) was computed for the crystal structure and the modeled structure, by using APBS [15], as described in Methods. The comparison of conformational energies is done only for the structures with the same number of atoms, where PRDa₃ is in the deprotonated state (PRD⁻). The crystal structure (including hydrogen atoms) is by 8.55kJ/mol more stable than the modeled structure. The crystal structures are often grown under dry conditions which may result in shorter distances between residues in salt-bridges. To relax the structure and eliminate the possible bias of the crystallization, MD simulations starting from the modeled and crystal structure were performed, where water was added in cavities as described before. Following the MD simulations, conformational energy was computed for time frames of the two trajectories, shown in Figure 3.6. The conformational energy shows significant fluctuations (up to 400kJ/mol) over time, for both trajectories starting from the crystal or the modeled structure. However, the trend of the computed conformational energies is similar in both cases, which indicates that conformations are of similar stability. Therefore, modeled structures with different protonation states of PRDa₃ and redox state of CcO were used for 25ns long MD simulations.

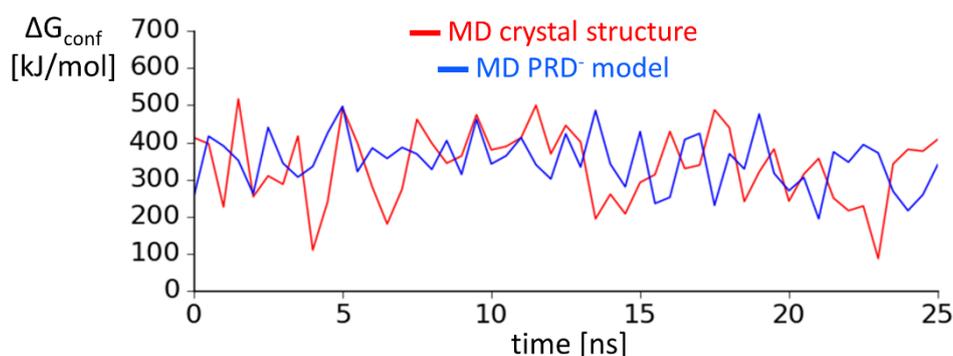


Figure 3.6. Conformational energy (ΔG_{conf}) of structures from time frames of MD simulations with deprotonated PRDa₃ (PRD⁻) starting from the crystal structure conformation (red line) and the modeled structure (blue line). The absolute values of energies are relative to a baseline value of 53200 kJ/mol.

Geometries of the PRDa₃-Arg481 salt-bridge are characterized as: “open (closed)” if both basic nitrogen atoms of Arg481 are more (less) than 3.8Å away from the center of mass of oxygen atoms of PRDa₃ (Figure 3.7). The geometry of the salt-bridge is named “semi-open” if only one nitrogen atom of Arg481 is less than 3.8Å away from the center of mass of oxygen atoms of PRDa₃ (Figure 3.7).

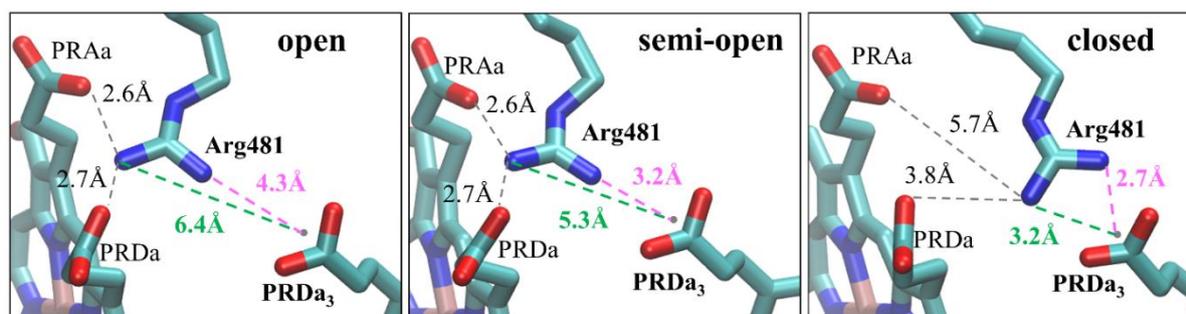


Figure 3.7. Three conformations of the PRDa₃-Arg481 salt-bridge from MD simulations of CcO (PF state), defined by the distance of nitrogen atoms of Arg481 (N₁-magenta, N₂-green) from the center of mass of oxygens atoms of deprotonated PRDa₃ over time. The interactions of Arg481 with PRDa and PRAa are also shown. **Left:** “open” conformation, after 16.5ns of the modeled structure MD simulation. **Middle:** “semi-open” conformation, after 22ns of the modeled structure MD simulation. **Right:** “closed” conformation, after 15ns of the crystal structure MD simulation.

The “closed” conformation of the PRDa₃-Arg481 salt-bridge is found in the crystal structure (PDB code: 2GSM[28]) and it was retained in MD simulations starting from the crystal structure. In MD simulations of modeled structures, the PRDa₃-Arg481 salt-bridge is in an “open” or a “semi-open” conformation, without large pK_A value fluctuations, e.g. in the Appendix B (Figure B.4) for deprotonated PRDa₃. Since the positive charge of a protonated arginine is distributed over five hydrogen atoms, only a completely closed salt-bridge conformation involving two guanidine nitrogens and both carboxylic oxygens prevents the protonation of PRDa₃ completely.

pK_A values of PRDa₃

The pK_A value of PRDa₃ computed for the modeled CcO structures (P_R state, shown in Figure 3.4(right)) is 2.26, a value higher than for the crystal structure, due to the opening of the salt-bridge with Arg481. In the following, the equilibrium pK_A values of titratable residues were computed with KB2⁺ [7] (KB2⁺MD procedure), using MD simulations of the modeled structures. Four redox states of CcO were considered (Figure 3.2). The information on PRDa₃ protonation states in trajectories, the considered time interval and computed Boltzmann factors used for pK_A computation are shown in the Appendix B (Table B.2). The obtained pK_A values of PRDa₃ for different CcO redox states are shown in Figure 3.8. The time course of the pK_A value of PRDa₃ for all considered trajectories (which have nonzero Boltzmann factors) are shown in the Appendix B (Figure B.5).

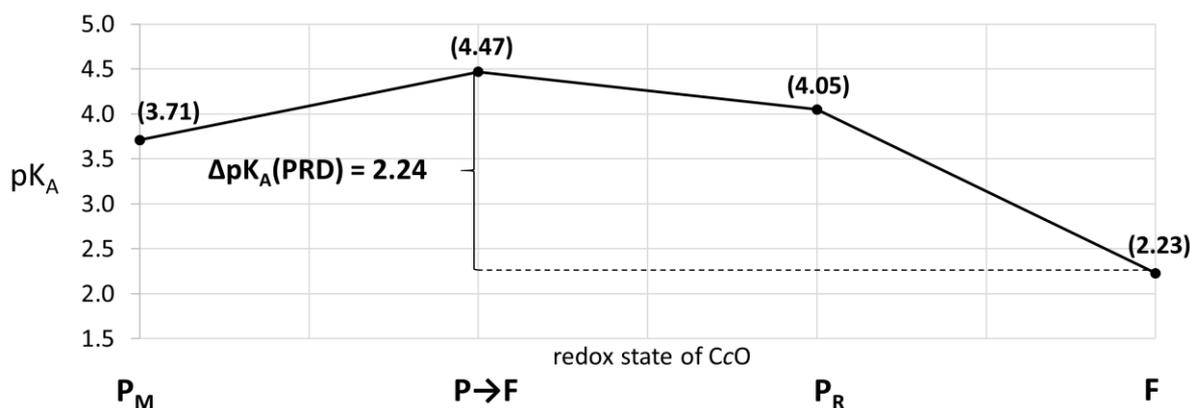


Figure 3.8. Computed pK_A values of PRD_{a3} from MD simulations of CcO in four states described in Figure 3.2. The software Karlsberg2⁺[7] was used for these computations (KB2⁺MD procedure). The difference in pK_A value (ΔpK_A) between the P→F and the F state is 2.24.

The pK_A values obtained from MD simulations are higher than the values from the crystal structure (< -10.00), for all considered CcO states. The trend of how the pK_A of PRD_{a3} changes in different redox states of the enzyme resembles the trend of the pK_A changes expected for a potential PLS. The proton can arrive at the PLS upon reduction of heme a (P_M state); the highest pK_A value of 4.47 is obtained in the P→F state where the proton is stabilized in the PLS and a much lower value of 2.23 occurs in the F state, when the proton is expected to leave the PLS [11,43,76,77].

The absolute pK_A values of PRD_{a3} are relatively low for an acidic residue. However, the energized proton arriving from the N-side of CcO at the PLS would not be released from the P-side of CcO, if its pK_A value is too high. The pK_A value of the PLS must be coupled to the electron-proton transfer from the cycle in Figure 3.2, as described in the introduction of this chapter. The pK_A difference of PRD_{a3} between the P→F state and the F state is 2.24 pH units (Figure 3.8), which is less than the difference of $\Delta pK_A \sim 5$ [41,42] needed for proton pumping in CcO. It is possible that other conformational or protonation changes contribute to the proton release process. Conformations of the PRD_{a3}-Arg481 salt-bridge are “semi-open” or “open” in MD simulations used for pK_A computations (Figure 3.7). Partial closing and opening of the salt-bridge could be coupled with the electron-proton transfer.

In other computational works, PRD_{a3} is suggested as the primary [42,77] or the secondary PLS after the proton is ejected from a His334 ligand of Cu_B [39,73] or considered as a part of the cluster of PLS residues [43,89]. Many experimental studies discuss the possibilities of the PLS

in the propionate region of heme a_3 [70,72,74,90,91]. Mutation experiments of Arg481 showed reduced CcO activity and proton pumping stoichiometry [91]. It may be possible that the mutation of Arg481 leads to conformational changes in the environment that influence the CcO function or that changes of the conformation of the PRDa₃-Arg481 salt-bridge regulate the PLS activity. Surface enhanced resonance Raman (SERR) experiments suggested that both PRDa₃ and PRAa₃ could act as the PLS [72]. It is often suggested that PRDa₃ is only involved in the loading of the PLS by conducting a proton to PRAa₃ [12,45,90,92–94]. PRDa₃ and PRAa₃ are bridged by a crystal water molecule (W6515), which is also connected to the His334 ligand of the Cu_B-complex. PRDa₃ can likely act as the PLS when the salt-bridge with Arg481 is not closed. However, it is not improbable that both propionates of heme a_3 are indeed involved in the PLS function, which is discussed further in the text.

PRAa₃ as PLS candidate

Propionic acid A on heme a_3 , PRAa₃ forms a hydrogen bond with Asp407 in the crystal structure, possibly sharing a proton (Figure 3.5, left). PRAa₃ also forms a hydrogen bond with His411 (Mg²⁺ ligand) and one crystal water molecule (W6515), connecting PRAa₃ with His334 (Cu_B ligand). As an acceptor of so many hydrogen bonds, the initial computation with KB2⁺ [7] assigned PRAa₃ to be deprotonated (while Asp407 is protonated) with the pK_A value of less than -10.00. For PRAa₃ to act as the PLS, the hydrogen bond with the Asp407 must be broken. Modeling of the conformation of Asp407 in the deprotonated state where it forms a salt-bridge with Arg408 was carried out, as described in Methods. The modeling result is shown in Figure 3.5 (right). Modeling was done for each protonation state of PRAa₃. The pK_A value from modeled structures of 0.92 is too low to assume that it can hold the proton. However, further investigation with MD sampling was carried out. First, conformational stability of the generated models was investigated, as discussed below.

Conformational stability of the model PRAa₃

The conformational energy (ΔG_{conf}) was computed for the crystal structure and the models, by using APBS [15], as described in Methods. Conformation energies can be compared only for the structures with the same number of atoms. In the modeled structures used for conformation energy computations PRAa₃ is in one of the protonated states (PRDH¹ or PRDH²) with Asp407 deprotonated while based on the crystal structure the computed protonation states are reversed:

deprotonated PRAa₃ (PRA⁻) and protonated Asp407. The models with PRDH¹ or PRDH² are more stable than the crystal structure conformation by 25.82 kJ/mol and 26.52 kJ/mol, respectively.

Conformational energy was computed for time frames from MD simulations, shown in Figure 3.9. Similarly to the results for PRDa₃, the conformational energy fluctuates over time (up to 400 kJ/mol), for both the crystal and the modeled structures, which indicates that conformations are similarly stable. Following the validation of the conformational stability of the model, MD simulations of 25 ns length with variation of PRAa₃ protonation state and CcO redox state were performed. The conformation where PRAa₃ is not forming a hydrogen bond with Asp407 is stable in all MD simulations except in one, in which the Asp407-Arg408 salt-bridge partially opens and PRAa₃ and Asp407 form a hydrogen bond again (Appendix B, Figure B.6).

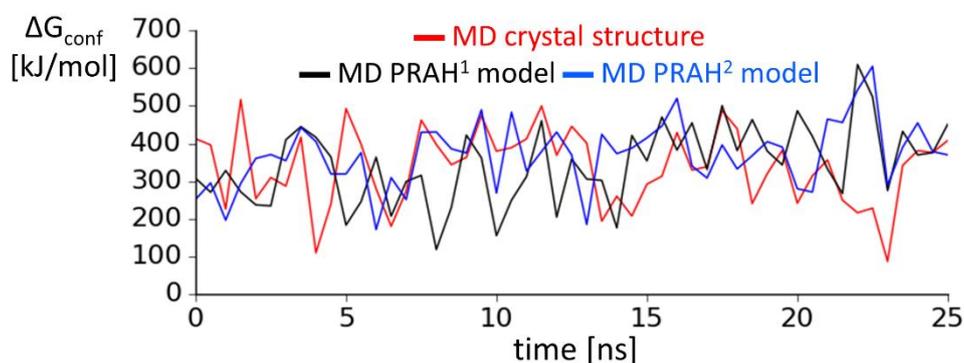


Figure 3.9. Conformational energy (ΔG_{conf}) of structures from time frames of MD simulation varying the states of PRAa₃. MD simulations starting from the crystal structure conformation are with deprotonated PRAa₃ (PRA⁻) and protonated Asp407 (forming a hydrogen bond). MD simulations starting from the modeled structure conformation are with deprotonated Asp407 (in a salt-bridge with Arg408) and protonated PRAa₃; where one of the carboxylic oxygen atoms is protonated PRAH¹ or PRAH² (black and blue line). The absolute values of energies are relative to a baseline value of 53200 kJ/mol.

pK_A values of PRAa₃

PRAa₃ has a pK_A value of 0.92, based on computations of the modeled CcO structures (P_R state, shown in Figure 3.5(right)). The low pK_A value implies that a proton cannot even transiently be located on PRAa₃, making it an unlikely PLS candidate. However, MD simulations were used for more precise pK_A computations. The equilibrium pK_A values of titratable residues were computed with KB2⁺ [7] (KB2⁺MD procedure), using MD simulations of modeled structures for four redox states of CcO (Figure 3.2). The protonation states of

PRAa₃ in trajectories, the considered time intervals and computed Boltzmann factors used for pK_A computation are presented in the Appendix B (Table B.3). The computed pK_A values of PRAa₃ for considered CcO states are shown in Figure 3.10. The course of the pK_A value of PRAa₃ over time for all considered trajectories (which have nonzero Boltzmann factors) are shown in Figure B.7 (see Appendix B).

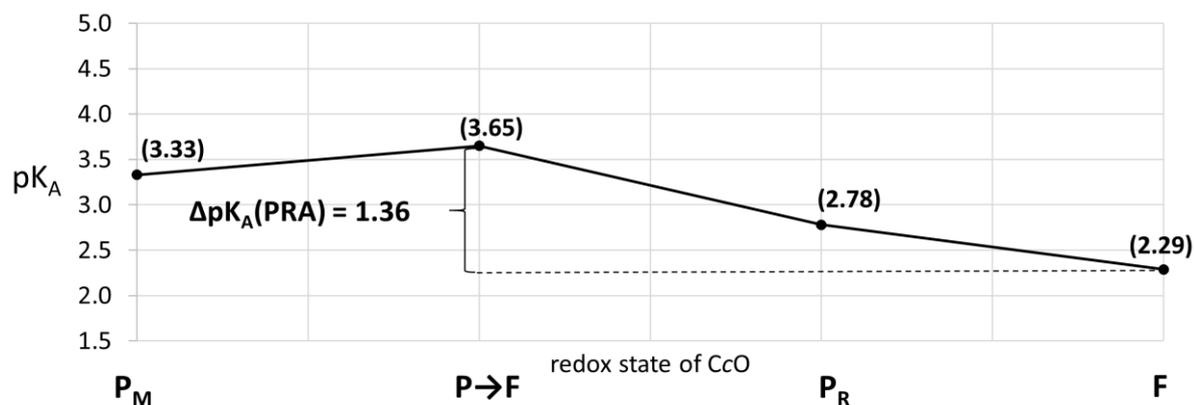


Figure 3.10. Computed pK_A values of PRAa₃ from MD simulations of CcO in four redox states from Figure 3.2. The software Karlsberg2⁺[7] was used for these computations (KB2⁺MD procedure). The difference in pK_A values (ΔpK_A) between the P→F and the F state is shown.

PRAa₃ as the PLS, loaded with an energized proton, is stabilized the most when the electron is on heme a₃ in the P→F state with a pK_A value of 3.65, while in the F state (Table 3.1), PRAa₃ has a low pK_A value of 2.29 and is ready to release its proton. The pK_A difference of 1.36 pH units between the P→F and the F states (Figure 3.2) is smaller than the requirements of the proton pump in CcO ($\Delta pK_A \sim 5$) [41,42]. The coupling between the pK_A value of PRAa₃ in the modeled structure and the redox state of BNC is shown in Figure 3.2. However, it is possible that other conformational changes take place to fully support the function of PRAa₃ as PLS.

In other works, PRAa₃ is often considered to be the PLS, with the proton being translocated from Glu286, via water chain and PRDa₃ to PRAa₃ [12,45,90,92–94] or by proton ejection from His334 (Cu_B ligand [39,73]). The computed pK_A values of PRAa₃ in other studies [45], were higher than in this work, however computations in theoretical studies may depend on the method applied. The pK_A difference of PRAa₃ in [45], upon changing the oxidation state of CcO cofactors is about 2 pH units, similar to the results obtained in this work. A different mechanism of PLS loading is proposed in [45], where after dissociation of the hydrogen bond of PRAa₃ and protonated Asp407 another proton arrives at PRAa₃. The pK_A values of the PLS were indirectly derived from time-resolved optical and electrometric experiments [74],

eliminating PRDa₃ as the PLS in favor of PRAa₃. In that study [74], the estimated pK_A values of the PLS are higher than the ones obtained in this work for the same redox states of CcO. This may again be the result of the approximations done in order to indirectly compute pK_A values. The relatively high pK_A values obtained in [74] may result in too much stabilization of the proton at the PLS. In this work, PRAa₃ could be proposed as the secondary PLS candidate, eventually receiving a proton via PRDa₃ (or His334), which is in some agreement with the other works [12,39,45,73,90,92–94]. The specific mechanism of the functioning of PRAa₃ as the PLS may need to be clarified further on. Subsequently, the PLS candidates are briefly discussed together and a few possible proton pathways involved in loading and unloading of the PLS are suggested.

PRDa as PLS candidate

Propionic acid D on heme a, PRDa may be a PLS candidate, being located on the upper side of the nonpolar cavity that connects PRDa₃ and Glu286 and it was suggested as a part of the PLS cluster of residues [43,89]. In the crystal structure PRDa is involved in a salt-bridge with Arg482 and a very low pK_A value of below -10.00, which was confirmed by MD simulations for various CcO redox states (Table 3.2). In this work, alternative conformations of the local environment were modeled for propionates of heme a and heme B. However, in all models and the already mentioned crystal structure PRDa is involved in a salt-bridge with Arg482 or Arg481, preventing it to be a proton acceptor. Additionally, in the model generated in Figure 3.4 (right), Arg482 moves back down to PRDa after a few nanoseconds in most MD simulations. This implies that PRDa likely cannot serve as an effective proton translocation residue to PRAa, eliminating both propionates of heme a as PLS candidates.

How could the PLS function?

From the MD simulations of modeled structures and subsequent pK_A computations presented in this work, PRDa₃ and PRAa₃ emerged as the PLS candidates. Here, several proton pathways starting from Glu286 to the PLS and the proton exit are proposed (Figure 3.11), combining results from this work and other sources.

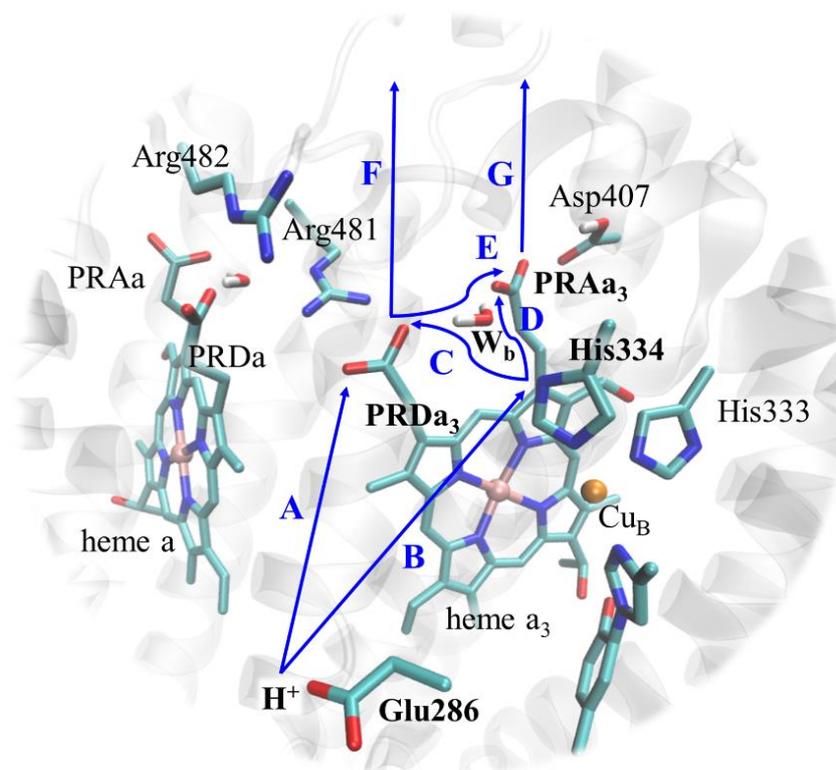


Figure 3.11. Schematic representation of hypothesized proton (H^+) pathways from the P-side to the PLS candidates and H^+ release to the N-side of CcO (PDB code: 2GSM[28]). The initial proton transfer is from Glu286 via water molecules in the cavity at the PLS. A) PRDa₃ acts as the PLS; B) His334 receives the H^+ ; C) The H^+ is transferred from His334 via the bridging water molecule labeled W_b (corresponding to crystal water W6515) to PRDa₃ or D) to PRAa₃ (secondary PLS). E) The H^+ is transferred from PRDa₃ to PRAa₃ which acts as the secondary PLS; The H^+ is released F) from PRDa₃ or G) from PRAa₃.

