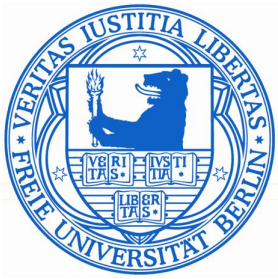
Towards Accurate Computations of Cofactor-Containing Biosystems

Suliman Adam

2017

Im Fachbereich Physik der Freien Universität Berlin eingereichte Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (*doctor rerum naturalium*)



1. Gutachter: Prof. Dr. Ana-Nicoleta Bondar
2. Gutachter: Prof. Dr. Roland Netz

Tag der Disputation: 07.05.2018

If a person who indulges in gluttony is a glutton,
and a person who commits a felony is a felon,
then God is an iron.

Spider Robinson

To those who *know*
how to pronounce iron

Abstract

In the cell, cofactor proteins take on a wide range of tasks. Photosystem II, for example, generates oxygen during photosynthesis and channelrhodopsin from the alga *Chlamydomonas reinhardtii* makes the organism move in response to light. For these proteins to function properly, the presence of their cofactor molecules is indispensable.

Theoretical attempts to studying cofactor-containing systems like photosystem II might struggle with the lack of reliable force field parameters. For a retinal cofactor, on the other hand, quantum mechanical description of its electrostatic interactions is necessary to correctly characterize its behaviour inside of the protein. Here, we used two different theoretical approaches to study cofactors of two proteins: classical mechanics and force field optimization to describe the iron-containing cofactors of photosystem II, and combined quantum and classical mechanics to analyse the environmental effects on proton transfer energetics in channelrhodopsin.

We successfully derived new CHARMM force field parameters for the hæm and non-hæm iron complexes of photosystem II. By comparing to quantum mechanical data and by using test simulations, we showed that these new parameters provide a greatly improved description of intramolecular and non-bonded interactions. The parameters presented here will facilitate reliable all-atom simulations of proteins that contain hæm and non-hæm iron complexes.

In channelrhodopsin, deprotonation of the Schiff base only occurs after the light-induced isomerization of the retinal. We used the Weighted Histogram Approach Method to generate proton transfer pathways to understand what prevents retinal deprotonation in the dark state of the protein and identified three important determinants that ensure the chromophore's continued protonation. These determinants were the presence of a positively charged lysine (K132), water molecules in the active site and the protonation state of the Schiff base counterion glutamate-162. Ultimately, we demonstrated that a combined quantum and classical mechanical approach can be applied to proteins embedded in a hydrated lipid membrane, and we created a protocol that can be transferred to other proteins to study their proton transfer pathways and energetics while taking the protein and lipid environment into account.

Zusammenfassung

Kofaktorproteine übernehmen innerhalb der Zelle eine Vielfalt von Aufgaben. So generiert Photosystem II Sauerstoff im Rahmen der Photosynthese und Kanalrhodopsin kontrolliert phototaxische Reaktionen der Alge *Chlamydomonas reinhardtii*. Um die korrekte Funktionsweise solcher Proteine zu garantieren, ist das Vorhandensein von bestimmten Kofaktormolekülen unabdingbar.

Versuche, Kofaktorsysteme wie Photosystem II zu untersuchen, können daran scheitern, dass keine geeigneten Kraftfeldparameter zur Verfügung stehen. Für Retinal als Kofaktor wiederum ist eine quantenmechanische Beschreibung der elektrostatischen Interaktionen notwendig, um zu gewährleisten, dass sein Verhalten innerhalb des Proteins korrekt dargestellt wird. In dieser Doktorarbeit wurden zwei verschiedene theoretische Ansätze angewandt, um Kofaktoren zweier Proteine zu untersuchen: klassische Mechanik und Kraftfeldoptimierung zur Beschreibung der eisenhaltigen Kofaktoren von Photosystem II sowie kombinierte Quantenmechanik/Molekularmechanik zur Analyse der Auswirkungen, die die Proteinumgebung auf den Protonentransfer in Kanalrhodopsin hat.

Wir konnten neue CHARMM-Kraftfeldparameter für die Häm- und Nichthäm-Eisenkomplexe in Photosystem II herleiten. Indem wir quantenmechanische Daten zum Vergleich heranzogen und Testsimulation nutzten, konnten wir zeigen, dass die neuen Parameter eine stark verbesserte Beschreibung der intra- und intermolekularen Interaktion bieten und zuverlässige Ganz-Atom-Simulationen von Proteinen ermöglichen, die Häm- und Nichthämeisenkomplexe enthalten.

In Kanalrhodopsin deprotoniert die Schiffsche Base erst, nachdem die lichtinduzierte Isomerisierung des Retinals eingetreten ist. Wir haben den gewichteten Histogramm-Ansatz (Weighted Histogram Approach Method) genutzt, um Protonentransferpfade herzuleiten, um zu verstehen, welche Faktoren die Deprotonierung im Grundzustand des Proteins verhindern. Wir konnten drei wichtige Bedingungsfaktoren identifiziere: das Vorhandensein eines positiv geladenen Lysins (K132), Wassermoleküle im aktiven Zentrum sowie der Protonierungszustand von Glutamat-162, einem Gegenion der Schiffschen Base. Letzten Endes konnten wir demonstrieren, dass ein kombiniert quantenmechanisch/molekularmechanischer Ansatz zur Beschreibung von Proteinen in einer hydratisierten Lipidmembran genutzt werden kann. Außerdem haben wir ein Protokoll entwickelt, welches zur Untersuchung von Protonentransferpfaden und -energien unter Berücksichtigung der Protein- und Lipidumgebung herangezogen werden kann.

Publications

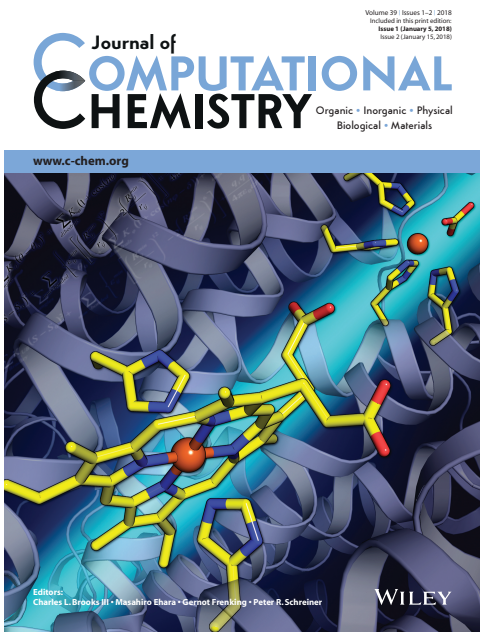
This dissertation is based on the following manuscripts, which have been published or are in preparation for publication in peer-reviewed journals:

Adam, S., Knapp-Mohammady, M., Yi, J. & Bondar, A.-N. Revised CHARMM force field parameters for iron-containing cofactors of photosystem II. *Journal of Computational Chemistry* **39**, 7–20 (2018). `http://dx.doi.org/10.1002/jcc.24918`

Adam, S. & Bondar, A.-N. Mechanism by Which Water and Protein Electrostatic Interactions Control Proton Transfer at the Active Site of Channelrhodopsin. PLOS ONE. Under Revision.

An additional paper was co-authored as a side project and is not included in the dissertation:

Guerra, F., **Adam, S.** & Bondar, A.-N. Revised force-field parameters for chlorophyll-a, pheophytin-a and plastoquinone-9. *Journal of Molecular Graphics and Modelling* **58**, 30–39 (2015).



My cover image design on the front page of issue 1 of the Journal of Computational Chemistry **39**, doi: 10.1002/jcc.25108

Si el poeta fuere casto en sus costumbres,
lo sera tambien en sus versos: la pluma es lengua del alma;
quales fueren los conceptos que en ella se engendraren,
tales seran sus escritos.

Miguel de Cervantes, *Don Quixote*

Acknowledgements

This PhD was funded by SFB 1078 - Protonation Dynamics in Protein Function. I thank the North-German Supercomputing Alliance (HLRN), the SOROBAN cluster of the FU and the Physics Department of the FU for providing HPC resources that have contributed to the research results of my thesis.

I thank my supervisor Ana-Nicoleta Bondar that introduced me to biophysics and who provided me with the support and freedom to pursue my research targets.

I thank Roland Netz, my second supervisor, for his support and advice.

I thank Jens Dreger and the other people of the IT Service of the Physics Department for setting up necessary hard- and software and for providing quick assistance whenever there was any problem with our local cluster or my computer.

I thank Joachim Heberle, Angelika Pasanec, Sylvia Luther and Kerstin Wagner for their support and understanding and for providing me with encouragement.

I especially thank Milan Hodoscek, who helped a lot with setting up software and who showed me how to add custom CHARMM commands—without which I would have run out of storage long before getting enough data on proton transfer.

I thank Stefan Milenkurac, who knows how to eat tuna. And god, does he know how to eat tuna. He really loves eating it! He's a pro at eating tuna—if you didn't know already. Also an acceptable human being. Thanks for your comments.

I thank Federico 'El Tafazzi' Guerra. He's had some tough years. Used to step in dog turds every now and then. Still knows how to enjoy a conference in Hünfeld.

I thank Christian Spakow, my authority on channelrhodopsin, punk music and K.I.Z. You, too, thanks for the feedback. But why Hohenschönhausen???

BTW—if you didn't know—I shared my office with Stefan, Fede and Christian for many years. Saw them more often than I saw my own reflection in the mirror. I very much have always liked them. They are very good guys. I know them. And they are very, very good guys. I remember that one time when Flo Flo almost got robbed by a bush. Thank god, Crafty Stefan was there.

I thank Frau Nadia (a real Knoppers® addict), Hamed, Nizar, Tina and Bouchra for fascinating conversations on PowerPoint and eating lunch together.

I thank Sebastian Lorch and Christian Nietner for reading my thesis. You won't regret it, once I'm rich and powerful! Unless you cross me, that is. Then you will.

Lastly, I like to thank Shani Bar-Tuvia for her patience and love for avocados.