In A-type CcO, Glu286 at the end of the D-channel is supplying all pumped protons. It was shown that Glu286 is transferring protons via a chain of water molecules that forms in the hydrophobic cavity in which Glu286 itself is located [9]. The location of the proton pathways (A-G) mentioned in the following are indicated with blue arrows in Figure 3.11. Following the reduction of heme a, the initial proton (H^+) transfer could be from Glu286 to PRDa₃ (A) [9,19] or a proton transfer directly to His334 (B) (Cu_B ligand) concerted with the deprotonation of the His334. Subsequent, the movement of the electron starts with the P_M state, as described in the cycle in Figure 3.2. The proton released from His334 can be transferred via bridging water molecule W_b (corresponding to crystal water W6515) to PRDa₃ (C) or PRAa₃ (D) [39,73]. If the proton is located on PRDa₃, it can be transferred to the secondary PLS, PRAa₃ (E) [12,45,90,92–94]. Finally, in the F state after the chemical proton had arrived at the BNC and

Glu286 carries another pumped proton (Figure 3.2), the PLS (PRDa₃ or PRAa₃) is deprotonated again and the proton leaves from PRDa₃ via (F) or from PRAa₃ via (G) [39]. It can be speculated that different residues play the role of the PLS in different parts of the CcO redox cycle (Figure 1.5, chapter 1.5) [36,73].

3.4. Summary and Conclusions

The main focus of this work was to identify Proton-Loading Site (PLS) candidates in CcO from *Rhodobacter sphaeroides*. The PLS is a molecular group in CcO, close to the BNC, which regulates the transfer of protons across the membrane against the electrochemical gradient by changing its pK_A value coupled to changes of charges at CcO cofactors [36]. Molecular modeling, MD simulations and pK_A computations were carried out in order to investigate the capability of propionates-A and -D of heme a (PRAa, PRDa, respectively) and propionates-A and -D of heme a₃ (PRAa₃, PRDa₃, respectively) to act as the PLS. CcO redox states examined in this work were from P_M (loaded PLS) via P→F and P_R to F (unloaded PLS), shown in Figure 3.2, where the first out of four protons is pumped across the membrane.

From MD simulations and pK_A computations based on the crystal structure, all four propionates are unlikely candidates for the PLS, having pK_A values below zero. Low pK_A values are due to the interaction of propionates with the environment preventing protonation. PRAa is involved in a system of hydrogen bonds with the environment; PRDa is in a salt-bridge with Arg482; PRDa₃ is in a salt-bridge with Arg481 and PRAa₃ forms a strong hydrogen bond with Asp407, sharing a proton with it. Modeling with CHARMM [21] was used to generate alternative conformations of propionates and their interaction partners. The pK_A value of PRAa stayed below zero even after relaxing its immediate environment, therefore it can be likely excluded as the PLS. PRDa is always involved in a salt-bridge with Arg482 or Arg481, be it in the crystal structure or in the models, eliminating it as a PLS candidate. The salt-bridge PRDa₃-Arg481 needs to be opened (Figure 3.4), to at least a “semi-open” conformation for PRDa₃ to be able to accept a proton. Similarly, when the hydrogen bond of PRAa₃ with Asp407 is broken (Figure 3.5), protonation of this propionate becomes probable. When the electron is located on heme a₃ (P→F state), the proton is stabilized in the PLS and when the electron is neutralized by the chemical proton and the arrival of the next pumped proton, the PLS releases the stored proton to the P-side. The pK_A values computed for PRDa₃ and PRAa₃ in different states are still too low for a permanent protonation but may be high enough for transient protonation. The

difference in pK_A values of propionates (ΔpK_A) in P \rightarrow F (when proton is stabilized in the PLS) and F state (when proton is released), are 2.24 for PRDa₃ and 1.36 for PRAa₃. These computed ΔpK_A values are lower than the experimentally estimated change in the pK_A values of the PLS, which is approximately 5 pH units. While both propionates can change their pK_A values significantly coupled with the change in charge on the CcO cofactors, it may be that additional conformational changes around the propionates take place to increase the stability of the proton in the PLS and prevent backflow. The possible conformational changes could be hydration and opening of the nearby cavity coupled to conformations of Glu286 [88]. Nevertheless, the results obtained in this work point to PRDa₃ and PRAa₃ as good PLS candidates, acting together with other nearby residues or separately. Possible proton pathways were postulated, from loading of the PLS (proton leaves Glu286) to the unloading of the PLS, as shown in Figure 3.11. These postulations above mostly agree with other works [12,39,45,73,90,92–94].

Future work could include the analysis of the events of loading of the proton in the PLS and its release, extending the investigation throughout the entire CcO redox cycle in order to determine if always the same groups act as PLS. More detailed computations on His334 (Cu_B ligand) are also useful to explore the proton transfer to the propionates in detail.

4. Cytochrome *c* Oxidase near the K-channel entry

This project was carried out in collaboration with the experimental group AG Alexiev from the Department of Physics at Freie Universität Berlin. Prof. Dr. Ernst Walter Knapp and Prof. Dr. Ulrike Alexiev supervised the work. Alexander Wolf is a doctoral student from AG Alexiev who contributed equally.

Own contributions:

- Computational work
- Generating all MD simulations and performing pK_A computations
- Analysis of results from computations
- Preparation and writing of manuscripts

Other contributions:

- Design and supervision of the project and construction of manuscripts (Ernst Walter Knapp and Ulrike Alexiev)
- Early computational work and system preparation (Anna Lena Wölke)
- Technical assistance with CHARMM (Milan Hodošček)
- Experimental work (Alexander Wolf)
- Analysis of results from experiments (Alexander Wolf)
- Preparation and writing of manuscripts (Alexander Wolf)

The following text is the modified presentation of two publications that are currently in preparation. The entire project from both manuscripts is merged in this chapter. The experimental results are presented in a simplified form to highlight the findings of this project.

4.1. Introduction

Cytochrome *c* oxidase (CcO) is discussed further in this chapter. In this project the protonation changes close to the K-channel entry of CcO from *Paracoccus denitrificans* were investigated.

As a proton-pump, CcO has several components that are needed to ensure the proper redox-coupled proton transport, as introduced in chapter 1.5. Protons can either be translocated to the catalytic center where the chemical reaction takes place, named chemical protons or they can be moved across the membrane, in which case they are pumped, named physical protons. Proton transport occurs via two input channels in A-type CcO, named the D- and K-channels. In one full redox cycle of CcO (Figure 1.5, chapter 1.5), eight protons in total are taken up, four of which are used for the chemical reaction, forming two water molecules and four are pumped across the membrane. Six protons enter via the D-channel and two protons via the K-

channel. The protons transported via the K-channel are chemical protons and this pathway is only active during the reductive part of the CcO cycle [8,38].

Details about the proton uptake, gating of the K-channel, the switch between the channels or channel communication are well investigated. [8,10,38,95–97] Nonetheless, open questions remain. The proton uptake from the surface of CcO on the N-side must be coupled with the activity of K-channel between O and R states. Earlier studies showed that there could be a proton-collecting antenna (PCA) that is activated to facilitate proton transfer close to entry of the K-channel [98]. The entry of the K-channel is not clearly defined, however, the glutamate (Glu78B in *P. denitrificans*) at the beginning of the K-channel has been suggested as the candidate [99]. Additionally, polarity changes on the surface of CcO were shown to be coupled to the redox state [97,100]. K-channel gating is described to be mainly governed by the protonation state and conformational change of the conserved Lys354 in order to bridge the proton translocation to Tyr280, i.e. the catalytic tyrosine [8,38]. Tyr280 and its covalently bound histidine (part of Cu_B-complex) are localized on Helix VI that connects the surface in the vicinity of the K-channel to the BNC. The connection between BNC and the CcO surface is interesting, since studies suggest that conformational changes of the enzyme take place to support the K-channel activity [101,102].

To shed more light on these subjects, the experimental group AG Alexiev used site-directed fluorescence labeling with sensitive fluorescent probes on the mutated CcO from *Paracoccus denitrificans*. We used molecular dynamics (MD) simulations to provide an atomistic picture of the supposed changes in CcO and to support the experimental findings. Labeling was done with the dye fluorescein. To observe the changes on the CcO surface close to the K-channel, a Lys299 was mutated into Cys299 which enables binding the fluorescein. The position 299 is part of the loop at the end of the Helix VI, which is may be important for the activity of the K-channel (Figure 4.1). Computational work, mainly pK_A computations and electrostatic energy computations were done to confirm that wild-type behavior of CcO is not disturbed upon labeling with fluorescein. To the best of our knowledge this is the first work that combines experiment and computations of the complete protein-membrane moiety of CcO including the fluorescein label. The functionally relevant residues His73 and His526 at the CcO surface, were also investigated. His73 may serve as entrance of the K-channel and is believed to be involved in the proton transfer. pK_A computations reveal conformational changes within CcO that occur upon reduction and that affect the H-bond pattern at the surface.

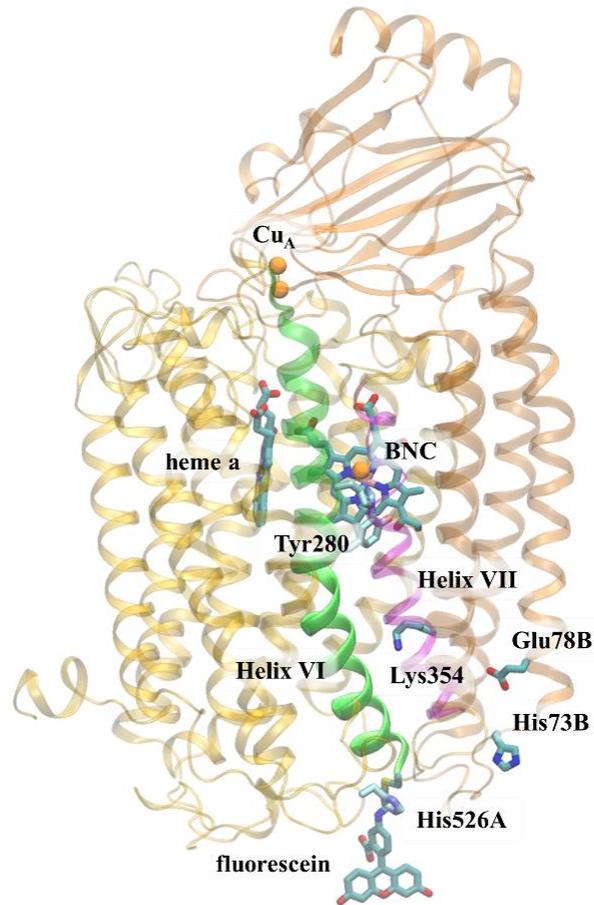


Figure 4.1. Modeled structure of subunits I (chain A) and II (chain B) of CcO from *Paracoccus denitrificans* (crystal structure PDB code: 3HB3 [30,31]), with K299C mutant and covalently bound fluorescein. Relevant residues and Helices VI and VII are highlighted. The membrane was omitted.

4.2. Methods

Crystal structure of CcO

The two subunit crystal structure of CcO from *R. denitrificans* bacteria was taken from the Protein Data Bank [24] (PDB code: 3HB3 [30,31]), as explained in chapter 1.5. Starting from the wild-type (named wt-CcO), a mutant K299C (named K299C-CcO) was modeled, which was used to model a structure with bound fluorescein (attached to Cys299) at the CcO surface (named flu-CcO).

Force field, redox states of CcO and atomic partial charges of CcO cofactors and fluorescein

Modeling structures and performing MD simulations of CcO embedded in the membrane was done with the CHARMM22 force field [21], the CHARMM36 extension for lipids [79] and additional (in-house) developed parameters for cofactors [9], as in chapter 3 and previous work [8,9]. Parameters of the fluorescein molecule were generated based on existing CHARMM force field parameters.

In conducted experiments, CcO was in the fully oxidized (O) or fully reduced (R) state. In order to reproduce these conditions, CcO was modeled in O (O_H state) or R state of the catalytic cycle (Figure 1.5, chapter 1.5). For the fully oxidized state, the binuclear center (BNC), i.e. heme a₃ and Cu_B-complex, has the following chemical composition: heme a₃(Fe^{III}-OH⁻)/Cu_B^{II}-H₂O/Tyr280-O* and heme a and Cu_A are also oxidized (heme a(Fe^{III})/Cu_A^{II}). For the fully reduced state, the BNC components are in the following states: heme a₃(Fe^{II}-H₂O)/Cu_B^I-H₂O/Tyr280-OH and heme a and Cu_A are also reduced, i.e. heme a(Fe^{II})/Cu_A^I. The reduced state produced in experiments and computations is with four electrons reducing all CcO cofactors, which differs from the physiological R state, in which two electrons reduce BNC at the end of the redox cycle but not heme a or Cu_A. Nevertheless, this state may be used to investigate the proton uptake to the K-channel.

The fluorescein atomic coordinates and partial charges were calculated quantum chemically as in previous work [103], using DFT from Jaguar v.7.7 [104] with the B3LYP functional and 6-31g** basis set [104]. Initial coordinates of fluorescein were obtained from the crystal structure of the human carbonic anhydrase II in complex with a fluorescent inhibitor (PDB code: 2F14 [105]). The fluorescein geometry was optimized quantum chemically, with no constraints on

hydrogens atoms while optimizing the other atoms with respect to bond-lengths and bond-angles and keeping the torsion angles fixed in the geometry of the crystal structure. The following step was the computation of the electrostatic potential with the same quantum chemical procedure used for geometry optimization. Finally, using a two-stage restraint-electrostatic-potential (RESP) [106,107] procedure the atomic partial charges were computed fitting the quantum chemically obtained electrostatic potential. The atomic partial charges of fluorescein are listed in the Appendix C (Table C.1).

Modeling of wild-type CcO

Wild-type CcO was prepared for MD simulations with CHARMM [21]. The protein was embedded in the phosphatidylcholine lipid bilayer and subsequently solvated in a TIP3P [84] water box, according to the steps described in chapter 3 (section “*Preparation of CcO structures and MD simulations*”). The obtained structure of the wt-CcO with equilibrated membrane and water was used as a template in subsequent modeling steps.

Computation of pK_A values of residues in crystal structure of wt-CcO

The initial protonation pattern is determined based on the crystal structure. Here, the crystal structure of CcO (PDB code: 3HB3) was used for both O and R states of the enzyme. Karlsberg²⁺ [7] was used to compute pK_A values of all titratable residues with the protocol introduced in chapter 1.4 (computational details presented in chapter 3 , section “*pK_A computations*”). This approach was used for the titration of all time frames of MD trajectories, described below.

The resulting protonation pattern obtained from the crystal structure of CcO, is presented in detail in the Results and Discussion section. This protonation pattern was applied for all MD simulations and kept fixed except for fluorescein and two relevant histidine residues, His526A and His73B, which will be referred to as His526 and His73, respectively. His526 is very close to the fluorescein (Figure 4.2) and His73 is close to the entrance of the K-channel and could be a shuttle in the proton pumping process [98]. Based on the crystal structure, both histidines are in the deprotonated state (oxidized CcO: pK_A(His73)=5.5, pK_A(His526)=4.5; reduced CcO: pK_A(His73)=5.5, pK_A(His526)=4.6). These residues may vary their protonation states, coupled with the redox change of CcO, Therefore, their protonation states were explored in detail.

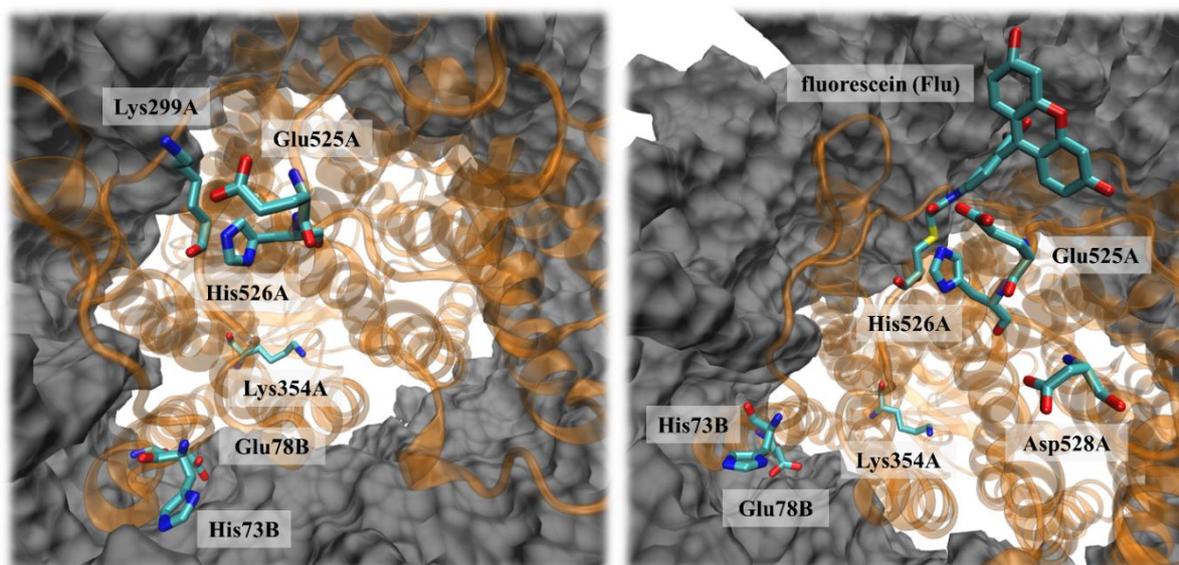


Figure 4.2. Overview of the N-side of the CcO protein-membrane moiety with focus on the K-channel entry, characterized by the conserved Lys354A. Transparent orange color represents the protein backbone; grey volume represents membrane slab. **Left:** wild type CcO (wt-CcO) with Lys299. **Right:** fluorescein environment attached at residue Cys354A, after K229 mutation.

Modeling of flu-CcO, addition of fluorescein to the protein surface

Modeling the bound fluorescein molecule on the surface of wt-CcO is a multi-step task. The previously prepared structure of wt-CcO with the equilibrated membrane was used as a template to add fluorescein to the system. The mutation of Lys299 into cysteine is the first step of the procedure, where the coordinates of the backbone atoms of Lys299 are kept and the sidechain of Cys rebuilt with CHARMM. In the following, the fluorescein conformation was taken from the crystal structure of the human carbonic anhydrase II in complex with a fluorescent inhibitor (PDB code: 2F14 [105]) and added to CcO forming the flu-CcO system. As seen in Figure 4.3, fluorescein has triple ring system orthogonal to the benzoic acid ring and it binds to cysteine by forming a C-S bond with the elimination of hydrogen.

Fluorescein has several protonation states. In this work, the carboxylic group of the benzoic acid was always kept deprotonated, however one the oxygen atoms of the triple-ring system could be protonated. If it is protonated, fluorescein has a total charge of -1 and if deprotonated the total charge is -2. The two protonated states of terminal oxygens are tautomeric, with a pK_A value in aqueous solution of 6.2 [108,109]. The triple-ring system may rotate, which would convert one tautomer into the other, however this was never observed in our MD simulations.

Therefore, there can be three different protonation states of fluorescein for purposes of computational work: a deprotonated state labeled as Flu and two protonated states, Flu(u) or Flu(d) depending the position of the hydrogen atom relative to the triple-ring system, as described in detail in Figure 4.3.

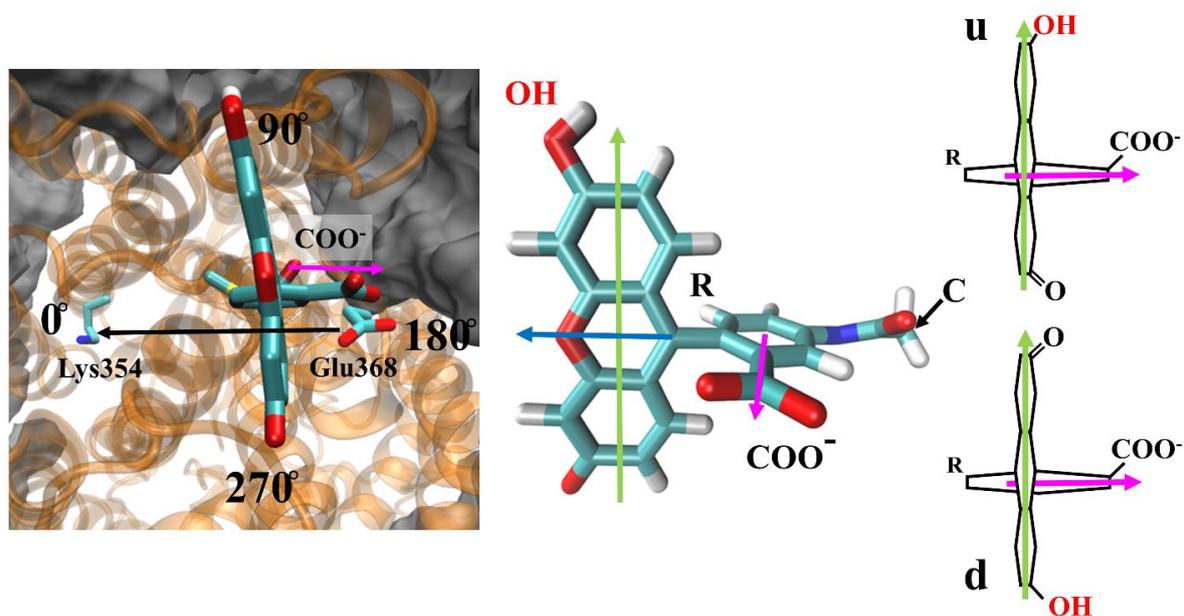


Figure 4.3. Left: Fluorescein orientation (in the initial model with the orientation angle of 180°), represented as the angle between the vector with the direction defined from Glu368A to Lys354A (black vector) and the vector pointing from the centroid of the aromatic ring to the carbon atom of the carboxylic group of the benzoic acid part of fluorescein (magenta vector). **Middle:** The structure of fluorescein and the vectors characterizing the orientation of fluorescein. The vector going through the triple-ring system is the vector characterizing fluorescence depolarization (green vector). The vector connecting the sulfur atom of fluorescein with the center oxygen atom of the triple-ring system (blue vector) is used to describe “cone-in-a-cone” movement of fluorescein as in previous works [110]. Ideally, blue, green and magenta vectors are pair-wise orthogonal. But, small fluctuations are possible. **Right:** Protonation states of fluorescein. Two tautomers of fluorescein can be discriminated as Flu(u) or Flu(d), depending what side is protonated relative to the magenta vector. If the fluorescein is deprotonated, it is denoted as Flu. Two different protonated states of fluorescein can be distinguished if there is no full rotation around the blue vector, which is however possible in experiments.

In the fluorescein structure was modeled with the triple-ring plane orthogonal (blue vector in Figure 4.3) to the CcO surface. The orientation and flexibility of fluorescein are important. However, the mobility of such a voluminous molecule can be greatly impaired by the presence of the protein surface. Previous work showed that for a small system, like fluorescein attached

to the AB-loop of bacteriorhodopsin, the fluorescein can fully reorient in about 1 ns [108]. In our case with CcO, fluorescein at the position C299 is surrounded by loops, which can slow down the reorientation time considerably. In order to find the best initial orientation, extensive preliminary MD simulations were carried out, as documented in the Appendix C. Four different initial orientations of fluorescein were modeled, labeled by the angle between the magenta and black vector (0° , 90° , 180° , 270°), as explained in Figure 4.3. To properly fit fluorescein into the water box, the overlapping water molecules were deleted subsequently after modeling fluorescein in each orientation. In the preliminary work, 5ns MD simulations were performed with CHARMM [21], as described in the Appendix C. The results from preliminary MD simulations showed that fluorescein adopts a preferred orientation with the angle of 180° . In subsequent steps, longer MD simulations of at least 32 ns were performed for various protonation states of relevant residues and in this longer time span fluorescein, however not as fast as in the case with bacteriorhodopsin where there are no steric impairments.

MD simulations

MD simulations were performed using NAMD [80]. General conditions for MD simulations, i.e. heating, temperature regulation, setting for the periodic boundary conditions, etc., were defined in chapter 3 (section “*Preparation of CcO structures and MD simulations*”). The total simulation times of generated trajectories in this work varied and different equilibration times were deducted from the entire run, depending on the quantities that were analyzed. In case of geometric quantities, e.g. root mean square deviation (RMSD) of coordinates of the protein backbone atoms, the equilibration time was shorter than in the case of energetic quantities, e.g. pK_A values.

For wt-CcO an equilibration time of 200ps was sufficient, since water and protein-membrane moiety were already equilibrated, as described above. In all time plots, the initial time $t = 0$ corresponds to the time after equilibration. The simulation time of each wt-CcO trajectory was 20ns. In these simulations, His73 and His526 were of main interest, especially by comparing the pK_A values to those residues in wt-CcO with flu-CcO, to investigate the influence of the fluorescein. Therefore, several MD simulations were performed, considering different protonation states of the two histidines and redox states of CcO (more details are shown in the Results and Discussion section).

In case of flu-CcO, the equilibration time was longer to allow full equilibration of fluorescein orientation. After fluorescein is added to the equilibrated wt-CcO structure, water molecules

and the membrane must be equilibrated again. MD simulations of the system flu-CcO with fluorescein orientation angle of 180° (as the preferred orientation) were performed with different protonation states of His 73, His526 and fluorescein for two redox states of CcO. For analysis of geometric quantities, like RMSD, equilibration time was 200 ps. The simulation time of flu-CcO trajectories was between 32 ns and 47 ns including equilibration, except for the ones done in the preliminary work for finding the preferred orientation of fluorescein (see Appendix C). In case of energetic quantities, like pK_A values longer equilibration times were considered. For both wt-CcO and flu-CcO, the last 15 ns of the trajectories were used to evaluate pK_A values. The analysis of fluorescence depolarization required an additional trajectory for the simulation of oxidized state of flu-CcO in order to match the protonation state of fluorescein in the experiment at pH 8, where fluorescein is deprotonated. The starting point of this MD simulation was not the modeled structure described above but the set of coordinates taken from the energetically most relevant trajectories (section “*Fluorescein orientations, Fluorescence depolarization and CcO conformational changes*”). For fluorescence depolarization, the equilibration time was 5ns.

The root-mean-square-deviations (RMSD) of the wt-CcO backbone atoms relative to the crystal structure (PDB code 3HB3 [30,31]), are below 1.4 \AA (Figure C.1, Appendix C) as it was observed for similar conditions in previous work [8,9]. MD simulations of flu-CcO show reasonable stability as well, as reflected in RMSD values below 1.6 \AA (Figure C.1, Appendix C).

Computation of pK_A values for MD trajectories of wt-CcO and flu-CcO

After generating trajectories with different protonation patterns for both wt-CcO and flu-CcO, pK_A values of relevant residues (His73, His526 and fluorescein) were computed from extracted time frames. For the electrostatic energy computations water molecules were eliminated (except for ligands of Cu_B in the R state). However, the atomic charges of membrane were considered explicitly, since the membrane is sufficiently close to His73, His526 and fluorescein (when present). Other residues were kept in the protonation state obtained from the titration of the crystal structure of CcO (PDB code: 3HB3 [30,31]), which shortened CPU time needed for pK_A computations.

pK_A values for MD trajectories were computed with software Karlsberg2⁺MD (KB2⁺MD) procedure [7], introduced in chapter 1.4 (computational details presented in chapter 3, section “*pK_A computations*”). There are two methods of evaluating pK_A values from MD trajectories.

After the titration is performed for each frame the obtained pK_A values can be averaged. The alternative would be to average the individual electrostatic energy terms for each trajectory, and use them for the computation of the pK_A values [7]. This is essentially the $KB2^+MD$ procedure described in chapter 1.4. Here, the goal was to compute the equilibrium pK_A values of the three relevant residues.

All pK_A values reported for the CcO fluorescein project were obtained with $KB2^+MD$ procedure, with few exceptions as noted in the text. However, a subtle deviation from the standard $KB2^+MD$ protocol was applied, described in chapter 3 (section “*Modification of the $KB2^+MD$ procedure*”). The Boltzmann factors obtained for each trajectory were used for computing precise pK_A values of titratable residues but also to weight trajectories [7], revealing which trajectories are energetically more relevant. In this work, trajectories which had normalized Boltzmann factors (yielding probabilities) below 0.01 were eliminated from pK_A computations and further analysis. From the remaining set of trajectories, the ones with the highest normalized Boltzmann factors are called *dominant* trajectories all others are named *significant* trajectories (terminology used in the manuscript in preparation).

Next, we discuss the computation of fluorescence depolarization of fluorescein.

Computing fluorescence depolarization

The time-dependence of fluorescence depolarization is described by the time correlation function

$$r(t) = \frac{2}{5} \langle P_2(\vec{\mu}_a(\tau) \cdot \vec{\mu}_b(\tau+t)) \rangle_\tau \quad (30)$$

where P_2 is the second order Legendre polynomial [110] and $\vec{\mu}_a$ and $\vec{\mu}_b$ are the unit length transition dipole vectors for absorption and emission spectroscopy, respectively. For fluorescein the two vectors are parallel to the axis that goes through the triple-ring system (green vector in Figure 4.3). The ensemble average represented by triangular brackets is replaced by the time average, denoted as $\langle \rangle_\tau$. Evaluating the time correlation function, eq. (30), the time interval $[0, t_{\max}-t]$ available for the average diminishes with the incremental time t . Therefore, the time correlation function is displayed only for a few nanoseconds. The fluorescence depolarization was investigated for both the oxidized and reduced state of CcO at pH 8, where fluorescein is deprotonated.

4.3. Results and Discussion

Protonation states of residues in the crystal structure

The first results are protonation states (based on pK_A values) of titratable residues from the crystal structure of CcO (PDB code: 3HB3 [30,31]). The majority of titratable residues have a “standard protonation” at pH 7 in both O and R state of CcO, which means they are in ionized states. N-terminus, Arg and Lys are protonated, with the exception of Lys354A. C-terminus, Asp, Glu and heme propionates are deprotonated, except for Glu278A, Glu481A and Asp399A, which are protonated. Histidines are a special case as they can be stable in both protonation states depending on the protein environment. At pH 7, His28A, His94A, His187A, His269A, His292A, His325A, His326A, His403A, His411A, His413A, His448A, His464A, **His73B** are in the deprotonated state as δ -tautomer; His59A, His119A, **His526A**, His36B, His181B, His224B are in the deprotonated state as ϵ -tautomer; His85A, His541A, His29B and His119B are in the protonated state.

This protonation pattern was applied in MD simulations of wt-CcO or flu-CcO for both O and R state of the enzyme, with the exception of protonation states of His73, His526 and fluorescein (when present).

Protonation states and pK_A values of residues from MD simulations of wt-CcO and flu-CcO

To obtain more precise pK_A values with the KB2⁺MD procedure, MD simulations of wt-CcO were performed. The length of these simulations was 20ns and the protonation states of His73 and His526 were varied. In order to save CPU time and simplify the investigation, only the ϵ -tautomer was used for His526 and only the δ -tautomer was used for His73 for the respective deprotonated states. This implies eight different charge patterns, i.e. two protonation states for each histidine (one protonated state and one deprotonated state) and two redox states of CcO. The following notation is introduced to describe MD trajectories: $wt_{00}(O/R)$, if both histidines are deprotonated; if only His73 is protonated, $wt_{01}(O/R)$, if only His526 is protonated, and $wt_{11}(O/R)$, if both histidines are protonated (Table 4.1).

Table 4.1. MD trajectories of oxidized (O) and reduced (R) CcO considered in further work. The dominant trajectories are the ones with the largest Boltzmann factors and are marked with bold letters **O** or **R** and an asterisk (*). Other trajectories that are of relevance for pK_A computations, but with smaller Boltzmann factors are significant trajectories and are presented with bold letters **O** or **R**.

wt-CcO trajectories ^{a, b}	flu-CcO trajectories ^{a, c}
wt ₀₀ (O/R): His ^δ 73/His ^ε 526	flu ₀₀₀ (O/R): His ^δ 73/His ^ε 526/Flu
wt ₀₁ (O/R): His ^δ 73/ His ⁺ 526	flu ₁₀₀ (O/R): His ⁺ 73/His ^ε 526/Flu
wt ₁₀ (O/R): His ⁺ 73/His ^ε 526	flu _{10u} (O/R): His ⁺ 73/His ^ε 526/Flu(u)
wt ₁₁ (O*/R*): His ⁺ 73/His ⁺ 526	flu _{10d} (O/R): His ⁺ 73/His ^ε 526/Flu(d)
	flu ₁₁₀ (O*/R): His ⁺ 73/His ⁺ 526/Flu
	flu _{11u} (O/R): His ⁺ 73/His ⁺ 526/Flu(u)
	flu _{11d} (O/ R*): His ⁺ 73/ His ⁺ 526/Flu(d)

^a The oxidation states of CcO are O for oxidized and R for reduced. Fluorescein protonation is denoted as follows: Flu for deprotonated states (with the total charge of -2) and Flu(u) or Flu(d) for protonated states (with the total charge of -1) as explained in Figure 4.3.

^b The time intervals of the trajectories used for analysis are 5-20ns for the O and R state of wt-CcO.

^c The time intervals of the trajectories used for analysis are 17-32ns for the O and R state of flu-CcO.

Table 4.2. Computed pK_A values and normalized Boltzmann factors (weights) of His73 and His526 obtained from MD simulations of for wt-CcO and flu-CcO selected for analysis as shown in Table 4.1. Karlsberg2⁺[7] was used for these computations (KB2⁺MD procedure).

oxidation state	wt-CcO trajectories ^{a,b}	flu-CcO trajectories ^{a,b}
	normalized Boltzmann factors pK _A (His73/His526)	normalized Boltzmann factors pK _A (His73/His526/Flu)
O	wt ₁₀ , wt₁₁ 0.05, 0.95 8.59/6.57	flu ₀₀₀ (O), flu₁₁₀(O) , flu _{11u} (O), flu ₁₀₀ (O) 0.01, 0.90 , 0.07, 0.02 8.41/6.77/5.55
R	wt ₀₁ , wt₁₁ 0.02, 0.98 7.63/6.09	flu ₀₀₀ (R), flu_{11d}(R) , flu _{11u} (R) 0.12, 0.82 , 0.06 7.60/6.64/6.25

^a Time intervals of the trajectories used for the pK_A computation are given in Table 4.3.

^b Trajectories of the set of eight for wt-CcO and of 14 for flu-CcO, which are not listed in the table, have normalized Boltzmann factors below 0.01.

The pK_A values and normalized Boltzmann factors of wt-CcO trajectories are shown in Table 4.2 (full list of Boltzmann factors is shown in the Appendix C, Table C.4). For both oxidized and reduced states, wt₁₁(O) and wt₁₁(R) are the dominant trajectories. The obtained pK_A values

deviate from the results for the crystal structure, much more in the case of His73, than for His526. From these results it can be concluded that at pH 7, His73 has a tendency to be protonated, while His526 has an ambiguous protonation.

The MD simulation time for pK_A computations were 32ns for flu-CcO. Considering that both His73 and His526 may be protonated or deprotonated, the charge states of the two were varied. In preliminary work (see Appendix C), deprotonated His73 was investigated in more detail and two trajectories with deprotonated His73 were selected to be used in the main work (see Table 4.1). In all other MD simulations of flu-CcO His73 was protonated, mostly based on results obtained from the computations for wt-CcO. Additionally, the protonation states of His526 and fluorescein were varied. Variation of protonation states of His525 in combination with the proximal fluorescein is important, due to the fact that a positively charged Lys299 was replaced by neutral cysteine with bound fluorescein which carries a negative charge (-1 or -2), causing a drastic change in local charge. Considering the two states of His526, three states of fluorescein and two redox states of CcO, 12 different MD simulations were carried out. Therefore, after adding the two trajectories from the preliminary work, a set of 14 trajectories was obtained. For these trajectories Boltzmann factors based on electrostatic energies were computed with Karlsberg2⁺. In all 14 MD simulations the initial orientation of fluorescein was with the angle of 180°. In preliminary MD simulations this orientation had the lowest energy.

In order to distinguish trajectories of flu-CcO with different protonation states the following notation is used (used in Table 4.1. and Table 4.2.): flu has three index symbols representing protonation states of His73, His526 and fluorescein respectively; **0** always indicates a deprotonated state, **1** indicates protonates states of histidines (only one tautomer is used for each histidine) and **u** or **d** are used for the protonated state tautomers of fluorescein (Figure 4.3.). Therefore, the trajectories from the set of 14 for the flu CcO are flu₀₀₀(O/R), flu₁₀₀(O/R), flu_{10u}(O/R), flu_{10d}(O/R), flu₁₁₀(O/R), flu_{11u}(O/R), flu_{11d}(O/R) as seen in Table 4.1. The two additional trajectories from preliminary work, denoted flu₀₀₀(O) and flu₀₀₀(R), were extended to the total of 32ns. In Table 4.2. all significant and dominant trajectories are presented that were used to obtain normalized Boltzmann factors and finally pK_A values (more details in the Appendix C, Table C.5). From these results two trajectories flu₁₁₀(O) and flu_{11d}(R) emerged as dominant. They were analyzed in more detail. However, significant trajectories are also used for the computation of pK_A values.

The influence of the presence of fluorescein on protonation states on the CcO surface was investigated by comparing the protonation states of His73 and His526 for wt-CcO and flu-CcO

(Table 4.2). The pK_A value of His73 shows a down-shift of -0.18, when adding fluorescein to the oxidized wt-CcO, while in the reduced state a marginal down-shift of -0.03 was obtained. However, no change in the protonation state of His73 was observed for these states. For His526, slightly larger upshifts were observed, which are +0.2 for oxidized and +0.55 for reduced CcO. Such pK_A shifts are not unusual, because of the proximity of the negatively charged fluorescein to His526. Similarly to His73, the pK_A values still indicate ambiguous protonation of His526. The obtained pK_A shifts show that the presence of fluorescein does not disturb wild-type features of CcO. However, analysis of the dynamics and conformations of these residues is needed to further confirm these claims.

His73 conformations and pK_A values

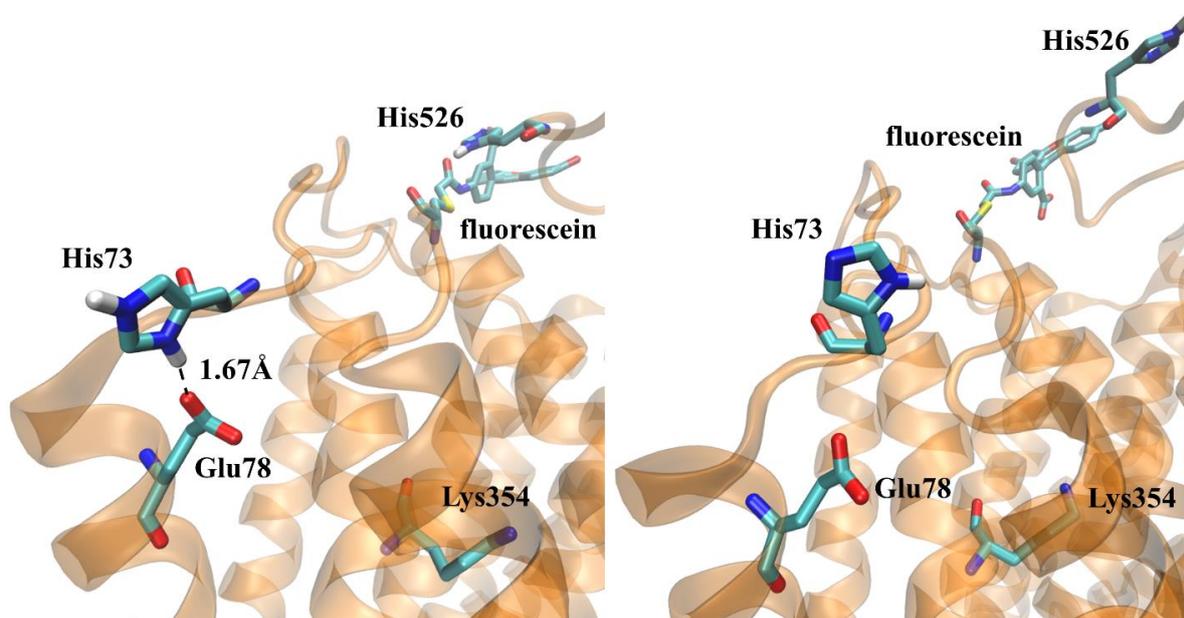


Figure 4.4. Two conformations of the residue His73 in the MD simulation of flu-CcO. **Left:** “Down” conformation of protonated His73 while forming a salt-bridge with Glu78 after 20ns in the trajectory flu₁₁₀(O). **Right:** “Up” conformation of the deprotonated His73 after 19.4ns in the trajectory flu₀₀₀(O). The membrane was omitted for better visibility.

MD simulations showed that His73 undergoes a conformational change, moving from an up to a down conformation where it forms a salt-bridge with Glu78B (referred to as Glu78). A similar salt-bridge can be found in other crystal structures of A-type CcO, e.g. *Rhodobacter Sphaeroides* (PDB code: 2GSM [28]), as documented in the Appendix C (Table C.6). This is not the case for the crystal structure used in this work (PDB code: 3HB3 [30,31]), where His73

adopts an up conformation. After inspecting the crystal unit cell, it was concluded that this is not due to crystal packing in this structure (Appendix C, Figure C.2). In all MD simulations, His73 was in the up conformation at the start. Deprotonated His73 prefers to be in the up conformation, which is observed in the trajectory flu₀₀₀(O) (Figure. 4.4, Right). In all trajectories with protonated His73, e.g. in the trajectory flu₁₁₀(O) (Figure. 4.4, Left) flipping to the down conformation is observed. The down conformation is also found in the dominant trajectories of wt-CcO [wt₁₁(O/R)] and of flu-CcO [flu₁₁₀(O), flu_{11u}(R)]. After a few nanoseconds, the protonated His73 flips into a down conformation, accompanied by an increase in the pK_A value of His73, which is expected upon formation of the salt-bridge. The shortest distances between the acidic oxygens of Glu78 and the nitrogen atoms of the imidazole ring of His73 and the pK_A values of His73 were shown in Figure 4.5 for the trajectories of wt-CcO. The salt-bridge is considered to be present at the distances shorter than 4.0 Å. The pK_A values of His73 for the dominant trajectories of flu-CcO are shown in Figure 4.6 and for the significant trajectories of both wt-CcO and flu-CcO in the Appendix C (Figure C.3, Figure C.4, respectively).

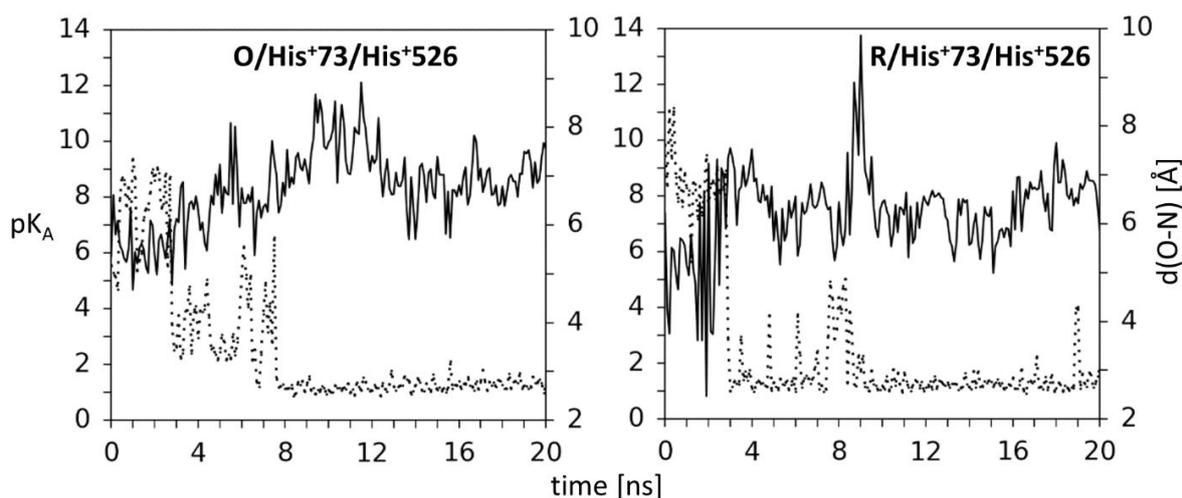


Figure 4.5. Time course of computed pK_A values of His73 (solid line, left scale) and shortest distance of acidic oxygens of Glu78 with nitrogen atoms of His73 (dotted line, right scale) from wt-CcO MD simulations for the dominant trajectories wt₁₁(O) (left part) and wt₁₁(R) (right part) listed in Table 4.2.

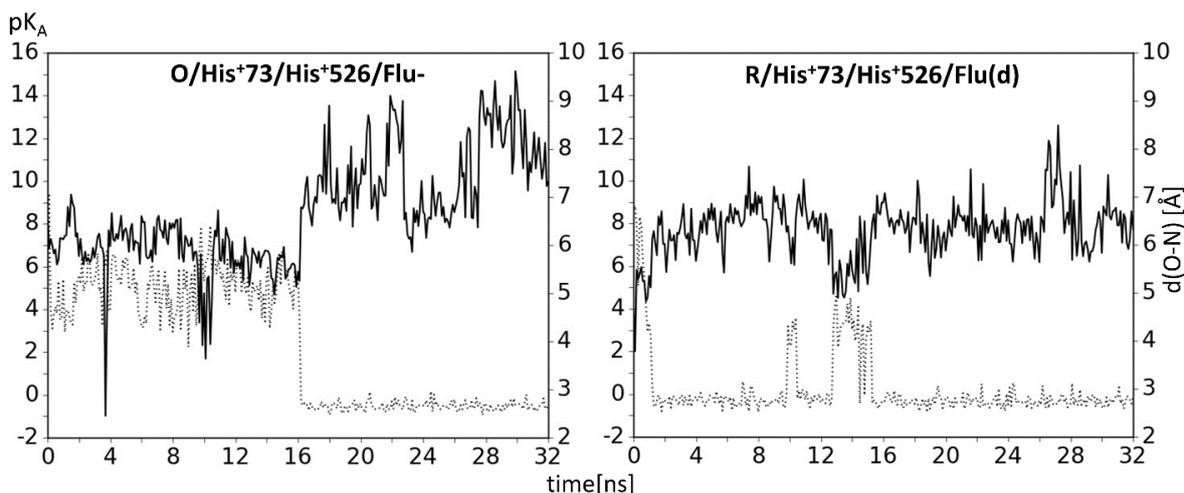


Figure 4.6. Time course of computed pK_A values of His73 (solid line, left scale) and shortest distance of acidic oxygens of Glu78 with nitrogen atoms of His73 (dotted line, right scale) from flu-CcO MD simulations for the dominant trajectories flu₁₁₀(O) (left part) and flu_{11d}(R) (right part) listed in Table 4.2.

His73 is located close to the entry point of the K-channel. Some studies suggest that an analogous glutamate, Glu101 from *Rhodobacter spheroides* is most likely the candidate for the starting point of the K-channel [98,99]. Glu78 (and the analogue Glu101) can participate in an H-bond network leading to the gating residue in the K-channel, a conserved Lys354 [8]. His73 is on the surface of CcO, exposed to the bulk water and in the proximity of the membrane on the N-side which enables it to collect an upcoming proton and transfer it into the channel via the salt-bridge with Glu78. Therefore, this residue could serve as a proton shuttle. His73 could also be part of the proton-collecting antenna, which is a pair (or a small cluster) of residues comprised typically of histidines and acidic residues that can collect upcoming protons. Some studies suggest that there are no proton-collecting antennas that are active the entire time at the surface of CcO close to the K-channel [98]. The activation could be regulated by opening and closing of this solvent exposed salt-bridge, possibly triggered by other protonation or redox changes within CcO. Moreover, the behavior of His73 is identical for wild-type CcO and if fluorescein is present, as seen from MD simulations of wt-CcO and flu-CcO.