Contents

List of Figures

List of Tables

Acronyms

4MEI	4-methylimidazole
B3LYP	Becke3,Lee–Yang–Parr functional
C1C2	channelrhodopsin-1–channelrhodopsin-2 chimæra
CGenFF	CHARMM General Force Field
CHARMM	Chemistry at Harvard Macromolecular Mechanics, a force field and an MD simulation and analysis software package
DFT	density functional theory
FEBIHIS	complex consisting of iron, bicarbonate and four 4-methylimidazoles
FEHIS	complex consisting of iron and four 4-methylimidazoles
HAEMHIS	complex consisting of hæm and two 4-methylimidazoles
HF	Hartree–Fock
MD	molecular dynamics
MM	molecular mechanics
MK	Merz–Singh–Kollman
MP2	Møller–Plesset perturbation theory
PES	potential energy scan
PMF	potential of mean force
QM	quantum mechanics
QM/MM	combined quantum and classical mechanics
RMSD	root-mean-square deviation
SCC-DFTB	self-consistent charge tight-binding density functional theory
TIP3P	a three-interaction-site water model
WHAM	Weighted Histogram Approach Method

Part One

Introduction

"I wanted to kill myself. I would have done it, too, if I had owned a gun. I was considering the gruesome alternatives—pills, slitting my wrists with a razor blade, jumping off a bridge—when another student called to ask me a detailed question on relativity. There was no way, after fifteen minutes of thinking about Mr. Einstein, that suicide was still a viable option. Divorce, certainly. Celibacy, highly likely. But death was out of the question. I could never have prematurely terminated my love affair with physics."

Arthur C. Clarke, *Rama II*

1 Background

When talking about proteins, the first thought that comes to mind might be of the fried egg one sometimes prepares for breakfast or of protein shakes that are popular with many fitness enthusiast[1]. Yet proteins are far more than just another food supplement.

Be it bacteria, be it plants, be it animals: Proteins can be found in all life forms and are an indispensable part of the machinery that keeps our bodies alive. They are macromolecules that are made up of chains of amino acid residues (Fig. 1.1) and perform a wide range of functions. Amongst others, proteins play a

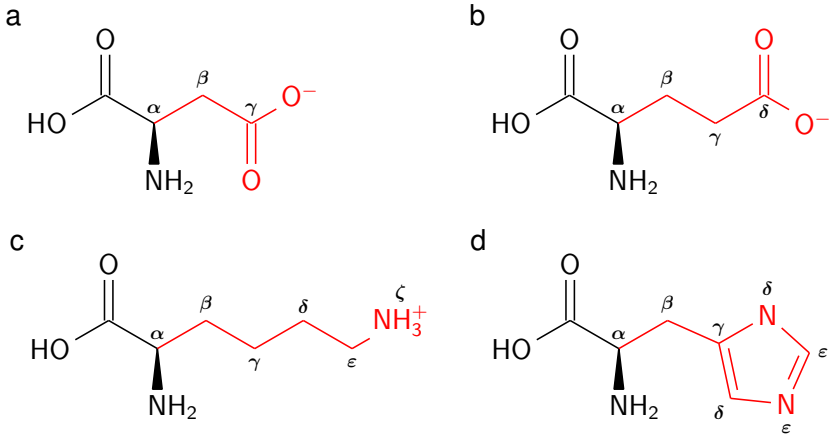


Figure 1.1: Skeletal formulae of the charged amino acids aspartate or aspartatic acid (a), glutamate or glutamic acid (b), lysine (c) and histidine (d). Amino acids can bind to each other by forming peptide bonds between their backbones (black), thus becoming amino acid residues in a peptide or protein. Each type of amino acid has a unique side chain (red), which, depending on the protonation state, can be either negatively (a,b) or positively charged (c) or neutral (d). The side chains are attached to the C_α carbon of the backbone, and their atoms are referred to by using Greek letters.

role in drug resistance and the removal of toxins from inside the cell[2,3,4,5] or in importing specific ions and regulating the cellular ionic concentration[6,7]. Proteins, however, are also involved in more well-known processes such as respiration[8], photosynthesis[9,10] or vision[11,12,13].

During photosynthesis, for instance, the protein complex photosystem II[9] splits water into molecular oxygen. For photosystem II to function properly, the presence of certain molecules like non-hæm iron and bicarbonate are essential[14,15]. In vision, retinal is necessary to absorb light, thereby activating a receptor protein[13]. These molecules or metal ions that are needed for a protein to fulfill its biological task are called cofactors[16].

1.1 | Describing Cofactor-Containing Proteins

To be able to characterize and understand the working of cofactor-containing proteins, it is of the utmost import to have a good understanding of their cofactors. In this work, we studied the following cofactors: the iron-containing complexes of photosystem II and retinal inside channelrhodopsin.

Iron-Containing Cofactors of Photosystem II

Photosystem II can be analysed using experimental methods[9,10,17,18,19,20,21] or theoretical approaches such as quantum mechanics (QM)[22], but reliable description

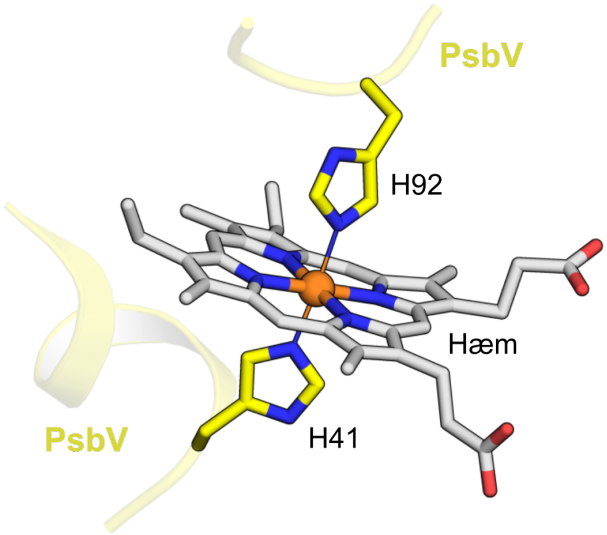


Figure 1.2: Bis-histidine-ligated hæm inside photosystem II (PDB ID: 3WU2[9]). Hæm b at the lumen side of photosystem II forms coordination bonds with N_ε of histidine-41 (H41) and histidine-92 (H92) of the subunit PsbV of photosystem II. Thin solid lines indicate coordination bonds between hæm (grey) and histidine (yellow).

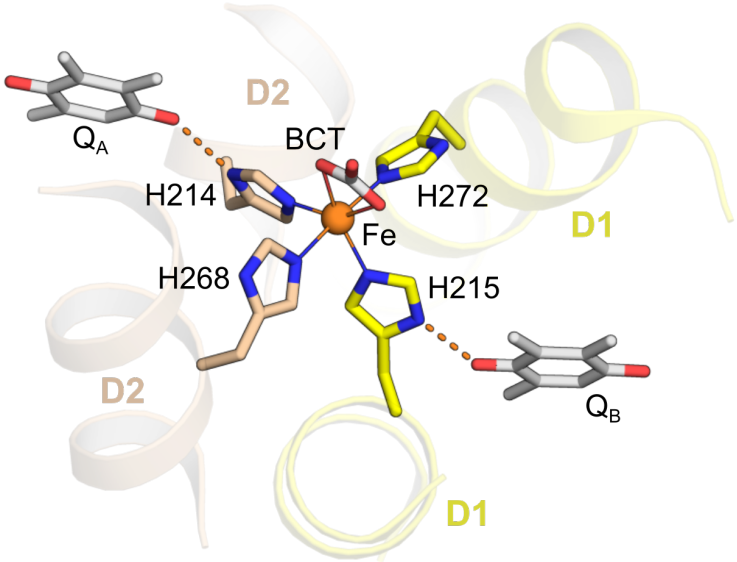


Figure 1.3: Non-hæm iron complex inside photosystem II (PDB ID: 3WU2[9]). The non-hæm iron is bound to bicarbonate (grey), to histidine-214 (H214) and histidine-268 (H268) of subunit D2 (wheat) and to histidine-215 (H215) and histidine-272 (H272) of subunit D1 (yellow). H214 and H215 connect the non-hæm iron–bicarbonate complex to the quinones Q_A and Q_B (BCT, grey) via a hydrogen-bonded network. Thin solid lines indicate coordination bonds between non-hæm iron (orange) and bicarbonate (grey) or histidine (yellow, wheat). Hydrogen bonds are shown as dashes.

of the conformational dynamics of photosystem II is required for comprehending how this membrane-embedded protein–cofactor complex generates oxygen from water. The usage of force fields and classical molecular dynamics (MD) is another theoretical tool that allows for studies with longer time scales than QM. In the case of photosystem II, however, a specific challenge is the lack of reliable force field parameters to describe the geometry and non-bonded interactions of the two iron-containing cofactors: (i) six-coordinated hæm with two histidine side chains bound to the hæm iron and (ii) the complex consisting of non-hæm iron bound to bicarbonate and four histidines.

The photosystem II dimer from *Thermosynechococcus vulcanus* contains two hæm b molecules bound to each monomer[9]. Each one of the hæms is bound to two histidines, i.e. the hæm is bis-histidine ligated (Fig. 1.2). Absence of the hæm or mutations in its vicinity can slow down the recovery rate of the oxygen activity of photosystem II[23]. Moreover, not having hæm results in a slower assembly of photosystem II[23], even though, for certain mutants, photoautotrophic growth and assembly are still possible[23,24]. These observations highlight the importance of understanding the conformational dynamics close to hæm.

The non-hæm iron in photosystem II is coordinated by four nitrogen atoms from histidine side chains contributed by subunits D1 and D2, and by the

carboxylate oxygen atoms of the bicarbonate molecule[9,25] (Fig. 1.3). In the crystal structure, the hydroxyl group of the bicarbonate molecule hydrogen-bonds to a water molecule and to two tyrosine side chains[9]. Two of the histidine side chains, H268 and H272, hydrogen-bond to nearby protein groups, whereas H214 and H215 hydrogen-bond to the quinone groups Q_A and Q_B, respectively (Fig. 1.3). This network of hydrogen bonds could be important for electron transfer at Q_A and Q_B[25].

To reliably describe these two complexes of photosystem II using classical MD, we derived new force field parameters for the CHARMM force field[26] using the CHARMM General Force Field (CGenFF) protocol[27].