Conformations and behavior of His526

His526 is located in the proximity of the two acidic residues Glu525A and Asp528A (referred to as Glu525 and Asp528). In MD simulations of wt-CcO, protonated His526 is stabilized by forming a weak salt-bridge with Glu525, which can be seen from the time course of distances between the nitrogen and oxygen atoms of this pair (Figure 4.7). The pK_A values of His526 are mostly between 6 and 7, meaning there is no clear protonation preference, as seen in Figure 4.8 (below) and Figure C.5 (see Appendix C). Here, the proximity of a positively charged Lys299 could lower the pK_A of His526.

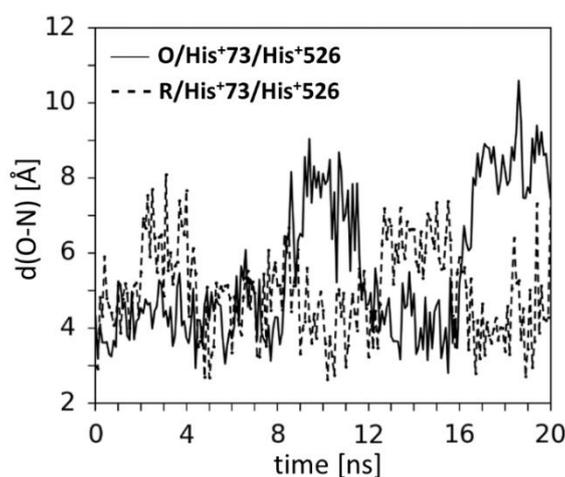


Figure 4.7. Time course of the shortest distance between the acidic oxygens of Glu525 and the nitrogens of His526, which are potential salt-bridge partners, based on the dominant trajectories of wt-CcO (dashed line, trajectory wt₁₁(O); solid line, trajectory wt₁₁(R)).

In the dominant trajectories flu₁₁₀(O) and flu_{11d}(R), the salt-bridge of His526 with Glu525 is not formed and there are no other strong electrostatic interactions with surrounding acidic residues. In some significant trajectories (Table 4.2), His526 interacts with Glu525 or Asp528. Interactions of His526 with fluorescein depend on the orientation of fluorescein and are generally rare in MD simulations. In trajectory flu_{11u}(R), His526 forms an H-bond with the oxygen atoms of the triple-ring system of fluorescein for only about 3 ns.

The charge pattern in flu-CcO is significantly different from wt-CcO, since there is a negatively charged fluorescein instead of a positively charged Lys299. The negative charge of fluorescein stabilizes the protonated state of His526, which results in a small upshift in the pK_A value (Figure 4.8, Table 4.2 and Figure C.6 (in the Appendix C)). The ambiguous pK_A values of His526 in all considered cases do not indicate an important functional role of this residue. However, by interacting with the nearby acids Asp528 and Glu525, His526 could hold a proton.

Hence, His526 paired with either Asp528 or Glu525 could be part of a proton-collecting antenna [98,111]. The functional role of His526, it could be affected by labeling the enzyme with fluorescein, but the pK_A values from Table 4.2 show a rather small influence.

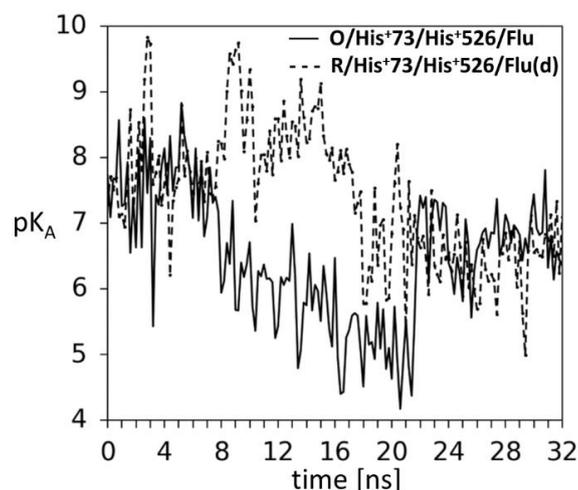


Figure 4.8. Time course of computed pK_A values of His526 in flu-CcO MD simulations (dashed line, trajectory flu₁₁₀(O); solid line, trajectory flu_{11d}(R)).

Fluorescein pK_A values and implications on the CcO conformational changes

Fluorescein pK_A values were obtained from MD simulations after allowing enough equilibration of the flu-CcO system. For evaluation of energetic quantities, such as pK_A values the last 15ns out of 32ns of the simulation have been used, which was explained in the Appendix C (Evaluating the fluorescein pK_A shift). In Figure 4.9 the time course of the fluorescein pK_A value from dominant trajectories flu₁₁₀(O) and flu_{11d}(R) is presented. Other plots for all flu-CcO trajectories from Table 4.2 are shown in Figure C.7 (Appendix C).

Fluorescein pK_A values computed from trajectories in Table 4.2 show a upshift from 5.55 in the oxidized to 6.25 in the reduced state of CcO, yielding the pK_A upshift $\Delta pK_A(R-O) = +0.70$ as seen in Figure 4.10(b). The computed pK_A upshift is large enough to be significant according to the computational accuracy. The pK_A value of fluorescein was measured experimentally, with methods described in the Appendix C. Titrations of samples with CcO labeled with fluorescein show a pK_A of 6.69 ± 0.04 when the enzyme is oxidized and 7.05 ± 0.05 in the reduced state, yielding a pK_A upshift of $\Delta pK_A(R-O) = +0.36$ (Figure 4.10(a)). In comparison to the experimental shift, the shift obtained from computations is about twice as large, however both approaches show an upshift of the fluorescein pK_A , when moving from O to R state of the enzyme.

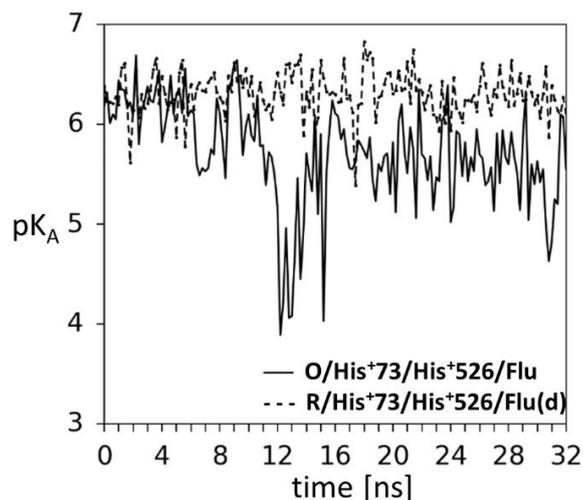


Figure 4.9. Time course of computed pK_A values of fluorescein for the dominant trajectories $flu_{110}(O)$ and $flu_{11d}(R)$ of $flu-CcO$ listed in Table 4.2.

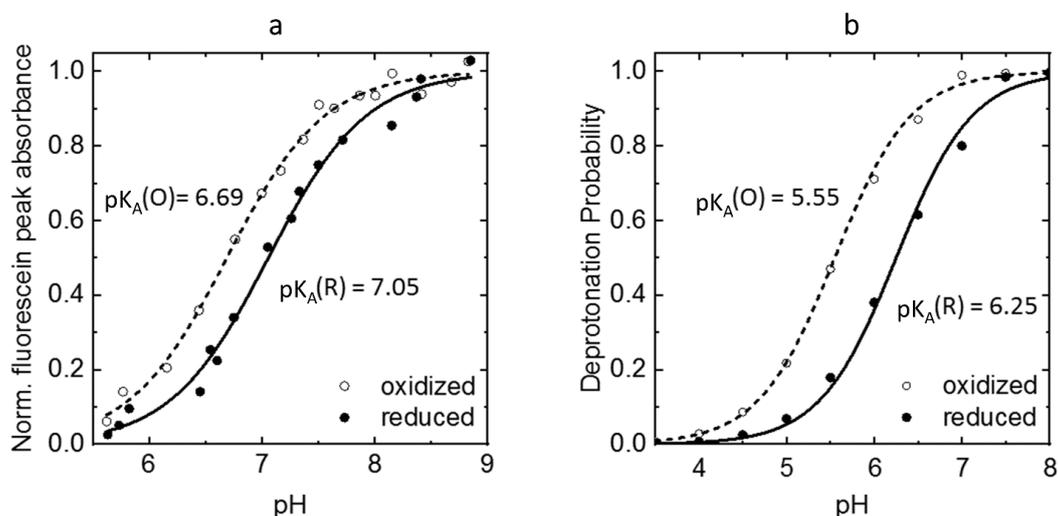


Figure 4.10. Titration curves of fluorescein **(a)** derived from the absorbance-based (UV-Vis spectroscopy) titration yielding an upshift of the pK_A value of +0.36 upon reduction of CcO (done by A.Wolf). **(b)** derived from the titration of time frames with Karlberg2⁺; yielding an upshift of the pK_A value of +0.70 (Table 4.2)(dashed line: oxidized CcO , solid line: reduced CcO).

Analysis of MD simulations can provide insight into the origin of the obtained pK_A shift. Three different factors can be discriminated: (i) a direct effect from the atomic partial charges that change with the redox state of CcO cofactors (BNC, heme_a or Cu_A); (ii) an indirect effect from the protonation changes of nearby residues (His73 and His526) induced by the redox change; (iii) and indirect effect of the orientation of fluorescein coupled with conformational changes of the protein induced by the change in redox state.

To understand all three factors influencing the pK_A values of fluorescein, they were computed using the time frames of dominant trajectories for one redox state (O or R) but by using atomic partial charges of Cu_A, heme_a and BNC from the corresponding other redox state (R or O). To investigate the influence of the factors (i) and (ii) a single time frame with the largest Boltzmann factor of the dominant trajectory was used. For the factor (i), only fluorescein was titrated, while keeping all other residues in the fixed protonation state obtained from the crystal structure, defined above. The combined effect of (i) and (ii) was estimated by titrating fluorescein and both His73 and His526 while keeping all other residues in the fixed protonation state. Lastly, if the conformational changes are included, the influence of all factors ((i), (ii) and (iii)) is obtained. For this purpose, all time frames of dominant trajectories were used. The results of these computations are shown in Table 4.3. For the factors (i) and (ii) the shifts when moving from one redox state to another is negligible. After including conformational changes (factor (iii)), i.e. if the coordinates of the protein atoms are from trajectories that originally had different redox states of CcO, the shift becomes significant. These pK_A values are different compared to the ones in Table 4.2, because only dominant trajectories are used.

Table 4.3. Fluorescein pK_A values computed with Karlsberg2⁺ using time frames of the dominant trajectories for oxidized and reduced CcO. pK_A values shown in bold digits correspond to charges of the proper redox state in MD simulations.

charges from	titration of Flu only ^a		titration of Flu, His73 and His526 ^a		with dominant trajectory only ^b	
	CcO(O) ^c	CcO(R) ^d	CcO(O) ^c	CcO(R) ^c	CcO(O) ^c	CcO(R) ^d
flu ₁₁₀ (O)	4.48	4.55	4.49	4.57	5.43	5.48
flu _{11d} (R)	6.10	6.12	6.32	6.33	6.27	6.29

^a Single time frame has been used for this pK_A computation.

^b In contrast to the pK_A values listed in Table2 only data from the dominant trajectory are used.

^c Charges of CcO cofactors correspond to the oxidized state.

^d Charges of CcO cofactors correspond to the reduced state.

The charge-charge interactions between fluorescein and the redox centers of CcO (Cu_A, heme_a and BNC) or with the charges of His73 and His526 have a marginal influence on the pK_A shift of fluorescein. The conformational changes within CcO induced by the redox changes of cofactors are the main influence behind the shift. It is not a global conformational change, as seen in low backbone RMSD values but rather a rearrangement of the H-bond network that involves fluorescein. The neighborhood of fluorescein suffers a local conformational change and rearrangement of the H-bond network with the change from O to R state of CcO, inducing

the pK_A shift. These changes are hard to observe directly in the experiment, but fluorescence depolarization investigation provided some more insight into the conformational changes around the fluorescein.

Fluorescein orientations, Fluorescence depolarization and CcO conformational changes

Fluorescein attached to the CcO surface has a tendency to change its orientation, which depends on the environment, as seen in Figure 4.11(left). This is a clear evidence that fluorescein is able to reorient completely, however due to the corrugated surface of CcO it is much slower than systems with little steric interferences [110]. The movement of fluorescein is not distorting the surface of CcO significantly, but has some influence on the properties of the nearby residues. Fluorescein is located at specific place on the CcO surface, attached to the cysteine in the position 299 on the loop that is an extension of the Helix VI to which catalytic Tyr280 belongs (Figure 4.1).

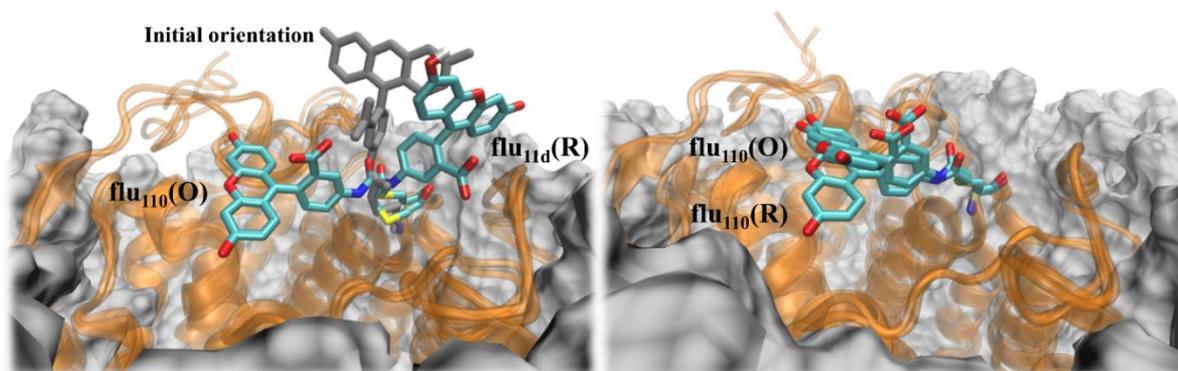


Figure 4.11. Left: Two conformations of the fluorescein bound to Cys299 in flu-CcO. Gray: initial orientation of 180° as defined in Figure 4.3. Green: final conformation of fluorescein after 32ns of MD simulations for the dominant trajectories flu₁₁₀(O) (with deprotonated fluorescein) and flu_{11d}(R) (with protonated fluorescein). **Right:** Conformations of deprotonated fluorescein at pH 8 in MD simulations of trajectories flu₁₁₀(O) and flu₁₁₀(R) after 32ns of MD simulations.

Figure 4.11(left) shows the preference for certain orientations of fluorescein in the dominant trajectories for the two redox states of CcO flu₁₁₀(O) and flu_{11d}(R). Similar observations were made for significant trajectories. Further work proved that it is rather the protonation of fluorescein that determines the orientation relative to the CcO surface. On the other hand, the

redox state of the CcO has an influence on the flexibility of the dye, which can be described by fluorescence depolarization shown as anisotropy decay curves.

In the experiments, the orientational dynamics of fluorescein in CcO using picosecond time-resolved fluorescence anisotropy was measured. From anisotropy decay curves, the information about global and local protein dynamics and conformational changes can be obtained [112–114]. Therefore, anisotropy decay curves can indirectly provide information on conformational changes that affect not only residues around fluorescein but more importantly Helix VI and Helix VII, which surround the K-channel. The data from MD simulations can also be used to plot anisotropy decay curves, as described in Methods. In the experiments, the measurements were done at pH 8 for both redox states of the enzyme. However, at this pH value, fluorescein is certainly deprotonated. Therefore, the dominant trajectory flu₁₁₀(O) used for the analysis was extended to the final time of 47 ns. The deprotonated state of fluorescein differs from the trajectory flu_{11d}(R), where fluorescein is protonated. Hence, a new trajectory was generated, named flu₁₁₀(R) with deprotonated fluorescein the conformation of fluorescein at the end of the trajectory flu₁₁₀(O) as starting conformation, which is more appropriate for the deprotonated state (Figure 4.11(right)). Finally, 30ns after equilibration of both trajectories (flu₁₁₀(R), flu₁₁₀(O)) were used to plot the anisotropy decay curves (Figure 4.12).

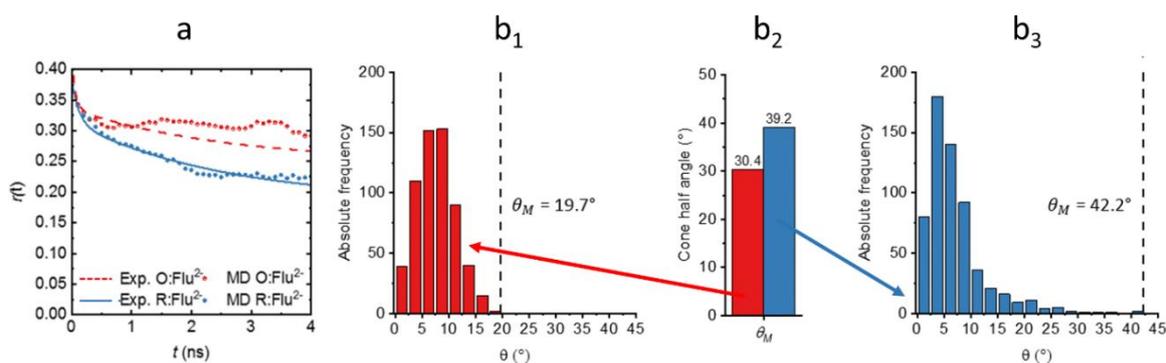


Figure 4.12. (a) Time-dependence of fluorescence anisotropy of oxidized (red) and reduced (blue) CcO, from experiment (dashed red line and full blue line) and MD simulation data (dotted lines) (done by A. Wolf). Fluorescein bound to the reduced CcO (solid line) has a faster anisotropy decay than for oxidized CcO. (b1–b2–b3) Angular distribution of “cone-in-a-cone” model. Based on the time-resolved anisotropy data the cone half angle θ_M increases upon reduction of CcO (b2). Angular frequency distributions from MD simulation data: (b1): oxidized CcO with a maximal semicone angle of 19.7°; (b3): reduced CcO with a maximal semicone angle of 42.2, showing agreement with experimental results from (b2). Experimental conditions: 20 mM potassium phosphate buffer solution at pH 8, 20 mM sodium chloride, 0.05 % β -dodecyl-maltoside, 20°C. The figure is adapted from the manuscript in preparation.

In MD simulation data, the anisotropy decay curves do not contain the component that originates from the protein tumbling, since the time scale of the simulation does not allow this. Therefore, the anisotropy decay stems from the dye itself and the conformational changes of the loop at the end of the Helix VI. The fluorescence depolarization (Figure 4.12) based on MD simulation data (blue and red dotted lines) agrees with experiment (blue full line and red dashed line). In theory and experiment the mobility of fluorescein is lower for the oxidized state. Fluorescein movement can be described using two models “cone-in-a-cone” and “wobbling-in-a-cone” [110], where the former shows the difference that comes from the motion of the immediate environment of the label and the label itself. Furthermore, the angular frequency distributions of the dye (Figure 4.12) give more insight into the motion. The vector used for this analysis is defined from the sulfur atom of Cys299 to the central oxygen of the triple-ring system of fluorescein (extension of the blue vectors from Figure 4.3). The cone half angle from computations can be compared to experiment. In the reduced state, the angle distribution is broader, indicating more flexibility. In addition, the maximal angles show a decent agreement of theory and experiment, more so in the reduced state. In the oxidized state, the maximal angle is smaller, which implies a more restricted movement of the dye.

In the experiments, different components of the anisotropic decay demonstrated that upon reduction, the flexibility of Helix VI and the loop flexibility increase. Both experimental work and the MD simulations show that a local conformational change takes place upon the reduction of the CcO cofactors. These suggested subtle conformational changes within CcO could be related to the previously observed changes in the H-bond pattern changing from O to R state, causing the pK_A shift of fluorescein for example. Studies have shown that flexibility of the loop between the Helix VI and Helix VII increases in the transition between O and R state of CcO [115]. With this, we can propose that the change in flexibility involves the end of the Helix VI at the CcO surface (where fluorescein is attached) as part of K-channel activity.

4.4. Summary and Conclusions

MD simulations and electrostatic energy computations were performed on the wild type CcO (wt-CcO) and the K299C mutant CcO with the attached fluorescein label (flu-CcO). These results support the experimental work based on fluorescence labeling of the mutated CcO from *Paracoccus denitrificans*. The main goal of this work was estimation of protonation changes at the surface of the CcO in relation to the activity of the K-channel in the O and R state.

Fluorescein bound on the surface of the CcO does not disrupt the wild type behavior, since only marginal changes in the protonation of the residues His73 and His526 are observed. His526 has an ambiguous pK_A value in the wild type with values between 6 and 7, which does not change after fluorescein is attached. The proximity of fluorescein is seen as a small upshift of the pK_A value of this residue. His526 is able to interact with Asp528 and Glu525, exhibiting behavior of a proton collecting antenna [111]. The pK_A values of His73 are affected to an even lesser extent than of His526. His73 may be functionally relevant. Namely, in MD simulations protonated His73 readily forms a salt-bridge with Glu78. A similar salt-bridge pair was found in the crystal structure of another A-type CcO from *Rhodobacter Sphaeroides* (PDB code: 2GSM [28]), with Glu101 being suggested as the entry point of the K-channel [99]. Deprotonated His73 is very close to the membrane surface and it is exposed to bulk water, where it can transfer an upcoming proton to Glu78 [98]. Thus, His73 could be a part of the proton collecting antenna at the surface that can be activated when necessary and act as a proton shuttle for the K-channel. This behavior is completely unchanged when fluorescein is present, proving that the CcO function remains unchanged in simulations and experiments upon fluorescence labeling.

The fluorescein pK_A value changes between O and R state of CcO. Computation of pK_A values from electrostatic energies of MD frames, showed an upshift of the pK_A value of fluorescein of +0.70 from O to R state of CcO (Figure 4.10). This is in agreement with the measured shift of +0.36 which is smaller, but still an upshift upon reduction. More detailed computations showed that the shift does not originate directly from the charges of CcO cofactors (BNC, Cu_A and heme a) or from the changes in protonation of the residues in the vicinity. It is rather a result of conformational changes of CcO in the proximity of fluorescein affecting the H-bond pattern due to changes of the charge pattern of cofactors (Table 4.2, Table 4.3). Fluorescence depolarization was investigated by experiment and computation, yielding good agreement (Figure 4.12). Lower fluorescein flexibility is observed in the O state of CcO, as seen from

anisotropy curves and the analysis of the cone half angle of a “cone-in-a-cone” movement of fluorescein (Figure 4.12). Results from experiments showed that in the reduced state a loop at the end of Helix VI becomes more flexible. Differences in flexibility of the loop of Helix VI that contains catalytic Tyr280, could indicate the opening or the activation of the K-channel, triggered by the redox change of CcO.

5. Conclusions and Outlook

This present dissertation covers three tasks. The first task was to improve the protocol for pK_A computations, Karlsberg2⁺ (KB2⁺). The other two tasks involve electrostatic energy computations and pK_A computations answering questions relevant for energy metabolism, i.e. to examine the function of cytochrome *c* oxidase (CcO).