Retinal and Retinal Proteins

Retinal is a molecule consisting of a long polyene chain and a six-membered carbon ring (Fig. 1.4). As a chromophore, it absorbs light of a specific frequency and gives the protein it is bound to its colour. Retinal can be found in many light-sensitive proteins[28,29,30,31,32].

In 1971, Dieter Oesterhelt showed that the cell membrane of *Halobacterium halobium* contained bacteriorhodopsin, which had retinal attached to it[33]. By pumping protons from the cytoplasmic to the extracellular side of the membrane, the cell could use the difference in the proton gradient to synthesize ATP—a simple form of photosynthesis—and recently, there have been suggestions for using light-sensitive proton pumps to generate bioenergy[34]. Despite the early identification of bacteriorhodopsin, it took nearly two decades before a three-dimensional atomic resolution map was published[35] and deposited in the Protein Data Bank (PDB)[36], a database for three-dimensional structures of proteins.

In this structure of bacteriorhodopsin from *H. halobium* (PDB ID: 1BRD)[35],

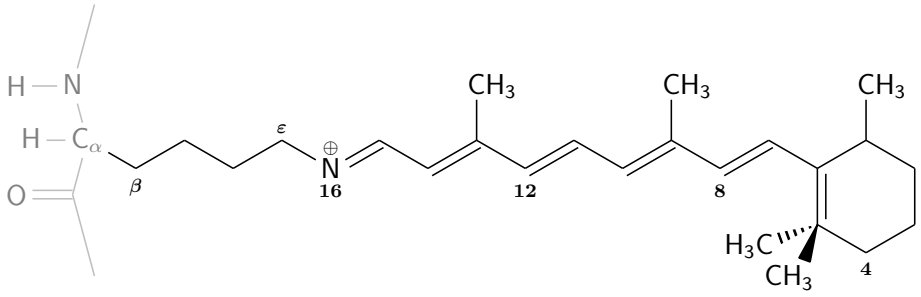


Figure 1.4: All-*trans* Retinal bound to lysine. The skeletal formula presents the naming convention used to describe retinal attached to a lysine amino acid residue of a protein. The Schiff base is located at C_{15}=N_{16}, and all bonds are shown in *trans* configuration. The protein backbone is shown in grey.

retinal could be seen inside of the protein. Retinal is attached to the protein by binding to lysine and forming a Schiff base linkage (Fig. 1.4). The absorption of light by the retinal and the subsequent deprotonation of the Schiff base initiate a photocycle, during which the proton-pumping activity takes place. In organic solvents, retinal absorbs at 440 nm[37]. This absorption maximum is changed by the protein environment[38], and research has been conducted to understand the mechanism behind this so-called opsin shift[31,37,39,40,41,42].

In the past, there had been difficulties deriving classical models for retinal, and only after continued work, conducted over many years by several research groups using bacteriorhodopsin as a reference system, did reliable retinal force field parameters emerge[37,43,44,45,46,47,48].

Initial steps in the theoretical analysis of retinal included division of the chromophore into smaller parts as full-molecule computations were too expensive at that time. These early attempts took existing parameters from small compounds like methylene and otherwise directly used values computed with QM[43] or focussed on retinal isomerization[44]. Later, more expensive QM methods and larger systems became feasible and the impact of the protein surrounding was taken into consideration[45,46,47]. Finally, it was shown that charge transfer and polarization effects are necessary to correctly describe the retinal pocket in bacteriorhodopsin[37,48].

The classical parameters that came out of these efforts to understand retinal in bacteriorhodopsin enabled more reliable classical MD simulations than before, but because deprotonation, like light-absorption, is a chemical process, and because charge transfer and polarizability are not properly included in classical approaches, the need for QM studies remained, and QM calculations have been used to characterize contributions to the reaction energetics of proton transfer involving retinal[45,49,50,51,52].

Channelrhodopsin

One of the main focusses of this PhD project was the study of channelrhodopsin, which is a seven-helical light-driven cation channel that contains a retinal chromophore[53,54]. Channelrhodopsin is a dimer[55], and retinal is bound to lysine-296 (K296) in, what we call here, the Schiff base region (Fig. 1.5).

Absorption of a photon by the all-*trans* chromophore (Fig. 1.4) leads to isomerization about the C_{13}=C_{14} bond, resulting in a 13-*cis* isomer. The isomerization starts the photocycle that leads to cations entering the cell and causing a depolarization of the plasma membrane[56].

Earlier studies of *Chlamydomonas reinhardtii* had indicated the existence of two phototaxis receptors[57], and the discovery of channelrhodopsin-1[58] and channelrhodopsin-2[59] confirmed these observations. Channelrhodopsin was the first light-gated cation channel ever discovered[60] and has found since application in the new-founded field of optogenetics[56,61,62,63]. Optogenetics expresses light-sensitive ion channels in cells, like neurons, and then uses these channels to control the altered cells or the organisms containing them.

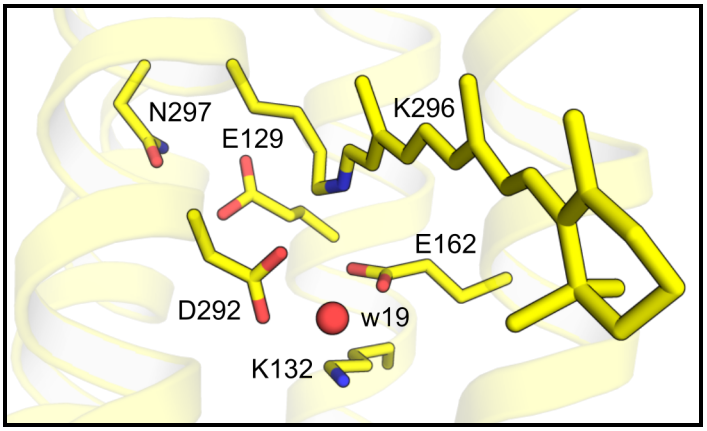


Figure 1.5: The Schiff base region in the crystal structure of channelrhodopsin (PDB ID: 3UG9[64]). Retinal is attached to lysine-296 (K296), and the negative counterions glutamate-162 (E162) and aspartate-292 (D292) are located within hydrogen-bonding distance of the retinal Schiff base. The positively charged lysine-132 (K132) closes off the Schiff base region on the extracellular end (bottom), while glutamate-129 (E129) forms a hydrogen bond with asparagine-297 (N297) and defines the intracellular end. In the crystal structure, w19 was the only water molecule resolved inside the Schiff base region. Throughout this dissertation channelrhodopsin-1 notation[64] will be used to refer to specific amino acid residues.

To create proteins with enhanced optogenetic properties, it has been of major interest to better understand what influences properties like frequency of light absorption or recovery times of channelrhodopsin. Channelrhodopsin-2, for instance, despite desensitizing under continuous illumination, has been used more frequently than channelrhodopsin-1 because the latter is more difficult to express in mammalian cells[65]. After mutagenetic studies yielded accelerated versions of channelrhodopsin-2, efforts were made to combine channelrhodopsin-1 and channelrhodopsin-2, in the hope of obtaining chimæras with improved optogenetic properties[65,66].

Finally, in 2012, Kato *et al.* succeeded in crystallizing a channelrhodopsin-1–channelrhodopsin-2 chimæra (C1C2) (PDB ID: 3UG9[64]) that had its first five transmembrane helices taken from channelrhodopsin-1 and the last two taken from channelrhodopsin-2[64]. Later, experimental studies supported by theoretical

calculations were used to design a C1C2 mutant with a desired blue-shift (PDB ID: 4YZI[31]). After conclusion of the work on this PhD project, the group of Gordely and coworkers published crystal structures for wild type channelrhodopsin-2—the protein used most often in optogenetics[60]—and for a mutant of channelrhodopsin-2 (PDB ID: 6EID and 6EIG, respectively[67]).

Despite the increasing amount of channelrhodopsin structures, there are still many unknowns regarding the functioning of channelrhodopsin. For example, channel-opening and proton-pumping activity in channelrhodopsin is initiated only after the retinal isomerization from all-*trans* to 13-*cis*. The protein environment of the retinal binding site must play an important role in ensuring the continued protonation of the Schiff base until photoisomerization occurs. Here, we use molecular mechanics (MM) and combined quantum and classical mechanics (QM/MM) to characterize determinants for the stability of the retinal Schiff base proton in the closed/dark state of the channelrhodopsin chimæra C1C2. Combined with the parametrization effort for the iron-containing cofactors of photosystem II, we hope that these different approaches to studying cofactors will provide important new insights into how cofactors influence and are influenced by their protein environment.

لا شَيْءْ

قبل أَنْ تَشْرَعَ في القِرَاءَةِ
ضَعْ القَلَمْ
وَانْظُرْ إلى الحِبْرِ السَائِلِ
كَمْ يَفْهَمُ في النَزْفِ.
تَفَقَّهْ
في مِسَاحَةِ الأُفُقِ
وضِيقِ البَصَرِ، إتِّسَاعِ البَصِيرَةِ
وَخُذْلانِ الأَصَابِعْ
ولا تَلُمْنِي/لا تَلُمْ أَحَدَاً
إنْ مِتُّ قبل أَنْ تَقْرَأَ لا شَيْءَ
قبل أَنْ نَفْهَمَ في النَزْفِ!

الصادق الرضي

2 Methods

In this dissertation, calculations will be presented that were performed utilizing a variety of quantum mechanics (QM) and molecular mechanics (MM) methods, such as the derivation of potential of mean force (PMF) profiles to analyse the proton transfer energetics in channelrhodopsin or the computation of water interaction energies for reduced model systems.

The methods chapter will introduce computational approaches and system setups used, and it will explain the philosophy behind the process of deriving force field parameters.

2.1 | Quantum Mechanical Methods

Where chemical reactions take place, QM methods become a necessity. The choice of QM method should be based on the system of interest as a computationally more expensive method is not guaranteed to yield the best results.

QM methods that derive their solutions from solving the Schrödinger equation[68] are referred to as *ab initio* quantum chemistry methods[69], a term that was initially used to indicate that QM calculations had been performed independently[70]. The Hartree–Fock (HF) or self-consistent field method is one such *ab initio* approach that forms the basis for many higher-level methods like Møller–Plesset perturbation theory (MP2)[71].

The Hartree-Fock Method

The simplest form of HF, restricted HF, describes the wave functions with a single Slater determinant. Here, all orbitals are occupied by an electron pair. Electrons are fermions, i.e. they are indistinguishable particles. To fulfill the Pauli exclusion principle, the electrons' wave function needs to be antisymmetric with respect to the exchange of two electrons. When considering the case for two electrons of same spin α, the ground state can be written as a product of both of their wave functions—the so-called Hartree product:

$$\psi_{\mathrm{HP}} = \psi_a(1)\alpha(1)\psi_b(2)\alpha(2). \tag{2.1}$$

A linear combination of two Hartree products is used to derive an antisymmetric wave function, the Slater determinant[72, pp. 119-124]:

$$\begin{aligned}\psi_{\mathrm{SD}} &= \frac{1}{\sqrt{2}}\left(\psi_a(1)\alpha(1)\psi_b(2)\alpha(2) - \psi_a(2)\alpha(2)\psi_b(1)\alpha(1)\right) \\ &= \begin{vmatrix}\chi_1(1) & \chi_2(2) \\ \chi_1(1) & \chi_2(2)\end{vmatrix},\end{aligned} \tag{2.2}$$

where $\chi_a(n) = \psi_a(n)\alpha(n)$.

This ansatz can be used to describe the molecular orbitals. The one-electron Fock operator, which is applied to the HF wave functions, is defined as the sum of the one-electron core Hamiltonian and the HF potential[72, p. 126]:

$$f_i = -\frac{1}{2}\nabla_i^2 - \sum_k^{\text{nuclei}} \frac{Z_k}{r_{ik}} + V_i^{\mathrm{HF}}\{j\}. \tag{2.3}$$

The HF potential term is derived from the Coulomb operator J_{ij} that describes Coulombic repulsion between the electrons and from the exchange operator K_{ij} that accounts for the antisymmetry of the wave function and corrects the Coulombic term:

$$V_i^{\mathrm{HF}}\{j\} = 2J_{ij} - K_{ij}, \tag{2.4}$$

where $J_{ij} = \left\langle \phi_i(1)\phi_j(2) \left| \frac{1}{r_{12}} \right| \phi_i(1)\phi_j(2) \right\rangle$ and $K_{ij} = \left\langle \phi_i(1)\phi_j(2) \left| \frac{1}{r_{12}} \right| \phi_i(2)\phi_j(1) \right\rangle$.

Next, using the Roothaan approach, the Hartree–Fock–Roothaan equations[73] are constructed, and the secular equation needs to be solved:

$$det(\mathbf{F} - \epsilon\mathbf{S}) = 0. \tag{2.5}$$

The overlap matrix $\mathbf{S}$ quantifies the overlaps of the atomic orbital basis functions, and the elements of the Fock matrix $\mathbf{F}$ are given by:

$$\begin{aligned} F_{\mu\nu} = \left\langle \mu \left| -\frac{1}{2}\nabla^2 \right| \nu \right\rangle - \sum_k^{\text{nuclei}} Z_k \left\langle \mu \left| \frac{1}{r_k} \right| \nu \right\rangle \\ + \sum_{\lambda\sigma} \mathbf{P}_{\lambda\sigma} \left[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma) \right], \end{aligned} \tag{2.6}$$

where $\mathbf{P}_{\lambda\sigma}$ is an element of the density matrix.

The two-electron integral is computed using the probability density ϕ of the electrons:

$$(\mu\nu|\lambda\sigma) = \iint \phi_\mu(1)\phi_\nu(1)\frac{1}{r_{12}}\phi_\lambda(2)\phi_\sigma(2)d\mathbf{r}(1)d\mathbf{r}(2). \tag{2.7}$$

The self-consistent field algorithm starts by defining a molecular guess geometry. Then the overlap, one-electron and two-electron integrals are calculated, and an initial density matrix is guessed. The Hartree–Fock–Roothaan equations are constructed, and the secular problem is solved. Afterwards, the density matrix is updated. This is repeated until the density matrices reach self-consistency. Once this is the case, the geometry can be optimized, and one goes back to calculating the integrals. After the geometry has reached a satisfactory level, the optimization process is finished[72, pp. 126-129].

Møller–Plesset Perturbation Theory

The HF method has the disadvantage that it never converges to the true energy E. Instead, it gives the energy in the HF limit (E_{HF}). The difference between these two values gives the electron correlation energy E_{corr}[72, pp. 165]:

$$E_{\mathrm{corr}} = E - E_{\mathrm{HF}}. \tag{2.8}$$

MP2 is a method, where electron correlation is considered. The approach is similar to the HF method. In the first step the one-electron Fock operator is applied to the HF wave function:

$$\mathbf{H}^{(0)}\mathbf{\Psi}^{(0)} = \sum_{i=1}^{n} f_i \mathbf{\Psi}^{(0)} = \sum_{i=1}^{\text{occ.}} \varepsilon_i \mathbf{\Psi}^{(0)}. \tag{2.9}$$

Because the repulsion between the electrons of a specific orbital and all other electrons is calculated for each orbital, the total repulsion is overestimated by a factor of two, and a correction term $\mathbf{V}$ needs to be applied:

$$\mathbf{V} = \sum_{i}^{\text{occ.}} \sum_{i>j}^{\text{occ.}} \frac{1}{r_{ij}} - \sum_{i}^{\text{occ.}} \sum_{j}^{\text{occ.}} J_{ij} - \frac{1}{2} K_{ij}. \tag{2.10}$$

The two contributions shown above are summed up to give the first-order corrected MP energy—the HF energy. The second-order correction term is calculated by again using the Rayleigh–Schrödinger perturbation theory approach, and is given by the following equation[72, pp. 216-222]:

$$a_o^{(2)} = \sum_{j>0} \frac{\left| \left\langle \mathbf{\Psi}_j^{(0)} \left| \mathbf{V} \right| \mathbf{\Psi}_0^{(0)} \right\rangle \right|^2}{a_0^{(0)} - a_j^{(0)}} = \sum_{i}^{\text{occ.}} \sum_{j>i}^{\text{occ.}} \sum_{a}^{\text{vir.}} \sum_{b>a}^{\text{vir.}} \frac{\left[(ij|ab) - (ia|jb)\right]^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}, \tag{2.11}$$

where $(ij|ab)$ is the two-electron integral as mentioned in Eq. (2.7).

When looking at these equations, it becomes apparent why MP2 is called a post-HF method: MP2 is a perturbation theory-based correlation energy correction applied to the HF energy. (The number after 'MP' gives the highest-order term.) Usage of the second-order correction term still allows for computationally efficient solving of the equation, and very good results can be obtained from MP2.

Nonetheless, MP2 remains an expensive method. In 2007, HF could be used with up to ∼5000 basis functions, while MP2 and MP4 had a limit of ∼800 and ∼300–400 basis functions, respectively[74, pp. 183-186]. Improvements in technology can raise these limits, but the scaling with system size remains the same, i.e. $\mathcal{O}(N^4)$ for HF, $\mathcal{O}(N^5)$ for MP2, $\mathcal{O}(N^7)$ for MP4 and $\mathcal{O}(N^9)$ for MP6[74, p. 184], where N is the number of basis functions.

Despite this dramatic increase in computational cost, going for higher-order MPn remains a potentially attractive possibility. We note, however, that improvements in accuracy gained can be very minor and that these higher-order calculations are not always convergent, with lower-level MPn possibly giving better results[75,76,77].

MP2 is the least expensive of the MPn methods and includes electron correlation

to a sufficient degree. For these reasons, MP2 is used for force field parametrization at stages where high-accuracy geometries, energies or frequencies are vital[27]. The most efficient way to use MP2 is to do a HF optimization beforehand, and use the geometry obtained for further MP2 calculations, thus reducing the number of MP2 steps required[71].

Density Functional Theory

Instead of using a wave function approach like HF, density functional theory (DFT) is based on the idea that a system's energy can be described by the electron density. Hohenberg and Kohn showed that the ground state energy of an electronic system is defined completely by its electron density[78]. DFT aims to derive functionals that describe the functional dependence between energy and density because the exact connection between both is not known[79]. The computational expense of DFT methods is comparable to HF, but the results obtained are of better quality[80, pp. 233-274].

Some DFT methods can be distinguished by the functional form they use to describe the exchange–correlation energy[80, pp. 244-253]. In the Local Density Approximation (LDA), the electron density is assumed to vary only slowly. More generally, for a spin-polarized system with the two spin densities ρ_α and ρ_β the Local Spin Density Approximation (LSDA) is used and its exchange energy is given by:

$$E_{\text{ex.}}^{\text{LSDA}} = \left(-2^{1/3}\right)\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\int\left(\rho_\alpha^{4/3}+\rho_\beta^{4/3}\right)d\mathbf{r}. \tag{2.12}$$

LSDA, however, significantly overestimates both exchange energy and electron correlation[80, p. 248]. Generalized Gradient Approximation (GGA) methods were created to improve LSDA, and they depend on the electron density as well as on its first derivative. The B88 correction[81] (Eq. (2.13)), for example, improves the error in the LSDA exchange energy by almost two orders of magnitude, while the Lee, Yang and Parr (LYP) functional[82,83] is a GGA functional that improves the correlation energy. The quality of the functional can be further improved by using both B88 and LYP to give the BLYP approach[80, pp. 249-250]:

$$\begin{aligned} E_{\text{ex.}}^{\text{B88}} &= E_{\text{ex.}}^{\text{LDA}} + \Delta E_{\text{ex.}}^{\text{B88}} \\ \Delta E_{\text{ex.}}^{\text{B88}} &= -\beta\int \rho^{4/3}\frac{x^2}{1+6\beta x\sinh^{-1}x}d\mathbf{r}. \end{aligned} \tag{2.13}$$

Hybrid functionals are another class of DFT functionals. They are created by adding the HF exchange energy. The Becke3,Lee–Yang–Parr func-

tional (B3LYP)[84,85] was made by combining a hybrid functional approach with the addition of the B88 exchange correction and the LYP correlation functional and uses three parameters a, b and c that have been fitted to experimental data:

$$E^{\mathrm{B3LYP}} = (1-a)E_{\mathrm{ex.}}^{\mathrm{LSDA}} + aE_{\mathrm{ex.}}^{\mathrm{HF}} + b\Delta E_{\mathrm{ex.}}^{\mathrm{B88}} + (1-c)E_{\mathrm{corr.}}^{\mathrm{LSDA}} + cE_{\mathrm{corr.}}^{\mathrm{LYP}}. \quad (2.14)$$

In B3LYP, the LSDA correlation energy is calculated using the Vosko, Wilk, and Nusair functional (VWN)[86]. VWN has different parametrizations, and as a consequence, there exist different definitions of B3LYP. B3LYP often gives good results, but one has to keep in mind that for DFT there are no standard methods like MP2, and the choice of functional should be made depending on the system[80, pp. 233-274].