The flexible, modular software KB2⁺ is considerably faster than the preceding version, Karlsberg⁺ (KB⁺)[16] due to more efficient modeling of PACs. In the procedure of KB⁺, modeling of PACs involves salt-bridge perturbation with many modeling steps, which is inefficient. In KB2⁺, the approach of “repelling charges” is applied for PAC generation and allows perturbation of all salt-bridges in only two modeling steps. Consequently, KB2⁺ is on the average 16 times faster than KB⁺. However, the precise CPU time reduction depends on the number of salt-bridges, which need to be perturbed. The accuracy of the KB2⁺ is equivalent to KB⁺. There was no improvement of results, since this approach is likely at the limits. Both, KB2⁺ and KB⁺, have in common that only local conformational variations are introduced to represent the pH dependency of the protein structure. In a newly developed KB2⁺MD procedure [7], MD simulations are used for PAC generation which yields a better agreement with experimental results at the cost in CPU time. Nevertheless, KB2⁺ is a good choice for fast pK_A computations based on protein crystal structures. There are still several ideas for improvements. The charges used for electrostatic energy computations can be refined to obtain better results and additional mild relaxation of the protein structure could be included in PAC modeling procedures.

In CcO, residues that may act as the proton loading site (PLS) were identified and characterized and the changes on the protein surface related to the activity of the K-channel were investigated. The proton loading site candidates investigated in this dissertation are the propionates A- (PRA) and D- (PRD) of heme a and heme a₃. All four propionates are involved in salt-bridges or hydrogen bonds in the crystal structure, rendering them incapable to accept a proton. Alternative stable conformations were found, where these interactions are not formed, and in which PRDa₃ and PRAa₃ may act as the PLS. The propionates of heme a are eliminated as PLS candidates. The change of pK_A values coupled with the redox state of CcO for PRDa₃ and PRAa₃ are somewhat less than the pump requirements. This could suggest that other structural changes in the vicinity of propionates of heme a₃ could happen in order to help the PLS

function. These questions remain to be answered. Details on proton uptake and release [9], hydration of the nearby cavity and even combination of several PLS candidates [12,39,42,43,45,46,71–73] may be systematically explored for the entire CcO redox cycle. The likely assisting residue to PRDa₃ and PRA₃ as PLS is a His ligand of the Cu_B-complex, which requires a QM approach for further investigation.

The protonation changes close to the K-channel entry were investigated computationally and experimentally. This work is the first study that connects experiments to simulations of the labelled CcO in the membrane. In the reductive part of the CcO cycle, when the K-channel is active, the measured pK_A of the fluorescent label on the CcO surface is shifted due to the redox change of CcO cofactors, which was confirmed by pK_A computations. The negligible direct influence of the redox change of the BNC on the pK_A values suggests that there likely subtle hydrogen bond rearrangements connecting the BNC with the CcO surface. Fluorescence depolarization experiments and computations showed that the loop at the end of Helix VI could be involved in the redox change induced conformational change. It was shown that labelling has little effect on the protonation of residues on the CcO surface compared to the wild type enzyme, which is of great significance for future trustworthiness of labeling experiments. Additionally, His73 is suggested as the proton entry point of the K-channel. While these results show interesting addition to the proposed K-channel activity [8], more computational work is needed to explain details concerning hydrogen bond rearrangements in wild type CcO. More labeling experiments may be carried out as well.

In this dissertation, it was shown that electrostatic interactions are tightly related to the protein function, e.g. salt-bridge opening and hydrogen bond rearrangement, as seen throughout the two projects. Electrostatic energy computations depend on the details of used structures, i.e. the conformations of the protein structure. General issue with pK_A computations is the quality of the physical model, or the approximation, used for the electrostatic energy computations. With the continuum approach, it can often happen that pK_A values are not in an ideal agreement with experimental results, however, pK_A differences show reasonable agreement, as seen in investigations of the CcO function. In general, the ultimate goal of research is to connect experiment and computation successfully, which was deemed successful in this work.

6. References

- [1] M.K. Gilson, A. Rashin, R. Fine, B. Honig, On the calculation of electrostatic interactions in proteins, *J. Mol. Biol.* 183 (1985) 503–516.
- [2] E. Alexov, E.L. Mehler, N. Baker, A. M. Baptista, Y. Huang, F. Milletti, J. Erik Nielsen, D. Farrell, T. Carstensen, M.H.M. Olsson, J.K. Shen, J. Warwicker, S. Williams, J.M. Word, Progress in the prediction of pKa values in proteins, *Proteins Struct. Funct. Bioinforma.* 79 (2011) 3260–3275.
- [3] A. Warshel, Electrostatic Basis of Structure-Function Correlation in Proteins, *Acc. Chem. Res.* 14 (1981) 284–290.
- [4] C.N. Pace, G.R. Grimsley, J.M. Scholtz, Protein ionizable groups: pK values and their contribution to protein stability and solubility, *J. Biol. Chem.* 284 (2009) 13285–13289.
- [5] H.-X.X. Zhou, X. Pang, Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation, *Chem. Rev.* 118 (2018) 1691–1741.
- [6] K. Bartik, C. Redfield, C.M. Dobson, Measurement of the individual pKa values of acidic residues of hen and turkey lysozymes by two-dimensional ¹H NMR, *Biophys. J.* 66 (1994) 1180–1184.
- [7] T. Meyer, E.W. Knapp, pKa Values in Proteins Determined by Electrostatics Applied to Molecular Dynamics Trajectories, *J. Chem. Theory Comput.* 11 (2015) 2827–2840.
- [8] A.L. Woelke, G. Galstyan, E.W. Knapp, Lysine 362 in cytochrome c oxidase regulates opening of the K-channel via changes in pK_A and conformation, *Biochim. Biophys. Acta - Bioenerg.* 1837 (2014) 1998–2003.
- [9] A.L. Woelke, G. Galstyan, A. Galstyan, T. Meyer, J. Heberle, E.W. Knapp, Exploring the Possible Role of Glu286 in C c O by Electrostatic Energy Computations Combined with Molecular Dynamics, *J. Phys. Chem. B.* 117 (2013) 12432–12441.
- [10] A.L. Woelke, A. Wagner, G. Galstyan, T. Meyer, E.W. Knapp, Proton Transfer in the K-Channel Analog of B-Type Cytochrome c Oxidase from *Thermus thermophilus*, *Biophys. J.* 107 (2014) 2177–2184.
- [11] V.R. Kaila, M.I. Verkhovsky, M. Wikström, Proton-coupled electron transfer in cytochrome oxidase., *Chem. Rev.* 110 (2010) 7062–7081.
- [12] M. Wikström, V. Sharma, V.R.I. Kaila, J.P. Hosler, G. Hummer, New Perspectives on Proton Pumping in Cellular Respiration, *Chem. Rev.* 115 (2015) 2196–2221.
- [13] C.N. Schutz, A. Warshel, What are the dielectric “constants” of proteins and how to validate electrostatic models?, *Proteins Struct. Funct. Genet.* 44 (2001) 400–417.
- [14] T. Simonson, C.L. Brooks, Charge screening and the dielectric constant of proteins: Insights from molecular dynamics, *J. Am. Chem. Soc.* 118 (1996) 8452–8458.
- [15] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome., *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 10037–41.

- [16] G. Kieseritzky, E.W. Knapp, Optimizing pK_A computation in proteins with pH adapted conformations, *Proteins Struct. Funct. Genet.* 71 (2008) 1335–1348.
- [17] B. Rabenstein, E.W. Knapp, Calculated pH-Dependent Population and Protonation of Carbon-Monoxo-Myoglobin Conformers, *Biophys. J.* 80 (2001) 1141–1150.
- [18] T.J. You, D. Bashford, Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility, *Biophys. J.* 69 (1995) 1721–1733.
- [19] G.M. Ullmann, E.W. Knapp, Electrostatic models for computing protonation and redox equilibria in proteins, *Eur. Biophys. J.* 28 (1999) 533–551.
- [20] D. Bashford, M. Karplus, pK_a's of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model, *Biochemistry.* 29 (1990) 10219–10225.
- [21] B.R. Brooks, C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: The biomolecular simulation program, *J. Comput. Chem.* 30 (2009) 1545–1614.
- [22] J.A. McCammon, B.R. Gelin, M. Karplus, Dynamics of folded proteins, *Nature*, 267 (1977) 585–590.
- [23] C.J. Cramer, *Essentials of Computational Chemistry Theories and Models*, 2004.
- [24] H.M. Berman, The Protein Data Bank, *Nucleic Acids Res.* 28 (2000) 235–242.
- [25] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †, *J. Phys. Chem. B.* 102 (1998) 3586–3616.
- [26] D.L. Nelson, A.L. Lehninger, M.M. Cox, *Lehninger principles of biochemistry*, Macmillan, 2008.
- [27] J.M. Berg, J.L. Tymoczko, G.J. Gatto Jr, *Stryer: Biochemistry*, WH Free. Co. 5 (2002) 306–307.
- [28] L. Qin, C. Hiser, A. Mulichak, R.M. Garavito, S. Ferguson-Miller, Identification of conserved lipiddetergent-binding sites in a high-resolution structure of the membrane protein cytochrome c oxidase, *Proc. Natl. Acad. Sci.* 103 (2006) 16117-16122.
- [29] D.M. Popovic, I. V. Leontyev, D.G. Beech, A.A. Stuchebrukhov, Similarity of cytochrome c oxidases in different organisms, *Proteins Struct. Funct. Bioinforma.* 78 (2010) 2691–2698.
- [30] C. Ostermeier, A. Harrenga, U. Ermler, H. Michel, Structure at 2.7 Å resolution of the *Paracoccus denitrificans* two-subunit cytochrome c oxidase complexed with an antibody FV fragment, *Proc. Natl. Acad. Sci.* 94 (1997) 10547–10553.

- [31] J. Koepke, E. Olkhova, H. Angerer, H. Müller, G. Peng, H. Michel, High resolution crystal structure of *Paracoccus denitrificans* cytochrome *c* oxidase: New insights into the active site and the proton transfer pathways, *BBA - Bioenerg.* 1787 (2009) 635–645.
- [32] G. Gilderson, L. Salomonsson, A. Aagaard, J. Gray, P. Brzezinski, J. Hosler, Subunit III of cytochrome *c* oxidase of *Rhodobacter sphaeroides* is required to maintain rapid proton uptake through the D pathway at physiologic pH, *Biochemistry.* 42 (2003) 7400–7409.
- [33] T. Tsukihara, K. Shimokata, Y. Katayama, H. Shimada, K. Muramoto, H. Aoyama, M. Mochizuki, K. Shinzawa-Itoh, E. Yamashita, M. Yao, Y. Ishimura, S. Yoshikawa, The low-spin heme of cytochrome *c* oxidase as the driving element of the proton-pumping process, *Proc. Natl. Acad. Sci.* 100 (2003) 15304–15309.
- [34] L. Qin, J. Liu, D.A. Mills, D.A. Proshlyakov, C. Hiser, S. Ferguson-Miller, Redox-Dependent Conformational Changes in Cytochrome *c* Oxidase Suggest a Gating Mechanism for Proton Uptake, *Biochemistry.* 48 (2009) 5121–5130.
- [35] M. Svensson-Ek, J. Abramson, G. Larsson, S. Törnroth, P. Brzezinski, S. Iwata, S. Tornroth, P. Brzezinski, S. Iwata, The X-ray crystal structures of wild-type and EQ(I-286) mutant cytochrome *c* oxidases from *Rhodobacter sphaeroides*, *J. Mol. Biol.* 321 (2002) 329–339.
- [36] A.A. Stuchebrukhov, Redox-Driven Proton Pumps of the Respiratory Chain, *Biophys. J.* 115 (2018) 830–840.
- [37] N. Agmon, The Grotthuss mechanism, *Chem. Phys. Lett.* 244 (1995) 456–462.
- [38] V. Sharma, M. Wikström, P. Bernardi, *Biochimica et Biophysica Acta* The role of the K-channel and the active-site tyrosine in the catalytic mechanism of cytochrome *c* oxidase, *BBA - Bioenerg.* 1857 (2016) 1111–1115.
- [39] D.M. Popovic, A.A. Stuchebrukhov, Proton Exit Channels in Bovine Cytochrome *c* Oxidase, *J. Phys. Chem.* 109 (2005) 1999–2006.
- [40] P. Brzezinski, R.B. Gennis, Cytochrome *c* oxidase: Exciting progress and remaining mysteries, *J. Bioenerg. Biomembr.* 40 (2008) 521–531.
- [41] Y.C. Kim, M. Wikstrom, G. Hummer, Kinetic gating of the proton pump in cytochrome *c* oxidase, *Proc. Natl. Acad. Sci.* 106 (2009) 13707–13712.
- [42] A. V. Pislakov, P.K. Sharma, Z.T. Chu, M. Haranczyk, A. Warshel, Electrostatic basis for the unidirectionality of the primary proton transfer in cytochrome *c* oxidase, *Proc. Natl. Acad. Sci.* 105 (2008) 7726–7731.
- [43] J. Lu, M.R. Gunner, Characterizing the proton loading site in cytochrome *c* oxidase, *Proc. Natl. Acad. Sci.* 111 (2014) 12414–12419.
- [44] D.M. Popovi, J. Quenneville, A.A. Stuchebrukhov, DFT/Electrostatic Calculations of pK_a Values in Cytochrome *c* Oxidase, *J. Phys. Chem.* 109 (2005) 3616–3626.
- [45] V.R.I. Kaila, V. Sharma, M. Wikström, The identity of the transient proton loading site of the proton-pumping mechanism of cytochrome *c* oxidase, *Biochim. Biophys. Acta - Bioenerg.* 1807 (2011) 80–84.

- [46] S. Supekar, A.P. Gamiz-Hernandez, V.R.I.I. Kaila, A Protonated Water Cluster as a Transient Proton-Loading Site in Cytochrome c Oxidase, *Angew. Chemie - Int. Ed.* 55 (2016) 11940–11944.
- [47] M. Wikström, C. Ribacka, M. Molin, L. Laakkonen, M. Verkhovsky, A. Puustinen, Gating of proton and water transfer in the respiratory enzyme cytochrome c oxidase., *Proc. Natl Acad. Sci. USA.* 102 (2005) 10478–81.
- [48] V.R.I. Kaila, M.I. Verkhovsky, G. Hummer, M. Wikström, Glutamic acid 242 is a valve in the proton pump of cytochrome c oxidase., *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 6255–6259.
- [49] K. Stierand, M. Rarey, Drawing the PDB: Protein–Ligand Complexes in Two Dimensions, *ACS Med. Chem. Lett.* 1 (2010) 540–545.
- [50] M.R. Gunner, N.A. Baker, Continuum Electrostatics Approaches to Calculating pK a s and E m s in Proteins, *Comput. Approaches Stud. Enzym. Mech. Part B.* 578 (2016) 1–20.
- [51] P. Beroza, D.A. Case, Including side chain flexibility in continuum electrostatic calculations of protein titration, *J. Phys. Chem.* 100 (1996) 20156–20163.
- [52] L. Nilsson, A. Karshikoff, Multiple pH regime molecular dynamics simulation for pK calculations, *PLoS One.* 6 (2011).
- [53] M. Marquart, J. Walter, J. Deisenhofer, W. Bode, R. Huber, The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen and its Complexes with Inhibitors, *Acta Crystallogr.,Sect.B.* 39 (1983) 480.
- [54] T. Gallagher, P. Alexander, P. Bryan, G.L. Gilliland, Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR., *Biochemistry.* 33 (1994) 4721–4729.
- [55] C. Martin, V. Richard, M. Salem, R. Hartley, Y. Mauguén, Refinement and structural analysis of barnase at 1.5 Å resolution., *Acta Crystallogr.,Sect.D.* 55 (1999) 386–398.
- [56] M. Ramanadham, L.C. Sieker, L.H. Jensen, Refinement of triclinic lysozyme: II. The method of stereochemically restrained least squares., *Acta Crystallogr.,Sect.B.* 46 (1990) 63–69.
- [57] B. Howlin, D.S. Moss, G.W. Harris, Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model., *Acta Crystallogr.,Sect.A.* 45 (1989) 851–861.
- [58] K. Katayanagi, M. Miyagawa, M. Matsushima, M. Ishikawa, S. Kanaya, H. Nakamura, M. Ikehara, T. Matsuzaki, K. Morikawa, Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution., *J.Mol.Biol.* 223 (1992) 1029–1052.
- [59] E.Y. Jones, S.J. Davis, A.F. Williams, K. Harlos, D.I. Stuart, Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2, *Nature.* 360 (1992) 232.
- [60] D.M. Szebenyi, K. Moffat, The refined structure of vitamin D-dependent calcium-binding protein from bovine intestine. Molecular details, ion binding, and implications for the structure of other calcium-binding proteins., *J.Biol.Chem.* 261 (1986) 8761–8777.

- [61] W. Bode, A.Z. Wei, R. Huber, E. Meyer, J. Travis, S. Neumann, X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor., *EMBO J.* 5 (1986) 2453–2458.
- [62] A. Weichsel, J.R. Gasdaska, G. Powis, W.R. Montfort, Crystal structures of reduced, oxidized, and mutated human thioredoxins: evidence for a regulatory homodimer, *Structure.* 4 (1996) 735–751.
- [63] R.L. Campbell, D.R. Rose, W.W. Wakarchuk, R.J. To, W. Sung, M. Yaguchi, High-Resolution Structures of Xylanases from *B. Circulans* and *T. Harzianum* Identify a New Folding Pattern and Implications for the Atomic Basis of the Catalysis, to be published
- [64] E.K. Oshea, J.D. Klemm, P.S. Kim, T. Alber, X-Ray Structure of the Gcn4 Leucine Zipper, a 2-Stranded, Parallel Coiled Coil, *Science* (80-.). 254 (1991) 539–544.
- [65] C.A. Castañeda, C.A. Fitch, A. Majumdar, V. Khangulov, J.L. Schlessman, B.E. García-Moreno, Molecular determinants of the pK_A values of Asp and Glu residues in staphylococcal nuclease, *Proteins Struct. Funct. Bioinforma.* 77 (2009) 570–588.
- [66] M.D. Joshi, A. Hedberg, L.P. McIntosh, Complete measurement of the pK_A values of the carboxyl and imidazole groups in *Bacillus circulans* xylanase, *Protein Sci.* 6 (2008) 2667–2670.
- [67] W.M. Matousek, B. Ciani, C.A. Fitch, B. Garcia-Moreno E., R.A. Kammerer, A.T. Alexandrescu, Electrostatic Contributions to the Stability of the GCN4 Leucine Zipper Structure, *J. Mol. Biol.* 374 (2007) 206–219.
- [68] W. Im, M.S. Lee, C.L. Brooks, Generalized Born Model with a Simple Smoothing Function, *J. Comput. Chem.* 24 (2003) 1691–1702.
- [69] M. Oliveberg, V.L. Arcus, A.R. Fersht, pK_A Values of Carboxyl Groups in the Native and Denatured States of Bamase: The pK_A Values of the Denatured State Are on Average 0.4 Units Lower Than Those of Model Compounds, *Biochemistry.* 34 (1995) 9424–9433.
- [70] I. Belevich, D.A. Bloch, N. Belevich, M. Wikström, M.I. Verkhovsky, Exploring the proton pump mechanism of cytochrome *c* oxidase in real time, *Proc. Natl. Acad. Sci.* 104 (2007) 2685-2690.
- [71] R. Sugitani, E.S. Medvedev, A. Stuchebrukhov, Cytochrome *c* oxidase upon single electron injection into the enzyme, *Proc. Natl. Acad. Sci.*, 1777 (2008) 1129–1139.
- [72] M. Sezer, A.L. Woelke, E.W. Knapp, R. Schlesinger, M.A. Mroginski, I.M. Weidinger, Redox induced protonation of heme propionates in cytochrome *c* oxidase: Insights from surface enhanced resonance Raman spectroscopy and QM/MM calculations, *Biochim. Biophys. Acta - Bioenerg.* 1858 (2017) 103–108.
- [73] J. Quenneville, D.M. Popović, A.A. Stuchebrukhov, Combined DFT and electrostatics study of the proton pumping mechanism in cytochrome *c* oxidase, *Biochim. Biophys. Acta - Bioenerg.* 1757 (2006) 1035–1046.
- [74] M. Wikström, M.I. Verkhovsky, Mechanism and energetics of proton translocation by the respiratory heme-copper oxidases, *Biochim. Biophys. Acta - Bioenerg.* 1767 (2007) 1200-1214.

- [75] P.E.M. Siegbahn, M.R.A. Blomberg, Energy diagrams and mechanism for proton pumping in cytochrome c oxidase, *1767* (2007) 1143-1156.
- [76] S. Yoshikawa, A. Shimada, Reaction Mechanism of Cytochrome c Oxidase, *Chem. Rev.* 115 (2015) 1936-1989.
- [77] C.Y. Son, A. Yethiraj, Q. Cui, Cavity hydration dynamics in cytochrome *c* oxidase and functional implications, *Proc. Natl. Acad. Sci.*, 114 (2017) E8830-E8836.
- [78] J.E. Morgan, M.I. Verkhovsky, G. Palmer, M. Wikström, Role of the PR intermediate in the reaction of cytochrome c oxidase with O₂, *Biochemistry.* 40 (2001) 6882–6892.
- [79] F.J.A. et al. Klauda JB Venable RM, Update of the CHARMM all-atom additive force field for lipids: Validation on Six lipid types., *Biophys J.* 72 (2010) 2002–2013.
- [80] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, Scalable molecular dynamics with NAMD, *J. Comput. Chem.* 26 (2005) 1781–1802.
- [81] J.-P. Ryckaert, G. Ciccotti, H.J.. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys.* 23 (1977) 327–341.
- [82] K. Blumhagen, I. Muegge, E.W. Knapp, Diffusion of two different water models and thermal conductivity in a protein—water system, *Int. J. Quantum Chem.* 59 (1996) 271–279.
- [83] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, *J. Mol. Graph.* 14 (1996) 33–38.
- [84] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (1983) 926.
- [85] A. Takiden, F. Velazquez-Escobar, J. Dragelj, A.L. Woelke, E.W. Knapp, P. Piwowarski, F. Bart, P. Hildebrandt, M.A. Mroginski, Structural and Vibrational Characterization of the Chromophore Binding Site of Bacterial Phytochrome Agp1, *Photochem. Photobiol.* 93 (2017) 713-723.
- [86] H. Batebi, J. Dragelj, P. Imhof, Role of AP-endonuclease (Ape1) active site residues in stabilization of the reactant enzyme-DNA complex, *86* (2018) 439-453.
- [87] T. Meyer, G. Kieseritzky, E.W. Knapp, Electrostatic pK_a computations in proteins: Role of internal cavities, *Proteins Struct. Funct. Bioinforma.* 79 (2011) 3320–3332.
- [88] P. Goyal, J. Lu, S. Yang, M.R. Gunner, Q. Cui, Changing hydration level in an internal cavity modulates the proton affinity of a key glutamate in cytochrome c oxidase., *Proc. Natl. Acad. Sci.* 110 (2013) 18886–91.
- [89] X. Cai, K. Haider, J. Lu, S. Radic, C.Y. Son, Q. Cui, M.R. Gunner, Network analysis of a proposed exit pathway for protons to the P-side of cytochrome c oxidase, *Biochim. Biophys. Acta - Bioenerg.* 1859 (2018) 997-1005.
- [90] A.L. Johansson, S. Chakrabarty, C.L. Berthold, M. Högbom, A. Warshel, P. Brzezinski, Proton-transport mechanisms in cytochrome c oxidase revealed by studies of kinetic isotope effects, *Biochim. Biophys. Acta - Bioenerg.* 1807 (2011) 1083–1094.