Tight-Binding Density Functional Theory

Semi-empirical methods have been derived to reduce the cost of *ab initio* methods. They are based on *ab initio* calculations, and to increase the speed of computations, they use drastic approximation, such as only considering valence electrons explicitly and describing them with a minimum basis set[87]. In addition, parameters are added to help overcome the limitations of the use of a minimal basis set[80, pp. 285-287].

Tight-binding DFT (DFTB) is a semi-empirical method that is based on DFT and can be used for large systems and systems involving transition metals. By adding corrections due to charge equilibration, second (DFTB2) and third order DFTB (DFTB3) is obtained. Because these extended DFTB approaches modify DFTB by a self-consistent redistribution of partial charges, they are also known as self-consistent charge tight-binding density functional theory (SCC-DFTB)[88,89,90,91]. Like DFT, DFTB does not account for dispersion[92], and for systems where dispersion has a significant effect additional corrective terms, like the Grimme dispersion correction[93,94], need to be added.

2.2 | The CHARMM Force Field

Highly accurate results can be obtained from QM, yet calculations can only be made for small system sizes. Describing a whole protein completely with QM is not possible at the moment. The MM approach is based on classical mechanics and enables simulations of small biological systems, e.g. a protein in a lipid membrane. The potential energy of the system is characterized by mathematical

$$
\begin{aligned}
U(\vec{R}) = & \sum_{\text{bonds}} K_b(b-b_0)^2 + \sum_{\text{angles}} K_\theta(\theta-\theta_0)^2 \\
& + \sum_{\text{Urey-Bradley}} K_{UB}(S-S_0)^2 \\
& + \sum_{\text{dihedrals}} K_\varphi(1+\cos(n\varphi-\delta)) + \sum_{\text{impropers}} K_\omega(\omega-\omega_0)^2 \\
& + \sum_{\text{residues}} U_{\text{CMAP}}(\varphi,\psi) \\
& + \sum_{\text{non-bonded}} \left\{ \varepsilon_{ij}^{min} \left[\left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \right\}
\end{aligned}
\tag{2.15}
$$

Equation 2.15: The CHARMM force field equation.

equations and parameters: the force field.

The CHARMM force field and software package[26,95] is one of the most widely utilized all-atom force fields and gives a good description of protein properties and interactions[96,97]. All-atom means that all of the atoms are represented explicitly without any of the hydrogens being summed into the heavier atoms. CHARMM uses a class I additive potential energy equation. The CHARMM equation contains the common potential terms, but also adds a Urey-Bradley term (UB) and a backbone torsional correction term (CMAP) (Eq. (2.15))[98]. These terms can be classified into two categories: intramolecular, i.e. internal or bonded, or intermolecular, i.e. external or non-bonded. Bonds, angles, UB, dihedrals, improper dihedrals (out-of-plane bending) and CMAP form the bonded terms. The non-bonded energy is calculated from the van der Waals and electrostatic interactions.

To create a force field, a set of corresponding parameters is needed. Bonds b, angles θ, impropers ω and UB distance S are described by harmonic potentials that include an equilibrium value (denoted by subscript 0 in Eq. (2.15)), and a force constant $K_{\text{parameter}}$. The dihedral term is a sum of cosine functions of multiplicity n and with a phase shift δ. CMAP was initially an extension to the CHARMM22[99] force field. It was introduced to improve the treatment of backbone dihedrals in peptides[100,101]. The UB term treats 1,3-non-bonded interactions and optimizes angle bending. This potential is used where the vibrational spectra require correction, as might be the case for separating symmetric and antisymmetric bond stretching modes in aliphatic molecules. The non-bonded parameters comprise the partial charges q_i, the depth of the Lennard–Jones (LJ) well ε_{ij}, and the minimum van der Waals interaction radius R_{ij}^{min}.

Because CHARMM uses fixed partial atomic charges, it is called an additive force field and lacks polarization. The dipole moment has to be overestimated

during the charge fitting procedure to mimic polarizability. Combining a partial charge model and van der Waals interactions gives a sufficient description of hydrogen bonding and dispenses with the need for hydrogen bonding parameters[102].

Even including CMAP and UB terms, CHARMM remains a class I force field. This means that cross-terms are not explicitly included in the force field equation. Cross-terms refer to the coupling terms between bonds, angles and torsional motion and are used in class II force fields, like the Merck Molecular Force Field[103]. Bond-angle coupling is an example for such a cross-term. It can be obtained from the product of the bonds and angles terms. The advantage of class II force fields is that they can reproduce the bonding behaviour and especially the vibrational spectra more accurately[104]. They require, however, a larger number of parameters. The advantage of class I force field like CHARMM, is that fewer parameters are required, and subsequently calculations become less expensive. For class I force fields, the derivation of transferable force field parameters becomes easier as well, and they can give satisfactorily describe biomolecules at room temperature[102].

The Water Model in CHARMM

CHARMM has been optimized for use with the TIP3P water model[105], an explicit solvation model, i.e. the water is represented by individual molecules. TIP3P is a three-interaction-site model, i.e. the water molecule consists of two hydrogens and one oxygen.

SPCE/E[106] is another widely used three-site model. It is a reparametrization of the single point charge (SPC) model[107]. It contains a correction for polarization and gives a better description of the hydration shells of water than TIP3P does[108]; however, the improvement in water-pair interactions comes at the expenses of

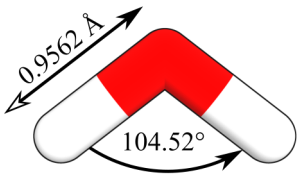


Figure 2.1: The TIP3P water model. TIP3P[105] is a three-interaction-site water model, where the water molecule consists of three particles: one oxygen and two hydrogens. The O–H distance is fixed at 0.9572 Å, and the angle ∠HOH has a value of 104.52°. The oxygen and hydrogen charge is −0.834 and +0.417, respectively.

the water-solute interactions.

The structural properties of the water molecules can be improved by separating mass and charge of the oxygen into two separate particles, which gave the four-site TIP4P model[105]. Whereas in a three-site model, $3 \times 3 = 9$ interactions have to be computed for two water molecules interacting with each other; in a four-site model, the splitting of the oxygen into charge and mass part adds an additional interaction, thereby increasing the total number of interaction distances to ten. Because water molecules claim most of the computation time in hydrated simulations, the usage of TIP4P can significantly increase the computation time of molecular dynamics (MD) calculations.

The CHARMM developers considered the TIP3P, SPC/E and TIP4P model before deciding on TIP3P[99]. Despite the CHARMM force field being optimized for TIP3P interaction, TIP4P and TIP5P[109], a five-site model, can give good results with CHARMM without force field reoptimization[110,111].

Combining Quantum and Classical Mechanics

Combined quantum mechanical/molecular mechanical (QM/MM) methods seek to combine the advantages of QM and MM; namely the ability of QM to describe chemical reactions and the speed of classical mechanics. One possible implementation of QM/MM is briefly described below[112].

In QM/MM, calculations on a region of interest, e.g. a chromophore in a light-activated protein, are performed using QM, while the rest of the system is described at the less detailed MM level. The entire system's Hamiltonian operator is then the sum of the MM region $\hat{H}_{\mathrm{MM}}$, the QM region $\hat{H}_{\mathrm{QM}}$ and the interaction between QM and MM region $\hat{H}_{\mathrm{QM/MM}}$[113]:

$$\hat{H} = \hat{H}^{\mathrm{QM}} + \hat{H}^{\mathrm{MM}} + \hat{H}^{\mathrm{QM/MM}}. \tag{2.16}$$

The QM and MM regions interact with each other via Coulombic and van der Waals interactions[112], and the total energy of the system is given by:

$$E = \langle \Psi | \hat{H}^{\mathrm{QM}} + \hat{H}^{\mathrm{QM/MM}}_{\mathrm{elec.}} | \Psi \rangle + E^{\mathrm{QM/MM}}_{\mathrm{vdW}} + E^{\mathrm{MM}}, \tag{2.17}$$

where $E^{\mathrm{QM/MM}}_{\mathrm{vdW}}$ is the van der Waals interaction between QM and MM, and the Coulombic interaction $\hat{H}^{\mathrm{QM/MM}}_{\mathrm{elec.}}$ is the one-electron operator given by:

$$\hat{H}^{\mathrm{QM/MM}}_{\mathrm{elec.}} = \sum_{A \in \mathrm{MM}} \sum_{B \in \mathrm{QM}} \frac{Q_A \Delta q_B}{|\mathbf{r}_A - \mathbf{r}_B|}. \tag{2.18}$$

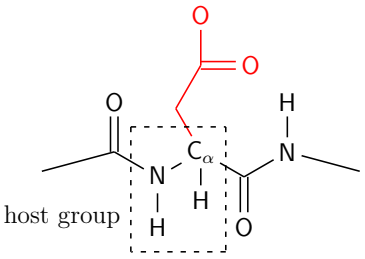


Figure 2.2: Illustration of the divided frontier charge scheme. The side chain atoms of aspartate are described using QM, while the rest of the protein is described using MM. The classical mechanical host atom is the atom binding to the QM region (C_α), and the host group is the MM group directly connected to the QM region (indicated by a dashed rectangle). In the divided frontier charge scheme, the charge of the host atom is redistributed over the remainder of the host group [113].