- [91] J.L. Hyun, L. Ojemyr, A. Vakkasoglu, P. Brzezinski, R.B. Gennis, Properties of Arg481 mutants of the aa3-type cytochrome c oxidase from *Rhodobacter sphaeroides* suggest that neither R481 nor the nearby D-propionate of heme a₃ is likely to be the proton loading site of the proton pump, *Biochemistry*. 48 (2009) 7123–7131.
- [92] S. Chakrabarty, I. Namslauer, P. Brzezinski, A. Warshel, Exploration of the cytochrome c oxidase pathway puzzle and examination of the origin of elusive mutational effects, *Biochim. Biophys. Acta - Bioenerg.* 1807 (2011) 413–426.
- [93] R. Liang, J.M.J. Swanson, Y. Peng, M. Wikström, G.A. Voth, Multiscale simulations reveal key features of the proton-pumping mechanism in cytochrome c oxidase, *Proc. Natl. Acad. Sci.* 113 (2016) 7420-7425.
- [94] M.R.A. Blomberg, P.E.M. Siegbahn, *Biochimica et Biophysica Acta Proton pumping in cytochrome c oxidase: Energetic requirements and the role of two proton channels*, *BBA - Bioenerg.* 1837 (2014) 1165–1177.
- [95] M.B. Helabad, T. Ghane, M. Reidelbach, A.L. Woelke, E.W. Knapp, P. Imhof, Protonation-State-Dependent Communication in Cytochrome c Oxidase, *Biophysj.* 113 (2017) 817–828.
- [96] S. Supekar, V.R.I. Kaila, Dewetting transitions coupled to K-channel activation in cytochrome c oxidase, *Chem. Sci.* 9 (2018) 6703–6710.
- [97] K. Kirchberg, H. Michel, U. Alexiev, Net proton uptake is preceded by multiple proton transfer steps upon electron injection into cytochrome c oxidase, *J. Biol. Chem.* 287 (2012) 8187-8193.
- [98] A. Wolf, C. Schneider, T.-Y. Kim, K. Kirchberg, P. Volz, U. Alexiev, A simulation-guided fluorescence correlation spectroscopy tool to investigate the protonation dynamics of cytochrome c oxidase, *Phys. Chem. Chem. Phys.* 18 (2016) 12877–12885.
- [99] M. Brändén, F. Tomson, R.B. Gennis, P. Brzezinski, The entry point of the K-proton-transfer pathway in cytochrome c oxidase, *Biochemistry*. 41 (2002) 10794–10798.
- [100] K. Kirchberg, H. Michel, U. Alexiev, Exploring the entrance of proton pathways in cytochrome c oxidase from *Paracoccus denitrificans*: Surface charge, buffer capacity and redox-dependent polarity changes at the internal surface, *Biochim. Biophys. Acta - Bioenerg.* 1827 (2013) 276-284.
- [101] A.S.F. Oliveira, S.R.R.R. Campos, A.M. Baptista, C.M. Soares, Coupling between protonation and conformation in cytochrome c oxidase: Insights from constant-pH MD simulations, *Biochim. Biophys. Acta - Bioenerg.* 1857 (2016) 759–771.
- [102] P.R. Magalhães, A.S.F. Oliveira, S.R.R.R. Campos, C.M. Soares, A.M. Baptista, M. Soares, A. So, F. Oliveira, S.R.R.R. Campos, Effect of a pH Gradient on the Protonation States of Cytochrome c Oxidase: A Continuum Electrostatics Study, *J. Chem. Inf. Model.* 57 (2017) 256–266.
- [103] H. Ishikita, E.W. Knapp, Redox Potential of Quinones in Both Electron Transfer Branches of Photosystem I, *J. Biol. Chem.* 278 (2003) 52002-52011.
- [104] Schrödinger LLC, Jaguar, Version 7.7, (2010) LLC, New York, NY.

- [105] V. Alterio, R.M. Vitale, S.M. Monti, C. Pedone, A. Scozzafava, A. Cecchi, G. De Simone, C.T. Supuran, Carbonic Anhydrase Inhibitors: X-ray and Molecular Modeling Study for the Interaction of a Fluorescent Antitumor Sulfonamide with Isozyme II and IX, *J. Am. Chem. Soc.* 128 (2006) 8329–8335.
- [106] C.C.I. Bayly, P. Cieplak, W.D. Cornell, P. a Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, *J. Phys.* 97 (1993) 10269–10280.
- [107] W.D. Cornell, P. Cieplak, C.I. Bayly, P. a Kollman, Application of RESP Charges To Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation, *J. Am. Chem. Soc.* 115 (1993) 9620–9631.
- [108] U. Alexiev, T. Marti, gJI P. Maarten Heyn, H. Gobind morana, P. Scherrer, Covalently Bound pH-Indicator Dyes at Selected Extracellular or Cytoplasmic Sites in Bacteriorhodopsin. 2. Rotational Orientation of Helices D and E and Kinetic Correlation between M Formation and Proton Release in Bacteriorhodopsin Micelles?, *Biochemistry.* 33 (1994) 13693–13699.
- [109] M. Möller, U. Alexiev, Surface charge changes upon formation of the signaling state in visual rhodopsin, *Photochem. Photobiol.* 85 (2009) 501–508.
- [110] G.F. Schröder, U. Alexiev, H. Grubmüller, Simulation of Fluorescence Anisotropy Experiments: Probing Protein Dynamics, *Biophys. J.* 89 (2005) 3757–3770.
- [111] Y. Marantz, E. Nachliel, A. Aagaard, P. Brzezinski, M. Gutman, The proton collecting function of the inner surface of cytochrome c oxidase from *Rhodobacter sphaeroides*, *Proc. Natl. Acad. Sci. USA.* 95 (1998) 8590–8595.
- [112] P. Volz, N. Krause, J. Balke, C. Schneider, M. Walter, F. Schneider, R. Schlesinger, U. Alexiev, Light and pH-induced changes in structure and accessibility of transmembrane Helix B and its immediate environment in channelrhodopsin-2, *J. Biol. Chem.* 291 (2016) 17382–17393.
- [113] K. Kirchberg, T.-Y. Kim, M. Moller, D. Skegro, G. Dasara Raju, J. Granzin, G. Buldt, R. Schlesinger, U. Alexiev, Conformational dynamics of helix 8 in the GPCR rhodopsin controls arrestin activation in the desensitization process, *Proc. Natl. Acad. Sci.* 108 (2011) 18690–18695.
- [114] U. Alexiev, D.L. Farrens, Fluorescence spectroscopy of rhodopsins: Insights and approaches, *Biochim. Biophys. Acta - Bioenerg.* 1837 (2014) 694–709.
- [115] L. Buhrow, S. Ferguson-Miller, L.A. Kuhn, From static structure to living protein: Computational analysis of cytochrome c oxidase main-chain flexibility, *Biophys. J.* 102 (2012) 2158–2166.
- [116] M.S. Lee, F.R. Salsbury, C.L. Brooks, Novel generalized Born methods, *J. Chem. Phys.* 116 (2002) 10606.

Appendix A

Supporting information for chapter 2

Table A.1. Computed pK_A values of titratable residues from the benchmark set containing 13 proteins and 194 pK_A values with $KB2^+$. (NTE - N-terminus; CTE - C-terminus)

Proteins														
2LZT			4PTI			3RN3			2RN2			1A2P		
residue	$KB2^+$	Exp.	residue	$KB2^+$	Exp.	residue	$KB2^+$	Exp.	residue	$KB2^+$	Exp.	residue	$KB2^+$	Exp.
LYS1A	9.21	10.80	NTE1A	5.43	7.90	NTE1A	6.45	7.60	GLU6A	3.72	4.50	ASP8A	2.29	3.06
GLU7A	3.26	2.85	ASP3A	3.66	3.55	GLU2A	2.36	2.80	GLU32A	2.25	3.60	ASP12A	1.70	3.65
LYS13A	10.59	10.50	GLU7A	3.96	3.85	GLU9A	4.94	4.00	GLU48A	2.12	4.40	HIS18A	6.57	7.72
HIS15A	5.48	5.36	TYR10A	9.43	9.46	HIS12A	5.64	5.80	GLU57A	2.24	3.20	ASP22A	3.55	3.30
ASP18A	1.98	2.66	LYS15A	10.44	10.43	ASP38A	3.10	3.10	GLU61A	2.24	3.90	GLU29A	3.69	3.75
TYR20A	8.10	10.30	TYR21A	10.35	9.94	GLU49A	4.78	4.70	HIS62A	7.04	7.00	ASP44A	4.29	3.50
TYR23A	10.24	9.80	TYR23A	11.31	11.00	ASP53A	3.49	3.90	GLU64A	4.29	4.40	ASP54A	1.13	2.20
LYS33A	10.69	10.40	LYS26A	10.47	10.44	ASP83A	2.38	3.50	ASP70A	4.91	2.60	GLU60A	1.40	3.20
GLU35A	5.18	6.20	TYR35A	8.59	10.60	GLU86A	4.51	4.10	HIS83A	5.34	5.50	GLU73A	1.28	2.10
ASP48A	1.89	1.60	LYS41A	10.29	10.75	HIS105A	6.61	6.60	ASP94A	3.40	3.20	ASP86A	4.16	4.20
ASP52A	2.56	3.68	LYS46A	10.22	10.35	GLU111A	4.12	3.50	ASP108A	2.25	3.20	ASP101A	1.27	2.00
TYR53A	12.72	12.10	GLU49A	4.63	3.91	HIS119A	6.02	6.10	GLU119A	2.16	4.10	CTE110A	-1.57	3.30
ASP66A	0.99	0.90	ASP50A	2.42	3.20	ASP121A	2.29	3.10	HIS124A	4.91	7.10	ASP8A	2.29	3.06
ASP87A	0.52	2.07	CTE58A	3.48	3.10	CTE124A	0.41	2.40	HIS127A	7.84	7.90			
LYS96A	11.48	10.80							GLU129A	2.32	3.60			
LYS97A	11.31	10.30							GLU131A	5.58	4.30			
ASP101A	4.41	4.08							ASP134A	2.29	4.10			
LYS116A	8.69	10.20							GLU135A	4.69	4.30			
ASP119A	3.05	3.20							GLU147A	4.47	4.20			
CTE129A	2.98	2.75							GLU154A	3.07	4.40			

1HNG			3ICB			1PPF			1ERT			1XNB		
residue	$KB2^+$	Exp.	residue	$KB2^+$	Exp.	residue	$KB2^+$	Exp.	residue	$KB2^+$	Exp.	residue	$KB2^+$	Exp.
ASP2A	3.11	3.50	LYS1A	10.20	10.60	NTE1I	6.41	8.00	GLU6A	3.24	4.80	ASP4A	1.65	3.00
ASP25A	3.72	3.53	GLU4A	4.76	3.80	GLU10I	4.27	4.10	GLU13A	5.49	4.40	ASP11A	2.16	2.50
ASP26A	4.44	3.58	GLU5A	3.45	3.40	TYR11I	9.78	10.20	ASP16A	3.23	4.00	GLU78A	4.20	4.60
ASP28A	3.23	3.57	LYS7A	10.60	11.35	LYS13I	11.89	9.90	ASP20A	3.19	3.80	ASP106A	4.19	2.70
GLU29A	5.30	4.42	GLU11A	5.80	4.70	GLU19I	4.23	3.20	ASP26A	13.61	9.90	ASP119A	3.98	3.20
GLU33A	4.12	4.20	LYS12A	10.28	10.82	TYR20I	10.67	11.10	GLU47A	2.00	4.10	ASP121A	4.40	3.60
GLU41A	6.01	6.53	LYS16A	10.82	10.09	LYS29I	11.38	11.10	GLU56A	3.25	3.30	GLU172A	6.36	6.70
GLU56A	4.34	3.95	GLU17A	3.60	3.62	LYS34I	11.48	10.10	ASP58A	1.35	3.60	ASP4A	1.65	3.00
ASP62A	4.55	4.18	LYS25A	12.31	11.81	GLU43I	4.80	4.80	ASP60A	5.07	3.30			
ASP71A	2.45	3.20	GLU26A	4.65	4.10	HIS52I	6.84	7.50	ASP61A	4.70	4.30			
ASP72A	3.98	4.14	LYS29A	10.91	10.97	LYS55I	10.19	11.10	ASP64A	3.35	3.20			
ASP94A	4.80	3.83	LYS41A	10.55	10.89				GLU68A	3.25	4.90			
CTE99A	3.68	3.11	ASP47A	3.46	3.00				GLU70A	3.25	4.60			
GLU99A	4.93	4.10	GLU48A	3.74	4.60				GLU88A	2.31	3.70			
			LYS55A	7.61	11.38				GLU95A	2.33	4.10			
			GLU64A	4.08	3.80				GLU98A	3.25	3.90			
			LYS71A	9.97	10.72				GLU103A	5.07	4.40			
			LYS72A	9.72	10.96									
			CTE75A	4.02	3.20									

1PGA			2ZTA			3BDC		
residue	KB2 ⁺	Exp.	residue	KB2 ⁺	Exp.	residue	KB2 ⁺	Exp.
LYS10A	11.38	11.00	NTE1A/B	5.85/5.62	7.88	GLU10A	2.78	2.82
LYS13A	11.87	11.00	LYS4A/B	10.47/10.73	10.78	ASP19A	2.27	2.21
GLU15A	3.44	4.40	GLU7A/B	4.81/4.59	4.60	ASP21A	11.76	6.54
GLU19A	3.32	3.70	ASP8A/B	5.20/5.30	3.48	ASP40A	3.25	3.87
ASP22A	3.29	2.90	LYS9A/B	10.68/11.37	11.26	GLU43A	5.74	4.32
GLU27A	2.43	4.50	GLU11A/B	6.49/6.70	3.94	GLU52A	4.87	3.93
LYS28A	11.33	10.90	GLU12A/B	4.30/4.72	4.05	GLU57A	4.46	3.49
TYR33A	10.66	11.00	LYS16A/B	11.21/11.29	10.62	GLU67A	2.79	3.76
ASP36A	5.35	3.80	TYR18A/B	11.15/11.53	9.82	GLU73A	3.66	3.31
ASP40A	4.23	4.00	HIS19A/B	7.11/5.76	6.24	GLU75A	2.65	3.26
GLU42A	4.97	4.40	GLU21A/B	4.62/5.11	4.38	ASP95A	0.49	2.16
ASP46A	3.33	3.60	GLU23A/B	3.84/4.04	4.20	GLU101A	4.09	3.81
ASP47A	3.23	3.40	LYS28A/B	10.86/11.48	11.10	GLU122A	2.85	3.89
CTE56A	4.00	4.00	LYS29A/B	10.99/11.06	10.64	GLU129A	2.19	3.75
GLU56A	2.94	4.00	CTE33A/B	3.65/2.42	4.03	GLU135A	2.67	3.76
			GLU33A/B	5.68/5.10	4.62			

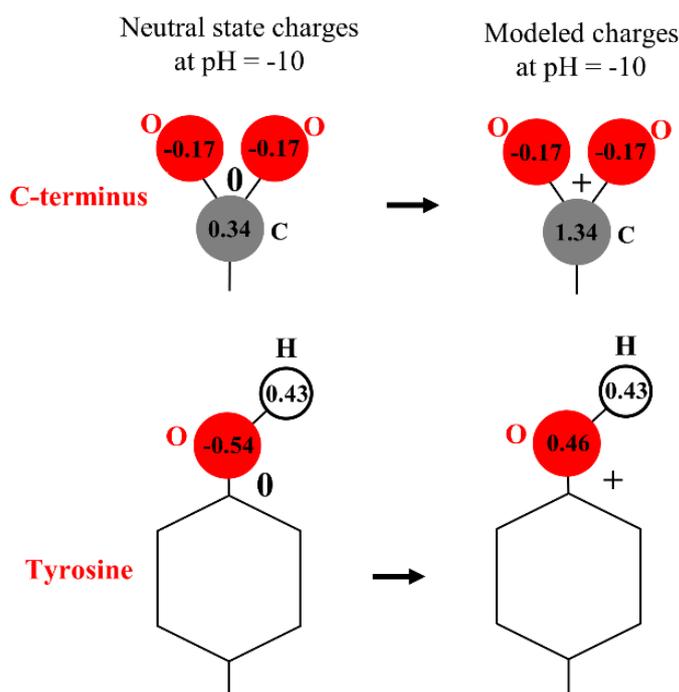


Figure A.1. Introduced changes of atomic partial charges of titratable residues for PAC at pH -10 by using the “repelling charges” modeling procedure for Tyr and C-terminus. Left are the template atomic partial charges of the neutral state and right are the modeled charges after the artificial charge is placed at the carboxylic carbon, resulting in the total positive charge of the residue.

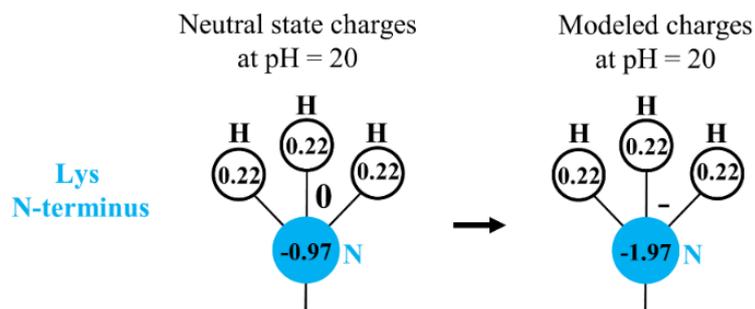


Figure A.2. . Introduced changes of atomic partial charges of titratable residues for PAC at pH 20 by using the “repelling charges” modeling procedure for basic residues Lys and N-terminus. Left are the template atomic partial charges of the neutral state and right are the modeled charges after the artificial charge is placed on the amino nitrogen, resulting in the total negative charge of the residue.

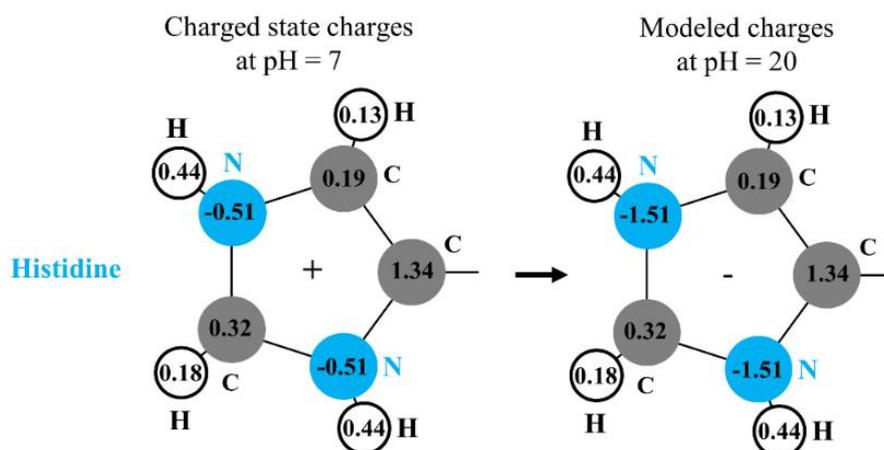


Figure A.3. Introduced changes of atomic partial charges for PAC at pH 20 of His residue by using the “repelling charges” modeling procedure. Left are the template atomic partial charges of the charged state at pH 7 and right are the modeled charges after putting a unit negative charge on the nitrogen atoms of imidazole ring, resulting in a total negative charge of the residue.

Appendix B

Supporting information for chapter 3

Table B.1. Transient “action” pK_A values of PRDa₃ computed with KB2⁺ in the modeled structure of CcO (Figure 3.4, right) with different conformations and protonation states of Glu286.

Glu286 protonation state	Glu286 conformation	CcO state			
		P _M	P _R	P→F	F
Glu286H	up	1.73	2.57	1.33	0.06
Glu286H	down	2.15	3.05	1.74	0.33
Glu286 ⁻	up	3.88	4.84	3.50	2.23
Glu286 ⁻	down	3.13	3.89	2.70	1.30

Table B.2. Results of the evaluation of electrostatic energies of PRDa₃ trajectories with KB2⁺. Bold characters denote the trajectories with the highest Boltzmann factors. Time interval used for evaluation is given in brackets.

PRDa ₃ protonation state	CcO state			
	P _M	P→F	P _R	F
PRD ⁻	1.000	0.973	0.936	0.983
	(1-25ns)	(1-25ns)	(1-25ns)	(1-25ns)
PRDH ¹	0.000	0.000	0.023	0.017
	(1-25ns)	(1-25ns)	(1-25ns)	(1-25ns)
PRDH ²	0.000	0.027	0.041	0.000
	(1-25ns)	(1-25ns)	(1-25ns)	(1-25ns)

Table B.3. Results of the evaluation of electrostatic energies of PRAa₃ trajectories with KB2⁺. Bold characters denote the trajectories with the highest Boltzmann factors. Time interval used for evaluation is given in brackets.

PRAa ₃ protonation state	CcO state			
	P _M	P→F	P _R	F
PRA ⁻	1.000 (1-25ns)	1.000 (1-25ns)	0.999 (1-25ns)	1.000 (1-25ns)
PRAH ¹	0.000 (1-25ns)	0.000 (1-25ns)	0.000 (1-25ns)	0.000 (1-25ns)
PRAH ²	0.000 (1-25ns)	0.000 (1-25ns)	0.001 (1-25ns)	0.000 (1-7ns)*

* Asp407 moves back towards protonated PRAa₃ after 7 ns (Figure B.6)

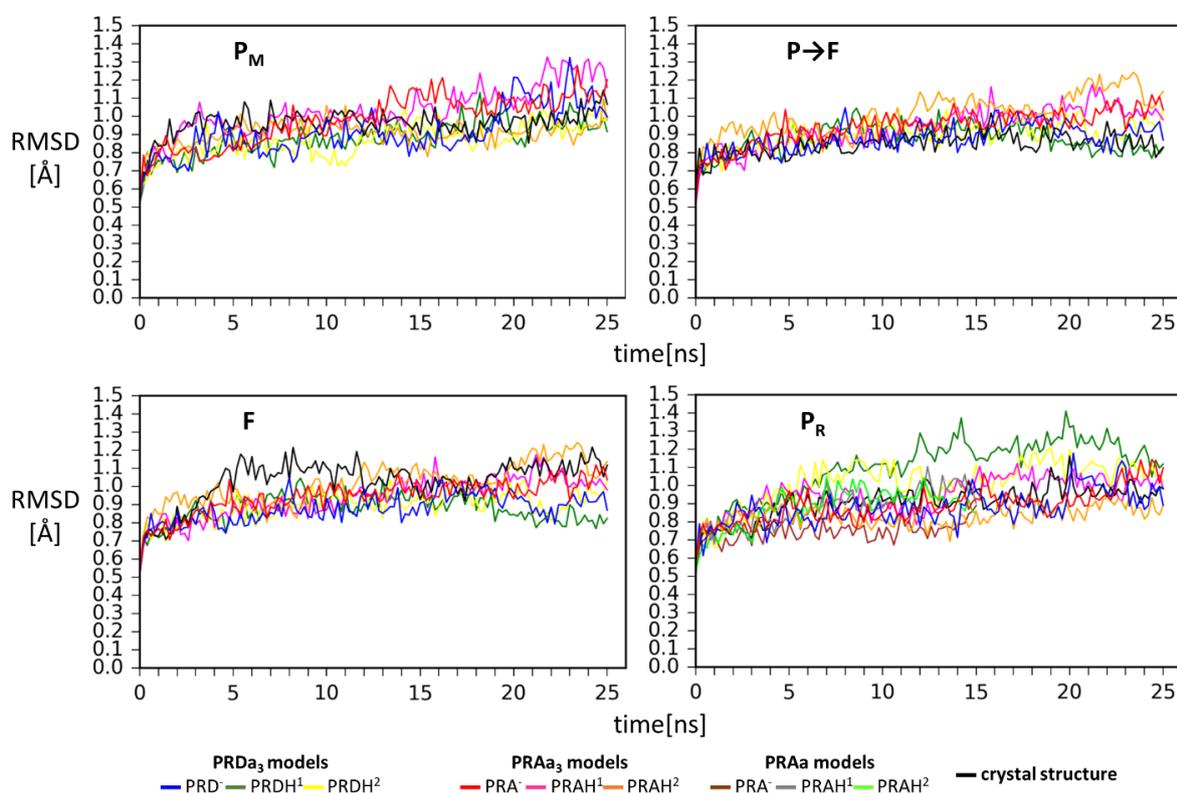


Figure B.1. Root-mean-square deviation (RMSD) of CcO backbone atoms relative to the crystal structure (PDB 2GSM[28]) obtained by MD simulations.

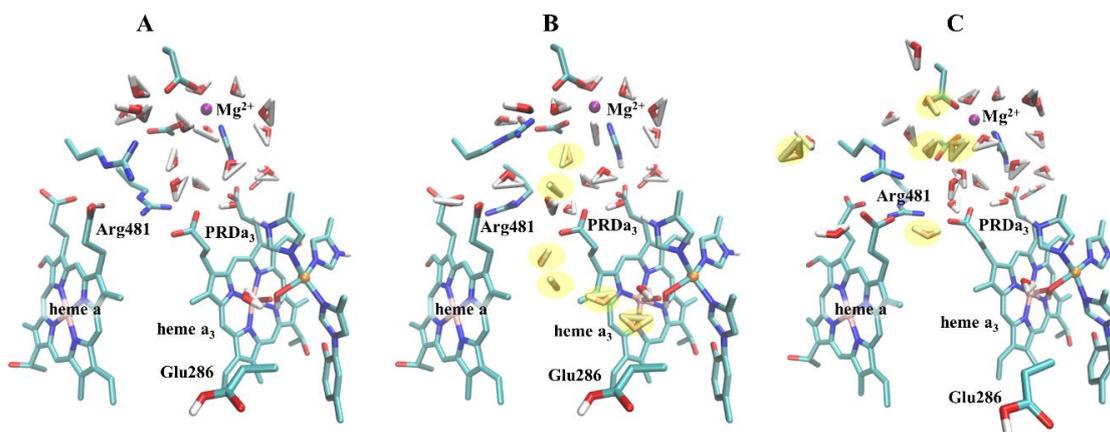


Figure B.2. Water molecules within 10\AA of PRDa₃: **A)** in the crystal structure (PDB: 2GSM[28]); **B)** in the modeled structure with an open salt-bridge between Arg481 and deprotonated PRDa₃ with six additional water molecules. Four water molecules are modeled in the cavity between Glu286 and PRDa₃, similarly to the work [9] and two water molecules are added to the hydrophobic cavity between heme a, heme a₃ and Mg²⁺ complex. Modeled water molecules are marked in yellow; **C)** after 10ns of the MD simulation starting from the structure in B), where all added water molecules leave the hydrophobic cavity between Glu286 and PRDa₃ after which five out of six water molecules reside in the cavity between heme a, heme a₃ and Mg²⁺ complex.

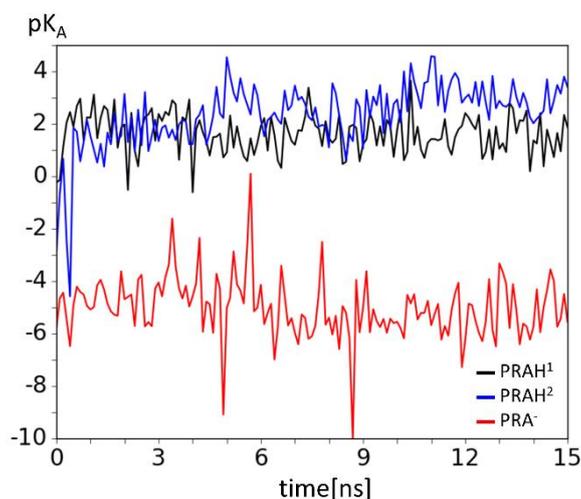


Figure B.3. Time course of computed pK_A values of PRAa from MD simulations of modeled structures with different PRAa protonation states and Boltzmann factors (in brackets) used for evaluation with KB²⁺: PRA⁻ (0.933), PRAH¹ (0.004), PRAH² (0.063). Time interval used for evaluation was 1-15ns.

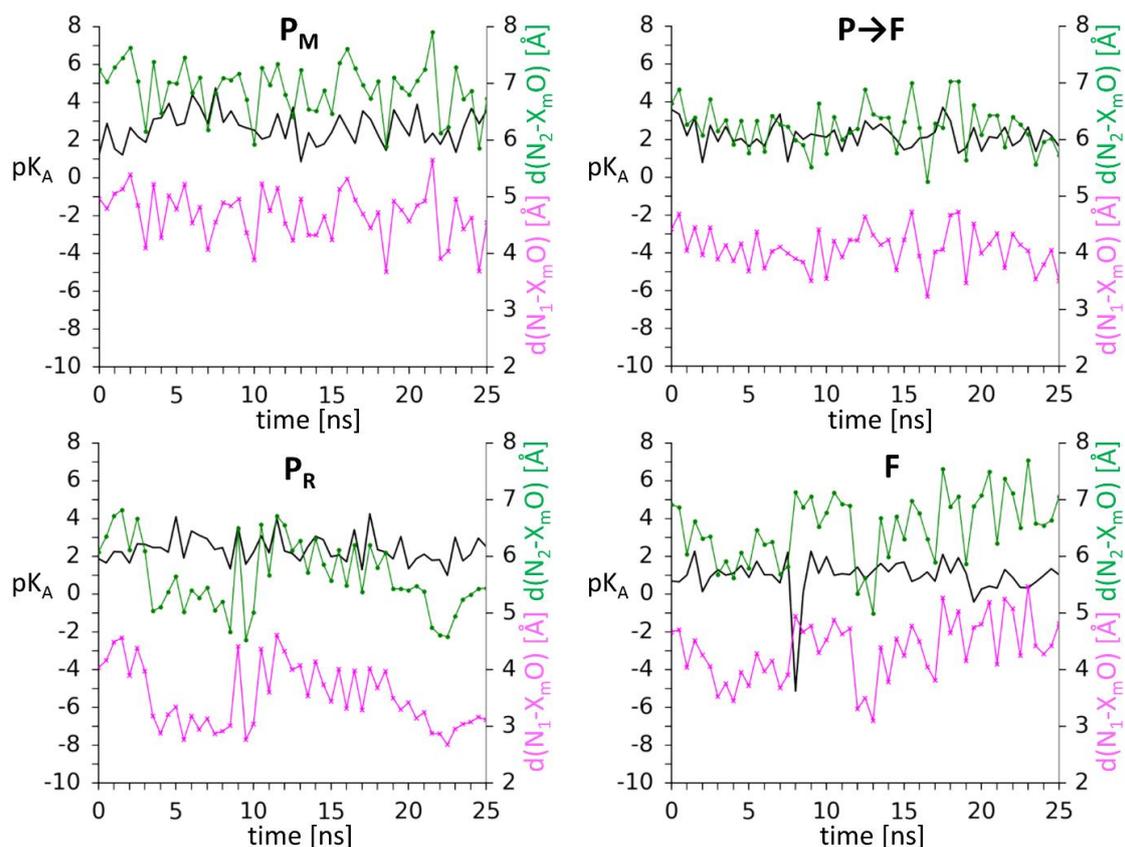


Figure B.4. The conformation of the Arg481-PRD_{a3}, featured by the distance of the nitrogen atoms of Arg481 (N₁-magenta, N₂-green) from the center of mass of the oxygens atoms of propionic acid sidechain over time with deprotonated PRD_{a3} and the pK_A value (black line) of PRD_{a3}. Data are presented for simulations of PRD⁻ for four redox states of CcO (Table 3.1).

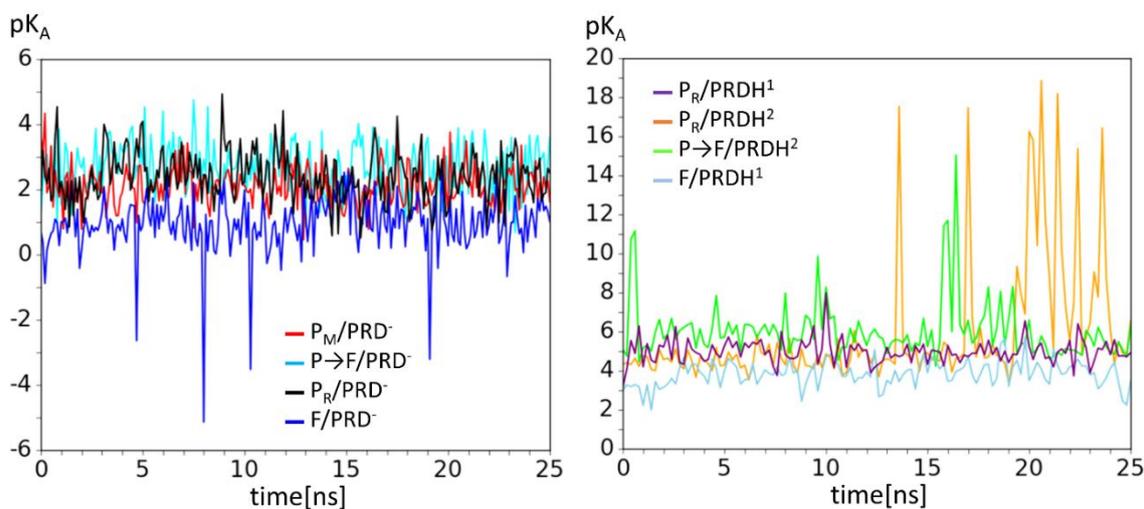


Figure B.5. Time course of computed pK_A values (with KB₂⁺) of PRD_{a3} from MD simulations of modeled structures: **Left:** Trajectories with the highest Boltzmann factors from Table B.2; **Right:** All other trajectories with nonzero Boltzmann factors from Table B.2.

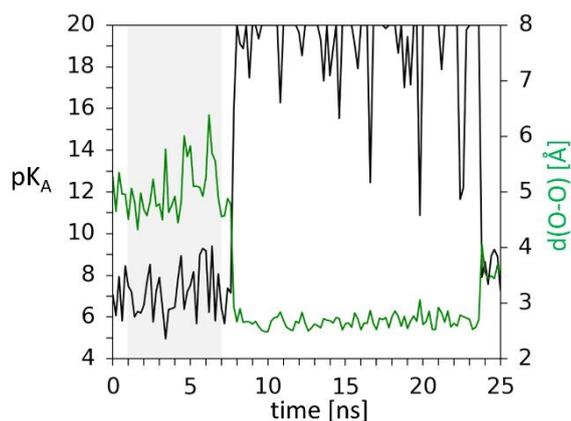


Figure B.6. Asp407 returns to the conformation to form a hydrogen bond with protonated PRAa₃ in the simulation with PRAH², represented by the closest distance of the oxygen atoms of Asp407 and PRAa₃ over time and the pK_A value of PRAa₃.

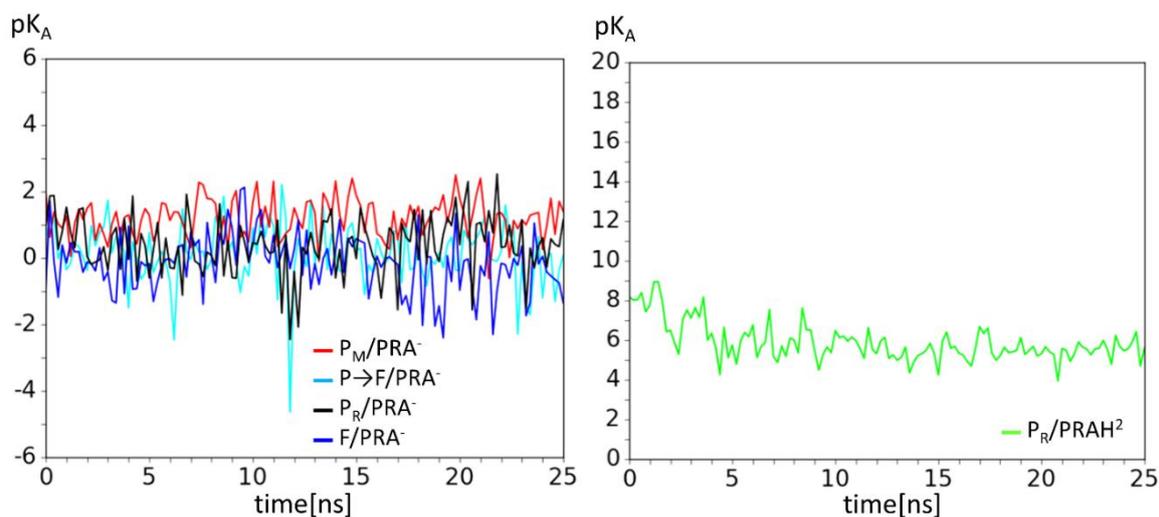


Figure B.7. Time course of computed pK_A values (with KB²⁺) of PRAa₃ from MD simulations of modeled structures: **Left:** Trajectories with the highest Boltzmann factors from Table B.3; **Right:** All other trajectories with nonzero Boltzmann factors from Table B.3.

Appendix C

Supporting information for chapter 4

Table C.1. Cartesian coordinates and charges of fluorescein from quantum chemical computations.

a) Optimized Cartesian coordinates and atomic partial charges for protonated fluorescein

	x	y	z	charge
C1	-3.8911926942	5.1939921606	13.5347723934	-0.250422
S2	-2.0812338340	5.1880656417	13.5454401274	-0.208491
C3	-1.5164171018	6.4061142484	12.8766055807	-0.310327
C4	-1.6710000000	7.8110000000	13.3790000000	0.744230
O5	-1.7239843944	8.0697289910	14.5095631177	-0.534312
N6	-1.7370000000	8.7430000000	12.3950000000	-0.612792
C7	-1.8810000000	10.1570000000	12.3350000000	0.350032
C8	-1.8400000000	10.9270000000	13.5150000000	-0.248783
C9	-1.9750000000	12.3220000000	13.4600000000	-0.130996
C10	-2.1570000000	12.9850000000	12.2010000000	-0.142257
C11	-2.2040000000	12.1930000000	10.9850000000	0.146226
C12	-2.0630000000	10.7960000000	11.0930000000	-0.356887
C13	-2.4000000000	12.7950000000	9.5900000000	0.550206
O14	-1.6190000000	13.7140000000	9.2340000000	-0.557834
O15	-3.3300000000	12.3500000000	8.8780000000	-0.637596
C16	-2.2900000000	14.5180000000	12.2490000000	0.276000
C17	0.2270000000	14.9040000000	12.7600000000	-0.030130
C18	1.2670000000	15.7830000000	13.0060000000	-0.386559
C19	1.0700000000	17.2170000000	13.0330000000	0.686744
C20	-0.2630000000	17.7010000000	12.7900000000	-0.577137
C21	-4.8660000000	17.2590000000	11.8590000000	-0.432467
C22	-6.0720000000	16.5200000000	11.5890000000	0.451724
C23	-6.0000000000	15.0520000000	11.5430000000	-0.432367
C24	-4.8090000000	14.4030000000	11.7520000000	0.008512
C25	-3.5590000000	15.1320000000	12.0250000000	-0.164640
C26	-3.6700000000	16.5820000000	12.0690000000	0.376525
C27	-1.1340000000	15.3640000000	12.5110000000	-0.198662
C28	-1.3030000000	16.8080000000	12.5460000000	0.434394
O29	-2.5510000000	17.3670000000	12.3210000000	-0.312221
O30	2.0890000000	18.0230000000	13.2700000000	-0.597389
O31	-7.1920000000	17.1060000000	11.3890000000	-0.546528
H32	-4.2609029087	4.2952745882	14.0284535480	0.137041
H33	-4.2552866906	6.0745557496	14.0640580700	0.123096
H34	-4.2489121032	5.2157202226	12.5053722206	0.093145
H35	-1.6782315341	8.4604963916	11.4270960902	0.306079
H36	-1.7037733155	10.4433044318	14.4709711752	0.171083
H37	-1.9422225261	12.9039590230	14.3692025807	0.090099
H38	-2.0979428968	10.2119512192	10.1852191809	0.154535
H39	0.4107439576	13.8398103065	12.7482333445	0.124614
H40	2.2626140416	15.4048975292	13.1854190674	0.131070
H41	-0.4615449572	18.7625751086	12.7961797037	0.174613

H42	-4.8911963534	18.3379943186	11.8981969920	0.193067
H43	-6.8899677751	14.4736975670	11.3431909529	0.157112
H44	-4.7687841390	13.3242581835	11.7188521815	0.124399
H45	-7.8729757241	16.4490203046	11.2269822755	0.386418
H46	-0.4314787067	6.3189907470	12.8181404258	0.153334
H47	-1.9281360003	6.4929767439	11.8710995114	0.124499

b) Optimized Cartesian coordinates and charges for deprotonated fluorescein

	x	y	z	charge
C1	-3.6830789064	5.4308113363	14.2326187456	-0.227424
S2	-1.9336703088	5.1203090229	13.8133697528	-0.247959
C3	-1.6458270526	6.4435252801	12.5991755581	-0.264389
C4	-1.8478711713	7.8844363383	13.1213825826	0.714179
O5	-1.9686388126	8.1551690752	14.3083992422	-0.543706
N6	-1.8671558618	8.7952813590	12.1025134597	-0.575418
C7	-1.9733451064	10.2107519398	12.1472256067	0.289919
C8	-2.0356892289	10.9470693111	13.3380657251	-0.225174
C9	-2.1191204330	12.3374534246	13.2610829630	-0.179265
C10	-2.1462893770	13.0123203201	12.0308442104	-0.028034
C11	-2.0852819965	12.2618276108	10.8464098371	0.101522
C12	-2.0011128058	10.8760103774	10.9188324980	-0.335922
C13	-2.0953838752	12.9592419424	9.4726214649	0.581215
O14	-2.1547171247	14.2191944227	9.5516627938	-0.607236
O15	-2.0532114286	12.2151426670	8.4710022323	-0.690407
C16	-2.2575433356	14.5049913424	12.0203047611	0.262089
C17	0.1392430338	14.8057889994	12.6516168928	-0.020489
C18	1.1542283192	15.6142012148	13.0458693199	-0.424395
C19	0.9706391729	17.0546712180	13.2773795951	0.662046
C20	-0.3746919470	17.5523372071	13.0378274961	-0.580822
C21	-4.9076022351	17.1811094097	12.0006323977	-0.578192
C22	-6.0424866048	16.4535797064	11.6612558239	0.661111
C23	-5.9584571965	15.0702256224	11.4396941431	-0.422589
C24	-4.7354956379	14.4296261434	11.5525677344	-0.028658
C25	-3.5642570900	15.1290998525	11.8920441229	-0.221702
C26	-3.6869598924	16.5156705716	12.1104002907	0.403234
C27	-1.1964756294	15.3027172497	12.4317134268	-0.231495
C28	-1.3768889291	16.7201352452	12.6374416351	0.408751
O29	-2.6200620492	17.2861431274	12.4536326892	-0.306219
O30	1.9042203513	17.7846673038	13.6551429516	-0.664917
O31	-7.2277481695	17.1350449206	11.5624668890	-0.665117
H32	-3.9731454758	4.6623980429	14.9526623428	0.121223
H33	-3.7830763883	6.4164288358	14.6883255414	0.125311
H34	-4.3239320546	5.3474416203	13.3499399817	0.086855
H35	-1.7659364286	8.4282198426	11.1660772461	0.290717
H36	-2.0147859397	10.4380740788	14.2913965317	0.155617
H37	-2.1617337818	12.9113260812	14.1832104709	0.088525
H38	-1.9488878403	10.3502803532	9.9666158954	0.145207
H39	0.3126236879	13.7486811390	12.4790313779	0.096859
H40	2.1559022141	15.2244458389	13.2033126068	0.112787
H41	-0.5698281174	18.6079978920	13.1906396787	0.149710
H42	-4.9618547006	18.2491381396	12.1744517210	0.149355

H43	-6.8497592630	14.5073510545	11.1705215192	0.112762
H44	-4.6655220337	13.3650342832	11.3646598063	0.101650
H46	-0.6015715649	6.3244839314	12.2883956065	0.136471
H47	-2.2651377986	6.2769013819	11.7094634612	0.112417

Preliminary work – short term MD simulations of flu-CcO

Due to the large number of different protonation combinations of investigated residues, optimization was required for the usage of computer resources. To identify the most probable protonation states of His73, His526 and fluorescein, short MD simulations with variation of protonation states of these residues were performed. Since His73 is more than 15Å from the position 299, it is assumed to be in the deprotonated state (δ -tautomer) found in the crystal structure for both O and R state of CcO. Therefore, only the protonation states of fluorescein and His526 were varied, since they are very close (within 5Å, closest atom pair distance). With the three protonation states of fluorescein, three protonation states of His526 and two redox states of CcO, a set of 18 different charge patterns of flu-CcO were assembled. In addition, four orientations of fluorescein were modeled for each charge pattern, resulting in a total of 72 MD simulations. These preliminary trajectories were at least 5ns long, with equilibration time of 2.2ns. Only the last 3ns were used for analysis of geometric and energetic quantities.

Preliminary work – pre-selection of trajectories

In short term trajectories of 5ns, changes of fluorescein orientation are small. The initial orientation and protonation have impact on the fluorescein movement. To explore this in detail, four orientations of fluorescein were modeled according to Figure 4.3. As explained above, 72 MD simulations of 5 ns length were performed, varying protonation states of His526 and fluorescein, redox states of CcO and fluorescein orientations.

The trajectories were then pre-selected, in terms of total electrostatic energy. Total electrostatic energies were evaluated with generalized Born (in CHARMM using gbmv[116]) and averaged for the time frames of the last 3ns of each trajectory. In Table C.2 below, the complete set of 72 trajectories and electrostatic energies is given. Only trajectories with the same number of atoms and for the same oxidation state of CcO were compared. The difference in the number of atoms comes from different protonation patterns, which is presented as a number of excess protons in Table C.2. For example, the trajectory His ^{δ} 73/His^e526/Flu has 0 excess protons, His ^{δ} 73/His ^{δ} 526/Flu(u) has one excess proton because fluorescein is protonated and

His⁺526/Flu(d) has 2 excess protons because both residues are protonated. In a further step of pre-selection, trajectories were selected if their total energy is less than 10 kcal/mol higher than the energy of the trajectory with the lowest energy (the latter energy was set to 0 in Table C.2). After the first step and the evaluation of electrostatic energies with Karlsberg2⁺ (KB2⁺MD)[7], these selected trajectories were extended to a total length of 20ns. Finally, four trajectories were selected for oxidized and three for the reduced state of CcO (bolded in Table C.2). Only one trajectory for each state was relevant for the pK_A computations, since the other trajectories with very low Boltzmann factors did not affect the final pK_A result of investigated residues (Table C.3). The two trajectories [flu₀₀₀(O) and flu₀₀₀(R)] were elongated to 32 ns and were evaluated together with trajectories described in the main work (Table C.5).