It is not always possible to include complete molecules in the QM region, and a decision has to be made, where to cut the MM description and start QM. The classical force field part does not mind if an atom gets deleted; for the QM region, however, one needs to ensure that the chemistry of the compounds is not changed. To preserve the QM region's behaviour as much as possible, the single link atom method [114] places a hydrogen that is only seen by the QM region on the interfacing bond between QM and MM region [115].The QM and the MM atom forming the bond connecting QM and MM region are called QM host atom and MM host atom, respectively. The description of the QM/MM interface can be further improved, by applying the divided frontier charge scheme [113]. The divided frontier charge scheme redistributes the charge of the MM host atom across its MM group (Fig. 2.2).

SCC-DFTB/MM

The SCCDFTB module in CHARMM [112] is used to perform SCC-DFTB/MM calculations [116]. A big advantage of the this module is the low computational cost of the SCC-DFTB method as well as the possibility to add cut-offs for the QM/MM electrostatic interactions and to use Ewald summation [117]. These options allow the application of SCC-DFTB/MM to large systems consisting of a protein in a hydrated lipid bilayer in a periodic boundary box.

2.3 | Deriving Force Field Parameters

A major part of this PhD project consisted of deriving force field parameters for the hæm and non-hæm iron complexes found in photosystem II (Figs. 1.2 and 1.3). These parameters were to be used with the CHARMM force field, and

therefore had to be compatible with the parametrization protocol used by the CHARMM developers[99].

The recently developed CHARMM General Force Field (CGenFF)[27] added parameters for a wide range of compounds, functional groups and drug-like molecules. It is an extension to the CHARMM force fields and fully compatible with the CHARMM protein force field. CGenFF also introduced a standardized parametrization algorithm to facilitate: adding new compounds or adding functional groups to molecules already present in the force field; as well as linking together existing molecules.

As an extensible force field, CGenFF was an ideal choice to derive parameters for the hæm and non-hæm iron-complexes of photosystem II: The hæm, histidine and bicarbonate moieties of the complexes were already present in the CHARMM36 protein force field[97], and the CGenFF approach could be used to connect these molecules. By using the CHARMM parameters for hæm[118,119], having to parametrize hæm from scratch could be avoided.

Model Systems and Starting Coordinates

Starting coordinates for the parametrization of the hæm and non-hæm iron complexes were taken from monomer A of the crystal structure of photosystem II (PDB ID: 3WU2[9]). To facilitate QM computations and the parametrization process, components of the complexes were replaced by simplified model moieties as described below.

In photosystem II, the bis-histidine–hæm complex consists of hæm b forming

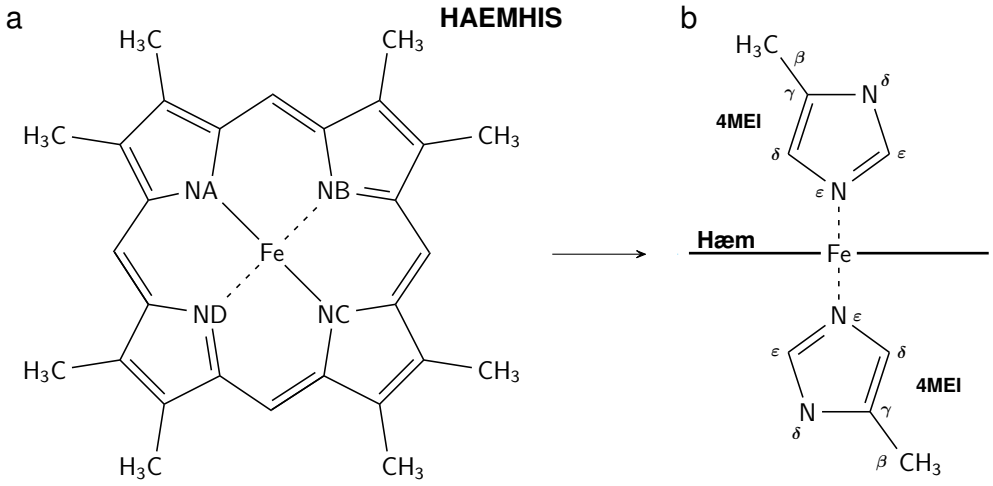


Figure 2.3: Chemical structure of the bis-histine–hæm model complex. (a) The propionic side chains of hæm b were replaced by methyl groups to obtain the hæm model moiety. (b) A side view of the complete HAEMHIS complex shows the N_ε atoms of the 4MEI groups forming a coordination bond with the hæm iron.

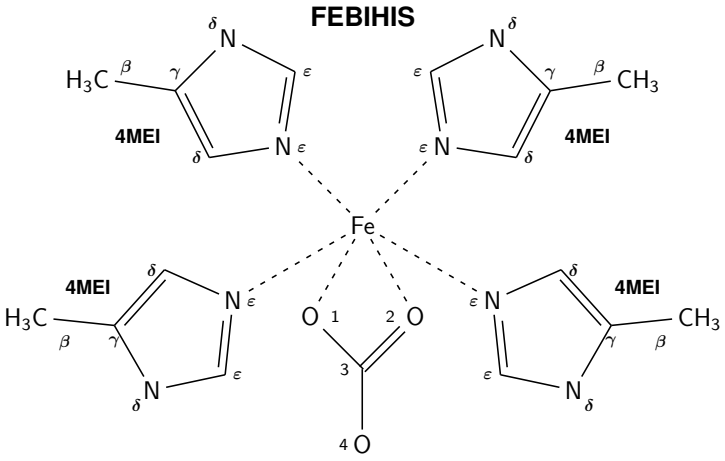


Figure 2.4: Chemical structure of the non-hæm iron–bicarbonate–histidine model complex. The FEBIHIS complex consists of a central iron forming coordination bonds with four 4MEI and bicarbonate. The 4MEI groups' N_δ is protonated.

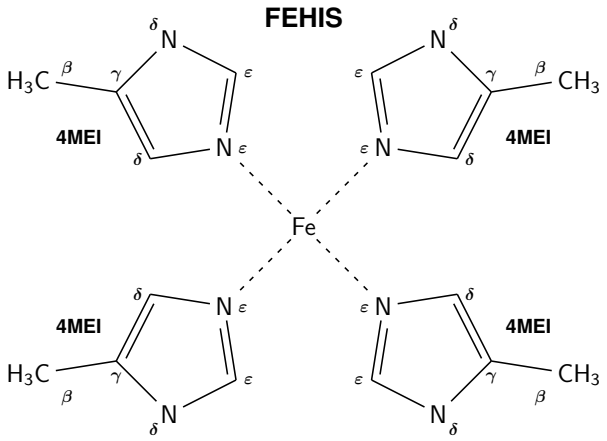


Figure 2.5: Chemical structure of the non-hæm iron–histidine model complex. The FEHIS complex has a central iron coordinating four 4MEI only. The 4MEI groups are δ-protonated.

a coordination bond with the ε-nitrogens of two δ-protonated histidine amino acid residues. We replaced the two negatively charged propionic groups and the ethyl group of hæm b with methyl groups (Fig. 2.3a). The histidines were represented by 4MEI (Fig. 2.3b). This simplified bis-histidine–hæm complex is denoted as HAEMHIS (Fig. 2.3).

Similarly, for the non-hæm iron–histidine-bicarbonate complex, we modelled each of the four histidine residues as 4MEI. The resulting complex was denoted as FEBIHIS (Fig. 2.4). We also derived a simplified non-hæm iron–histidine complex that did not contain a bicarbonate and denoted it as FEHIS (Fig. 2.5).

The iron in both bis-histidine-ligated hæm and the non-hæm iron complex of photosystem II can exist in its ferrous (Fe^{2+}) or ferric (Fe^{3+}) form[19,22,120]. To enable simulations with these two redox states of the iron, we derived parameters

for HAEMHIS, FEBIHIS and FEHIS containing ferrous and ferric iron. The total charge of ferrous and ferric HAEMHIS is 0 and +1, respectively; the total charge of ferrous and ferric FEBIHIS is +1 and +2, respectively; and the total charge of ferrous and ferric FEHIS is +2 and +3, respectively.

All 4MEI groups had N_δ protonated and N_ε unprotonated. We used the HBUILD command of CHARMM[26] to construct hydrogen atoms.

The CHARMM General Force Field Algorithm

The basic idea behind parametrizing a force field is to make MM calculations reproduce QM or experimental data. To retain compatibility within a force field, however, it is important to use the same parametrization algorithm for each addition to the force field.

The CGenFF[27] algorithm (Fig. 2.6) used here relies, where possible, on *ab initio* generated target data and is as follows: In the first step, files containing the topology of a compound as well as initial guess parameters need to be generated, and the partial charges are optimized using the QM geometry. After this, the iterative phase of the parametrization can begin. First, bonds and angles are optimized, then dihedrals and impropers. These are the so-called intramolecular parameters. Once convergence has been reached, the partial charges will be reoptimized—now using the MM geometry.

Partial charges and Lennard–Jones parameters form the intermolecular part of the force field parameters. No optimization of Lennard–Jones parameters was performed as fitting of Lennard–Jones parameters requires either experimental data or high-level QM calculations[27] that were not possible for complexes as large as HAEMHIS, FEBIHIS or complex consisting of iron and four 4-methylimidazoles (FEHIS). Moreover, Lennard–Jones parameters already present in standard CHARMM are transferable if used in similar chemical environments. After the partial charge optimization, the intramolecular parameters are rechecked,

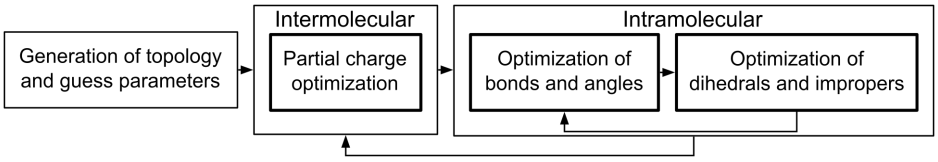


Figure 2.6: Illustration of the CGenFF algorithm. The CGenFF force field optimization is an iterative process. After obtaining starting parameters, these parameters will be repeatedly refined until the electrostatic and geometric properties of the MM compound agree with the QM target data. Optimization of Lennard–Jones interaction was not necessary for any of the complexes parametrized in here; therefore, it has been omitted from the parametrization protocol.

and the whole parametrization algorithm is repeated until self-convergence has been achieved.