Table C.2. Time averages of the total electrostatic energy relative to the lowest energy average for protonation patterns of flu-CcO (the third column). The energies were computed with the gbmV method [15,116] as implemented in CHARMM[21]. In brackets are the absolute energies.

state of CcO	No. of excess H	protonation pattern (protonation side, starting position)	total electrostatic energy [kcal/mol]			
			Starting orientation			
			0°	90°	180°	270°
O	0	His ^δ 73/His ^δ 526/Flu	5.12 (-12323.63)	39.11 (-12289.64)	49.70 (-12279.05)	75.82 (-12252.93)
		His ^δ 73/His ^ε 526/Flu	62.35 (-12266.40)	46.49 (-12282.26)	0.00 (-12328.75)	64.21 (-12264.54)
	1	His ^δ 73/His ^δ 526/Flu(d)	157.54 (-12174.45)	23.56 (-12308.43)	110.40 (-12221.59)	92.67 (-12239.32)
		His ^δ 73/His ^δ 526/Flu(u)	87.60 (-12244.39)	40.59 (-12291.40)	54.81 (-12277.18)	140.50 (-12191.49)
		His ^δ 73/His ^ε 526/Flu(d)	30.77 (-12301.22)	64.82 (-12267.17)	11.19 (-12320.80)	45.36 (-12286.63)
		His ^δ 73/His ^ε 526/Flu(u)	145.71 (-12186.28)	98.96 (-12233.03)	59.26 (-12272.73)	0.00 (-12331.99)
		His ^δ 73/His ⁺ 526/Flu	86.55 (-12245.44)	61.43 (-12270.56)	41.15 (-12290.84)	71.13 (-12260.86)
		2	His ^δ 73/His ⁺ 526/Flu(d)	0.00 (-12352.76)	55.98 (-12296.78)	106.81 (-12245.95)
	His ^δ 73/His ⁺ 526/Flu(u)		42.65 (-12310.11)	108.89 (-12243.87)	47.63 (-12305.13)	106.40 (-12246.36)
	R	0	His ^δ 73/His ^δ 526/Flu	72.93 (-12288.77)	106.86 (-12254.84)	74.21 (-12287.49)
His ^δ 73/His ^ε 526/Flu			135.96 (-12225.74)	93.72 (-12267.98)	0.00 (-12361.70)	82.31 (-12279.39)
1		His ^δ 73/His ^δ 526/Flu(d)	98.06 (-12243.11)	152.19 (-12188.98)	88.92 (-12252.25)	46.62 (-12294.55)
		His ^δ 73/His ^δ 526/Flu(u)	67.55 (-12273.62)	64.06 (-12277.11)	89.28 (-12251.89)	53.91 (-12287.26)
		His ^δ 73/His ^ε 526/Flu(d)	77.30 (-12263.87)	82.71 (-12258.46)	76.08 (-12265.09)	101.20 (-12239.97)
		His ^δ 73/His ^ε 526/Flu(u)	100.02 (-12241.15)	99.69 (-12241.48)	52.86 (-12288.31)	137.82 (-12203.35)
		His ^δ 73/His ⁺ 526/Flu	39.56 (-12301.61)	73.37 (-12267.80)	42.42 (-12298.75)	0.00 (-12341.17)
		2	His ^δ 73/His ⁺ 526/Flu(d)	272.88 (-12092.74)	58.97 (-12306.65)	65.79 (-12299.83)
His ^δ 73/His ⁺ 526/Flu(u)			48.49 (-12317.13)	98.82 (-12266.80)	102.48 (-12263.14)	0.00 (-12365.62)

Table C.3. Results of the evaluation of electrostatic energies of pre-selected trajectories with KB2⁺. Bold characters denote the dominant trajectories of flu-CcO from the set from Table C.2 used for the analysis, as explained above.

state of CcO	protonation pattern (protonation side, starting position)	occupancy ^a for short term trajectory	long term time interval	occupancy ^a for long term trajectory
O	His ^{δ73} /His ^{ε526} /Flu (180°)	0.76	5-20ns	0.96
	His ^{δ73} /His ^{ε526} /Flu(u, 270°)	0.16	5-20ns	0.01
	His ^{δ73} /His ⁺⁵²⁶ /Flu(d, 0°)	0.08	5-20ns	0.04
R	His ^{δ73} /His ^{ε526} /Flu (180°)	0.44	5-20ns	0.98
	His ^{δ73} /His ⁺⁵²⁶ /Flu (270°)	0.50	5-20ns	0.02
	His ^{δ73} /His ⁺⁵²⁶ /Flu(u, 270°)	0.06	5-20ns	0.00

^aThe occupancy is evaluated by the normalized Boltzmann factor using the total electrostatic energy.

Table C.4. Results of the evaluation of electrostatic energies of trajectories of wt-CcO with Karlsberg²⁺. Bold characters denote the dominant trajectories used for the further analysis.

State of CcO	Protonation pattern	wild type		
		label	Time used	Occupancy
O	His ^{δ73} /His ^{ε526}	wt ₀₀ (O)	5-20ns	0.00
	His ^{δ73} /His ⁺⁵²⁶	wt ₀₁ (O)	5-20ns	0.00
	His ⁺⁷³ /His ^{ε526}	W ₁₀ (O)	5-20ns	0.05
	His ⁺⁷³ /His ⁺⁵²⁶	W₁₁(O)	5-20ns	0.95
R	His ^{δ73} /His ^{ε526}	wt ₀₀ (R)	5-20ns	0.00
	His ^{δ73} /His ⁺⁵²⁶	wt ₀₁ (R)	5-20ns	0.02
	His ⁺⁷³ /His ^{ε526}	W ₁₀ (R)	5-20ns	0.00
	His ⁺⁷³ /His ⁺⁵²⁶	W₁₁(R)	5-20ns	0.98

Table C.5. Results of the final evaluation of electrostatic energies of long term trajectories with KB2⁺. Bold characters denote the dominant trajectories of flu-CcO.

state of CcO	Protonation pattern (protonation side, starting position)	label	Time used	Occupancy
O	His ^{δ73} /His ^{ε526} /Flu	flu ₀₀₀ (O)	17-32ns	0.01
	His ⁺⁷³ /His ^{ε526} /Flu	flu ₁₀₀ (O)	17-32ns	0.02
	His⁺⁷³/His⁺⁵²⁶/Flu	flu₁₁₀(O)	17-32ns	0.90
	His ⁺⁷³ /His ⁺⁵²⁶ /Flu(u)	flu _{11u} (O)	17-32ns	0.07
R	His ^{δ73} /His ^{ε526} /Flu	Flu ₀₀₀ (R)	17-32ns	0.12
	His ⁺⁷³ /His ⁺⁵²⁶ /Flu(u)	Flu _{11u} (R)	17-32ns	0.06
	His⁺⁷³/His⁺⁵²⁶/Flu(d)	flu_{11d}(R)	17-32ns	0.82

Table C.6. Distances of the histidine residues His73 with the potential glutamic acid salt-bridge partners Glu78 (and Glu101) in various CcO crystal structures.

residue pair	O-N distance [Å]	Species	PDB code	resolution [Å]
His73-Glu78	7.66	<i>Paracoccus denitrificans</i>	3HB3[31]	2.25
His73-Glu78	6.86	<i>Paracoccus denitrificans</i>	1AR1[30]	2.70
His96-Glu101	3.97	<i>Rhodobacter spheroides</i> (oxidized)	2GSM[28]	2.00
His96-Glu101	3.98	<i>Rhodobacter spheroides</i> (reduced)	3FYE[34]	2.15
His96-Glu101	7.16	<i>Rhodobacter spheroides</i>	1M56[35]	2.30
His52-His328	5.60 (N-N)	Bovine	1V55[33]	1.90

Evaluating the fluorescein pK_A shift

The system flu-CcO may need longer equilibration time than wt-CcO, therefore energetic quantities (pK_A computations for example) need to be evaluated in an appropriate time interval. The pK_A shift was evaluated for different time intervals of the trajectories from Table 4.2, as shown in Table C.7. The positive pK_A shift is in agreement with experiment only if later parts of trajectories are used. The last 15 ns of trajectories have been used to compute the pK_A value of fluorescein and the time before was considered as equilibration for energetic quantities.

Table C.7. Fluorescein pK_A shift dependence on the time interval used for evaluation the trajectories are taken from Table 4.2.

time interval	2-12ns	7-17ns	12-17ns	12-22ns	17-22ns	17-27ns	22-32ns
pK _A shift of fluorescein	-0.08	-0.39	-0.16	0.15	0.53	0.44	0.76

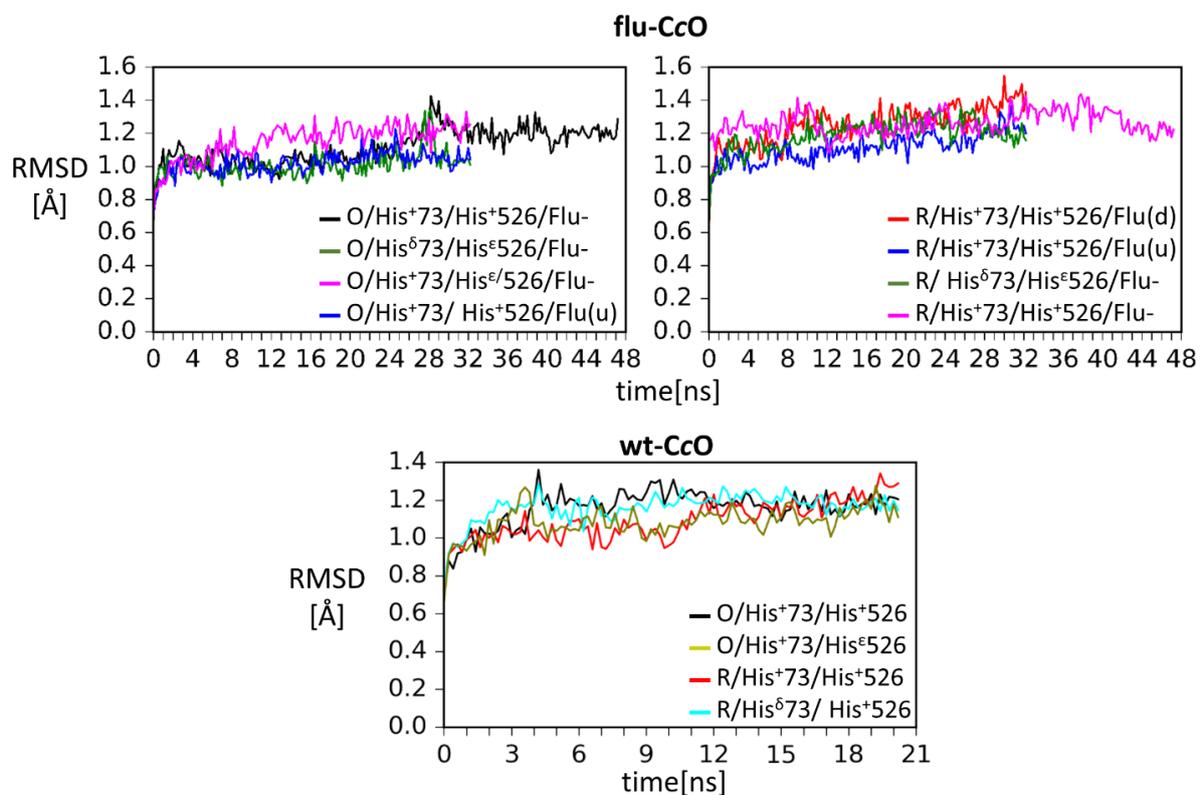


Figure C.1. Root-mean-square deviation (RMSD) of CcO backbone atoms relative to the crystal structure PDB code 3HB3[31] obtained by MD simulations. The RMSD was calculated after excluding mobile loop close to fluorescein (residues 517 to 529 in chain A) and a mobile terminal loop (535 to 545 in chain A).

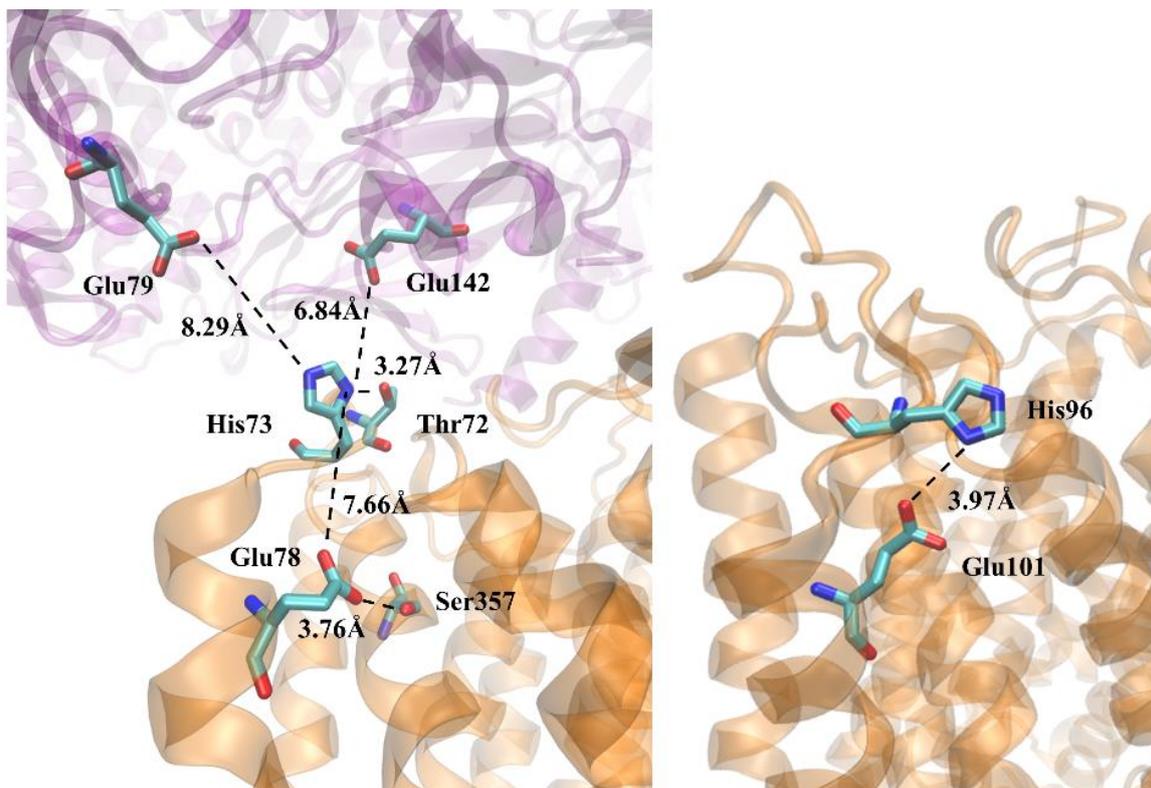


Figure C.2. Left: Crystal structure of two CcO monomers (gold and purple) from *Paracoccus Denitrificans* (PDB code 3HB3[30,31]) showing crystal contacts of the His73 in the “up” conformation. **Right:** Crystal structure of CcO from *Rhodobacter Sphaeroides* (PDB code: 2GSM[28]) His96 in the “down” conformation.

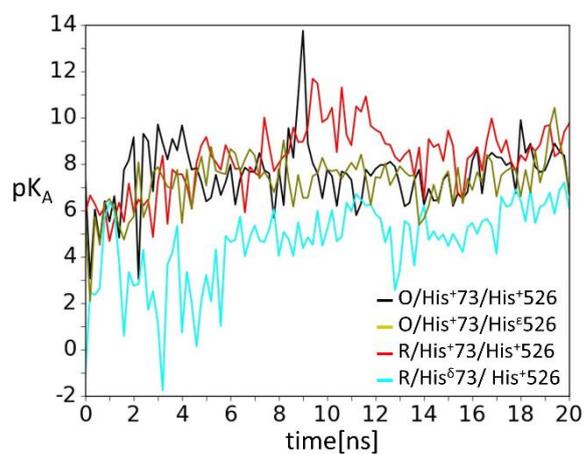


Figure C.3. Time course of computed pK_A values of His73 from MD simulations for all trajectories of wt-CcO as listed in Table 4.2.

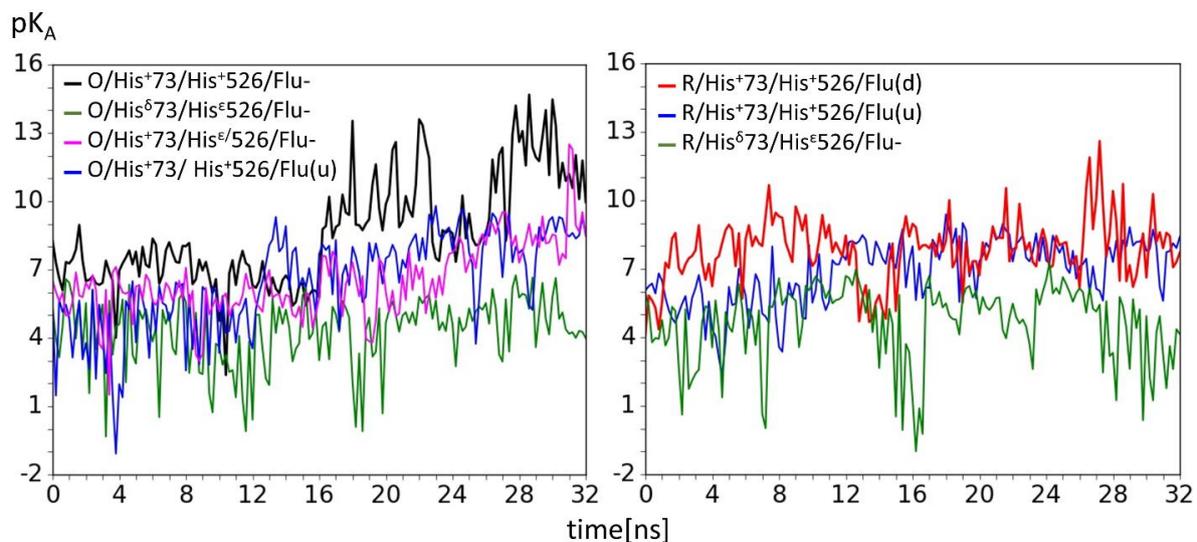


Figure C.4. Time course of computed pK_A values of His73 from MD simulations for all trajectories of: **Left:** oxidized flu-CcO and **Right:** reduced flu-CcO as listed in Table 4.2.

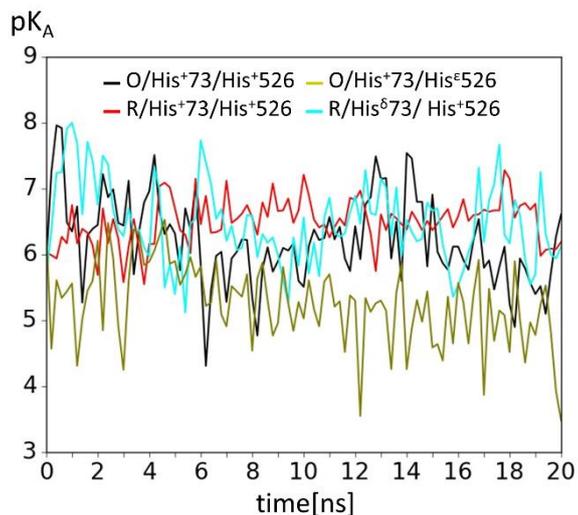


Figure C.5. Time course of computed pK_A values of His526 from MD simulations of wt-CcO as listed in Table 4.2.

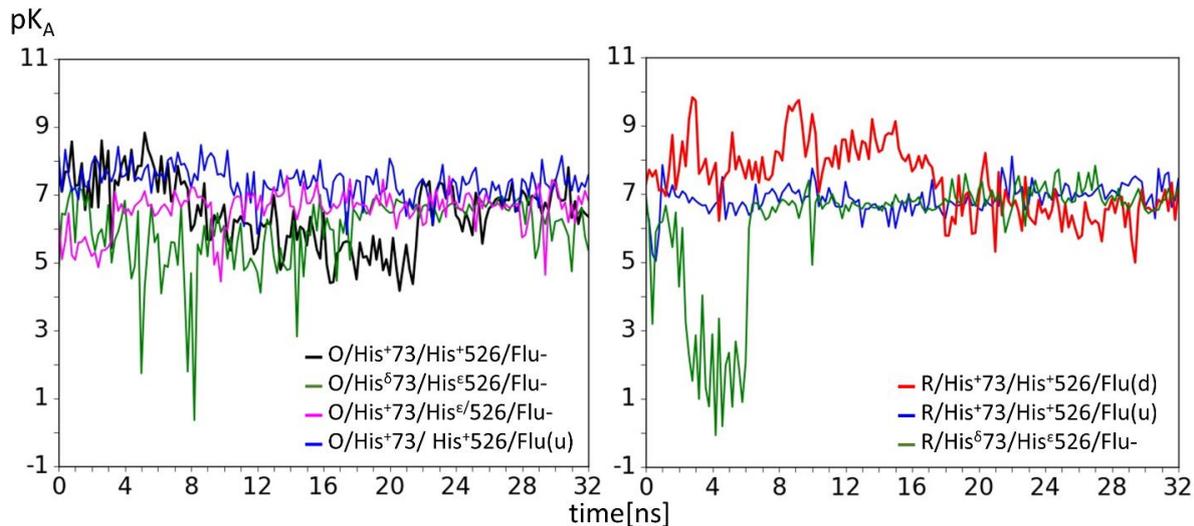


Figure C.6. Time course of computed pK_A values of His526 from MD simulations for all trajectories of: **Left:** oxidized flu-CcO and **Right:** reduced flu-CcO as listed in Table 4.2.

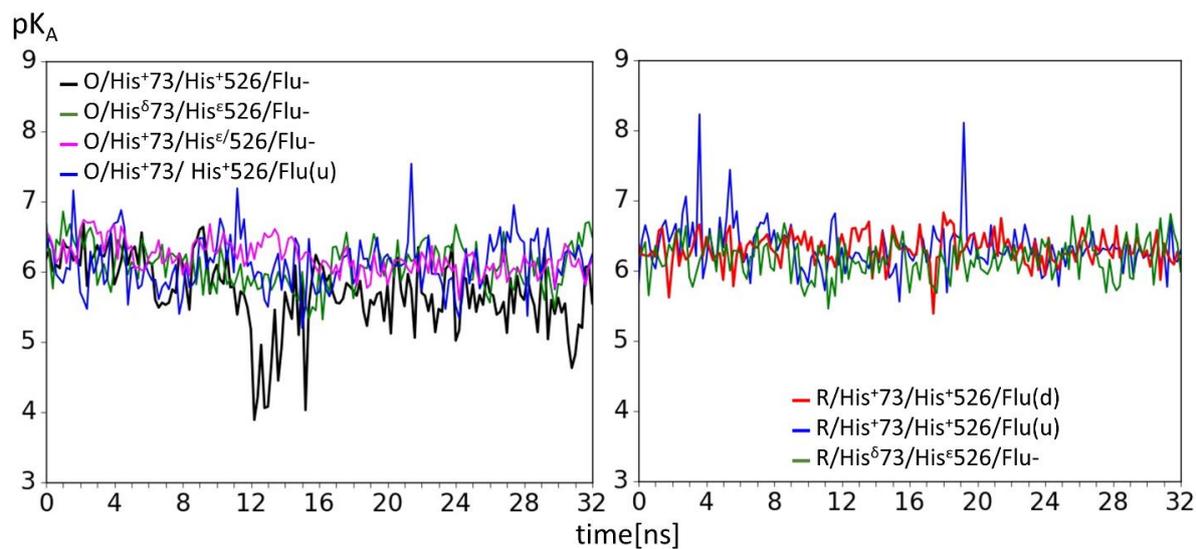


Figure C.7. Time course of computed pK_A values of fluorescein from MD simulations for all trajectories of: **Left:** oxidized flu-CcO and **Right:** reduced flu-CcO as listed in Table 4.2.

CURRICULUM VITAE

JOVAN DRAGELJ

For reasons of data protection, the curriculum vitae is not published in the electronic version.