New Chemical Types and CHARMM-Compatibility

Depending on the oxidation state and the system, we introduced new chemical types for iron. By doing this, none of the parameters already present in CHARMM was overwritten, and full CHARMM-compatibility was retained. The parameters presented in this dissertation are an extension to the CHARMM36 protein force field[97] and need to be used in conjunction with the protein force field and its hæm extension[118,119]. All Lennard–Jones parameters were taken from the CHARMM36 protein force field and were not modified.

QM Geometry Optimization

QM-optimized geometries of the isolated model complexes serve as geometrical target data to validate the parameters. For the ferrous FEBIHIS and FEHIS complex, we used the MP2 method. Usage of MP2 for HAEMHIS was computationally very expensive, and we performed B3LYP calculations instead. The ferric systems required the use of unrestricted QM methods, and calculations using unrestricted MP2 were not possible. Therefore, unrestricted B3LYP (UB3LYP) was used for QM calculations involving ferric complexes. In accordance with the CGenFF protocol[27], we used the 6-31G*[121,122,123] basis set in all QM computations.

Partial Charge Optimization

The aim of the optimization of the atomic partial charges was to reproduce water interaction energies E between TIP3P water and the compound. Using the example of FEBIHIS, the water interaction energy E was defined as the difference between the energy of the FEBIHIS–water complex and the sum of energies of the isolated FEBIHIS complex and a single TIP3P water molecule, as given in the following equation:

$$E = E(\text{FEBIHIS+TIP3P}) - E(\text{FEBIHIS}) - E(\text{TIP3P}). \tag{2.19}$$

The calculation of water interaction energies was done at the HF/6-31G* level of theory to retain compatibility with the rest of the CHARMM additive force fields[27]. The results of the QM calculations were compared to water interaction calculations performed using MM, and the partial charges of individual atom were adjusted until the QM target energy could be reproduced. Using higher level

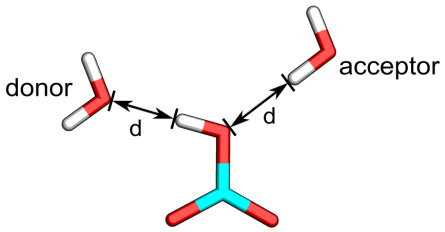


Figure 2.7: Optimization of TIP3P water interactions. Water interaction energies were calculated for donor (interaction with water oxygen) and acceptor (interaction with water hydrogen) sites. A TIP3P water molecule was placed in a planar interaction geometry. The interaction energy E was obtained by optimizing the complex, while only allowing the orientation and interaction distance d of the water molecule to change. Two water molecules are shown for illustrative purposes only as water interaction sites were sampled one at a time.

QM methods can give a better description of the water interactions[124,125], but in doing so might cause an imbalance in the treatment of non-bonded interactions in different parts of the force field.

Following the CGenFF parametrization protocol[27], compounds with zero total charge, like ferrous HAEMHIS, required the application of a scaling factor of 1.16[126] to the QM water interaction energies that were used during the derivation of atomic partial charges. With the exception of ferrous HAEMHIS, the compounds presented in here had a non-zero total charge and did not require scaling of the interaction energies.

In the first step of the optimization of the FEBIHIS water interactions, QM interaction energies were computed using the QM-optimized geometry of the FEBIHIS complex and a TIP3P water molecule. We used the Force Field Toolkit[127] to generate input files for the calculation of water interactions. To calculate a specific water interaction energy, a water molecule was constrained to an ideal linear interaction geometry with the target interaction site (Fig. 2.7). For donor interactions, the interaction distance d was the distance between water oxygen and the hydrogen atom at the interaction site; for acceptor interactions, it was the distance between water hydrogen and interacting heavy atom (Fig. 2.7).

Next, we QM-optimized the distance between the two interacting atoms as well as the orientation of the TIP3P water molecules relative to the FEBIHIS complex. This optimization gave the HF/6-31G* equilibrium distance d_{QM} and the interaction energy E_{QM} of the water molecule at that specific interaction site. This process was repeated for all accessible interaction sites, using one water molecule at a time.

The MM values for distance and energy, d_{MM} and E_{MM}, were computed using the MM atomic partial charges. The initial guesses for the partial charges were

based on Merz–Singh–Kollman (MK) partial charges[128] calculated with HF/6-31G*, with the geometry of the isolated complex optimized with MP2/6-31G* for ferrous FEBIHIS and FEHIS, B3LYP/6-31G* for ferrous HAEMHIS, and UB3LYP/6-31G* for the ferric systems.

For the MM calculations of water interaction energies at each interaction site considered during the first iteration of the CGenFF algorithm (Fig. 2.6), we again placed the respective TIP3P water molecule in an ideal interaction geometry and used the QM-optimized FEBIHIS geometry. Later iterations always used the MM-optimized FEBIHIS geometry instead.

To determine the MM interaction position d_{MM}, we constrained the orientation of the TIP3P water molecule relative to FEBIHIS to that from the QM-optimized geometry and scanned the distance between water and the FEBIHIS interaction site in steps of 0.01 Å, until the energy minimum was reached.

We manually changed the partial charges of the FEBIHIS complex and recalculated the water interaction energies. These computations were performed separately for ferrous and ferric FEBIHIS. With the exception of the non-polar hydrogens of the 4MEI moieties (Fig. 2.4), where the partial charge was kept at the standard CHARMM value for methyl hydrogen atoms of 0.09, the partial charges of the iron, bicarbonate and 4MEI atoms were optimized until E_{MM} and E_{QM} agreed to within the recommended convergence criterion of 0.2 kcal/mol[27]. Ideally, water interaction distances for the polar sites are offset by −0.2 Å[129].

The partial charge optimization of the FEHIS and the HAEMHIS systems was performed in the same manner, with the exception that, for HAEMHIS, partial atomic charges taken from the CHARMM36 protein force field were used as initial charge guesses for 4MEI and ferrous hæm.

Optimization of Bonds and Angles

We compared the equilibrium positions of the bonds and valence angles of the QM- and MM-optimized structures of HAEMHIS, FEBIHIS and FEHIS. To improve the agreement between MM- and QM-optimized geometries, we refined the bonds and angles involving iron coordination bonds. For all complexes, we optimized the iron–histidine coordination bond Fe−N_ε, and the valence angles Fe−N_ε−C_δ and Fe−N_ε−C_ε (see Figs. 2.3b, 2.4 and 2.5 for atom names).

For HAEMHIS, we additionally modified bonds between the iron and the coordinating nitrogen atoms NA, NB, NC, and ND of the hæm plane (Fig. 2.3a). We refined the equilibrium positions until the deviation between MM and QM values converged to a value less than 0.03 Å for bonds, and 3° for valence angles[27].

For the bicarbonate compound of FEBIHIS and FEHIS, the starting parameters were taken from the carbonate ion parameters of the CHARMM36 protein force field[97]. In here, we derived new bonds and angles parameters for the hydroxyl group.

For all ferric complexes, we started from the bonded parameters optimized for the corresponding ferrous complexes and then further tested and refined specific bond and angle parameters that had been previously refined for the ferrous complexes; that is, for both ferrous and ferric complexes, we refined only bonded parameters involving the iron ion and kept the remaining bonded CHARMM parameters unchanged.

Optimization of Force Constants for Bonds and Angles

After comparing the QM-optimized geometries of the complexes with the initial MM-optimized geometry, it became apparent that a number of bond and angle parameters of ferrous HAEMHIS, FEBIHIS and FEHIS required further refinement. Ideally, force constants for bonds and valence angles are optimized by analysing the vibrational modes of model compounds[27]. For large molecules, like the HAEMHIS, FEBIHIS and FEHIS complexes studied here, however, there are complications: First of all, depending on the system size and the elements of the atoms involved, the system might be too complex for the QM computation of vibrational modes to converge. Notwithstanding the above, for a molecule composed of N atoms there will be $(3N - 6)$ vibrational degrees of freedom, i.e. even if the vibrational calculations converge, there might be simply too many eigenmodes and decomposition of the frequencies will not be feasible. In situations like these, CGenFF recommends potential energy scans (PES) of small perturbations of bonds and angles[27]. We used such an approach to optimize the force constants of bonds and angles involving iron in HAEMHIS, FEBIHIS and FEHIS.

In the first step, we used the QM-optimized geometries of ferrous HAEMHIS, FEBIHIS and FEHIS to measure the equilibrium positions for bond lengths and the valence angles subject to refinement. In the second step, we performed PES separately for ferrous HAEMHIS, FEBIHIS and FEHIS. To refine force constants for bonds, we started with an offset of -0.01 Å from the equilibrium length, and conducted a three-step PES with a step size of 0.01 Å. For valence angles, we used an offset of $-1°$ and a step size of $1°$. We adjusted the MM force constants until the agreement between the QM and MM PES could no longer be improved.

The refined force constants obtained for the ferrous systems were used as starting force constants for parametrizing specific bonded force constants of ferric

HAEMHIS, FEBIHIS and FEHIS.

Optimization of Dihedrals

To derive parameters for describing the orientation of the two 4MEI groups with respect to the hæm plane, we performed a PES for rigid and flexible dihedral angles defined as follows: Rigid dihedrals were dihedrals describing ring torsions, i.e. the central two atoms of such dihedrals formed part of a ring, and the dihedral angle was associated with a relatively high energy barrier; all other dihedral angles were denoted as flexible dihedral angles.

Taking the the position of the minimized structure as starting point, we used QM to derive the PES for rigid dihedrals, once with a step size of 5° and once with a step size of −5° for six steps each. For all flexible dihedrals, we used 5° increments to perform an 18-step scan in each direction.

The PES was then computed with MM using the same step size and number of steps as in the corresponding QM PES described above. In optimizing the MM description of the dihedral angle, we aimed to reproduce the height of energy barriers and the location of energy minima. Reproducing the energy profile in the region of the energy minima is particularly important for MD simulations because these are the regions that are sampled most often[27]. We adjusted the dihedral parameters by changing their multiplicity, phase shift and force constant. Following the CHARMM protocol[27], we used only 0° and 180° for the phase shift.

For both ferric and ferrous complexes, the parametrization steps described above were performed iteratively until self-consistency had been reached.

Testing the New Parameters with Classical MD

The classical MM calculations performed utilizing the HAEMHIS parameters derived in here used the CHARMM36 protein[97,99,101] force field, the ion parameters of Roux and coworkers[130] and the TIP3P water model[105].

We used human brain neuroglobin to test the ferrous HAEMHIS parameters derived. Neuroglobin is a small soluble protein of 151 amino acid residues that contains a ferrous hæm b cofactor coordinated by two histidine residues. The starting coordinates of the protein were taken from chain B of the crystal structure (PDB ID: 1OJ6[131]).

We used the HBUILD command in CHARMM[26] to construct hydrogen atoms. The protein was placed at the center of the Cartesian coordinate system, and it was overlapped with a box of TIP3P water molecules. We deleted water molecules that were overlapping with crystal structure coordinates. The simulation setup

consisted of the protein, the hæm molecule, 42 crystal structure water molecules, 24 506 bulk water molecules and six sodium ions added for charge neutrality in a simulation box of size $90 \times 90 \times 90$ Å^3 . The system was assembled using CHARMM-GUI[132,133] Solvator[134].

To further test the impact of the optimization of partial charges, we performed independent simulations using the ferrous force field parameter sets refined for HAEMHIS in here. For the non-bonded interactions, we used a switch[135] distance of 10 Å, a cut-off distance of 12 Å and a pair list distance of 14 Å.

We used NAMD[136,137] with a Langevin dynamics scheme and a Nosé–Hoover Langevin piston[138,139] to perform simulations in an isothermal–isobaric ensemble (NpT) with isotropic cell fluctuations at a temperature of 300 K and a pressure of 1 bar, and with a Langevin damping coefficient of 5.0 ps^{-1}. We used SHAKE[140] to constrain covalent bonds involving hydrogen atoms. Heating was performed with velocity rescaling in a canonical ensemble (NVT). During heating and the first 2 ns of equilibration, we used an integration time step of 1 fs. For the rest of the simulation, we used the reversible multiple time step integration scheme[141,142] with steps of 1 fs for the bonded forces, 2 fs for short-range non-bonded and 4 fs for long-range non-bonded interactions. Simulation were prolonged to 100 ns and coordinate sets were saved every 10 ps^{-1}. Unless specified otherwise, average values and histograms were computed using 5000 equally spaced coordinate snapshots from the last 50 ns of a simulation.

2.4 | Studying Channelrhodopsin

To analyse the channelrhodopsin chimæra C1C2, MM and QM/MM calculations were performed using the CHARMM36 protein[97,99,101] and lipid[143] force fields, the ion parameters of Roux and coworkers[130] and the TIP3P water model[105]. The MM parameters used in here for retinal were based on studies performed for retinal in bacteriorhodopsin, in which parameters for partial charges[37,48], van der Waals interactions[43,44], bond lengths and valance angle[45,46] and dihedrals[47] had been derived. As QM method, we employed the SCC-DFTB method[88,89,90,91], which is known for giving a good description of retinal's geometry and its proton transfer energetics[50,51,52,144].

MM Simulations of Channelrhodopsin with All-*trans* Retinal

As starting coordinates, we used the crystal structure of the channelrhodopsin-1 and channelrhodopsin-2 chimæra C1C2 (PDB ID: 3UG9[64]). Part of the loop that

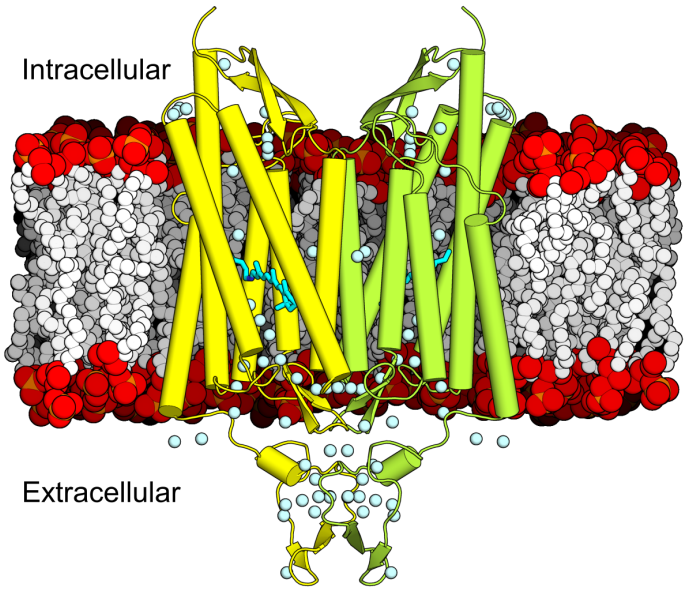


Figure 2.8: Dimeric setup of channelrhodopsin C1C2 (PDB ID: 3UG9[64]) inside a POPC membrane. The C1C2 monomers are shown in yellow (monomer 1) and lime (monomer 2) with crystal waters highlighted in light blue and retinal in cyan. In this dissertation, all illustrations of channelrhodopsin will use the same orientation.

is located between helix 1 and helix 2 was missing from the crystal structure, and we used the Phyre2 web portal[145] to model the missing amino acid residues 110–117.

We assembled the protein dimer with PISA[146] and linked the monomers with three disulfide bridges for C66, C73 and C75. We used standard protonation for all amino acid residues, except for E122, E129 and D195, which were considered protonated. In channelrhodopsin-2, E162 was thought to be deprotonated[147,148,149,150,151], and studies of C1C2 also favoured unprotonated E162[41,152]. To explore the effects of E162 protonation, we generated systems with neutral E162 and systems with negatively charged states of E162. Hydrogen atoms were constructed using HBUILD in CHARMM[26]. The protein dimer was placed inside a hydrated 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer using the CHARMM-GUI[132,133] Membrane Builder[153,154,155] (Fig. 2.8). The complete system consisted of the protein dimer, 86 waters found in the crystal structure (43 per monomer), 300 POPC lipid molecules and ∼28 500 TIP3P water molecules in a simulation box of size $110 \times 110 \times 120$ Å^3. To setups with protonated E162, we added two chloride ions to preserve charge neutrality of the system.

The MM simulations were performed using NAMD[136,137] with a Langevin dynamics scheme and a Nosé–Hoover Langevin piston[138,139] at a temperature of 300 K. We constrained covalent bonds involving hydrogen atoms using the

SHAKE algorithm[140].

The first nanosecond of heating and equilibration was done following the CHARMM-GUI protocol[132,133], while maintaining an integration time step of 1 fs and reassigning velocities every 0.5 ps. This was followed by two additional steps of equilibration of 2 ns each, where we placed on the protein and lipid heavy atoms a harmonic constraint of 5 kcal/mol/Å^2. Starting from the last 2 ns of the equilibration, we used to the reversible multiple time step integration scheme[141,142] with steps of 1 fs for the bonded, 2 fs for short-range non-bonded and 4 fs for long-range non-bonded interactions. For the first 100 ps, a canonical ensemble (NVT) was used, before switching an isothermal–isobaric ensemble (NpT) with isotropic cell fluctuations at a pressure of 1 bar. After 5 ns of equilibration, production runs were started. The Langevin damping coefficient was 1 ps^{-1} during equilibration and 5 ps^{-1} during production runs. For each of the MM simulations of all-*trans* C1C2, we generated trajectories with a length of 250 ns. Coordinate snapshots were saved every 10 ps.

Modelling and MM Simulations of Channelrhodopsin Mutants

To understand the impact selected key residues have on the hydrogen-bonding networks and interactions in the Schiff base region, we performed additional calculations with various mutants of C1C2.

For the mutant computations, we used the wild type starting coordinates for C1C2 with protonated and unprotonated E162. We performed the equilibration as described above for wild type C1C2, with the only exception being that the 5 kcal/mol/Å^2 harmonic constraint placed on the protein atoms was not relaxed. The equilibration was followed by 20 ns of simulation with the protein still constrained, after which we used CHARMM to perform the mutation in both monomers of C1C2. We then repeated the equilibration procedure using the same approach as described above for the wild type simulations, before prolonging the trajectories of the production runs to 110 ns each.

The mutants are briefly introduced below.

K132A: The K132A mutant has faster kinetics than the wild type and changes the ion selectivity of the channel from an unselective one to a potassium channel[64].

E129A: E129 protonation has been experimentally confirmed, and the residue is thought to deprotonate during the photocycle[150,156]. Theoretical studies showed that the presence of protonated E129, which forms a stable hydrogen bond with N297, is important in the closed/dark state of channelrhodopsin to prevent ion leakage[157]. E129A, like K132A, affects ion selectivity[64,158], but has a photocurrent

similar to wild type C1C2[64].

E162D: The negative counterion E162 is necessary for a fast and efficient isomerization of the retinal and stabilizes the positive charge on the C_{13}=C_{14} double bond[149]. By mutating E162 to aspartate, the missing methylene group changes the local geometry and possibly the hydrogen-bonding network in the Schiff base region, resulting in the disappearance of the K-like intermediate and faster photodynamics, but also in a tenfold decrease in photocurrent and a slower peak recovery[147,149]. E162D has the same negative Schiff base counterions as found in bacteriorhodopsin[159].

Modelling and MM Simulations of Channelrhodopsin with 13-*cis* Retinal

To further analyse the possible effects of the conformational change of all-trans retinal to 13-*cis*,15-*anti* retinal, we performed additional MM MD simulations of wild type C1C2 with a model containing isomerized retinal, using the same simulation settings as described above for the all-*trans* simulations.

As a starting point, we used the structure from the end of the all-*trans* simulations, and we applied a harmonic biasing potential to constrain the dihedral angle C_{12}−C_{13}=C_{14}−C_{15} as follows: Using a force constant of 250 kcal/mol/° and beginning at a starting value ω_{ref} of −180°, we increased ω_{ref} in steps of 10° until we reached 0°, i.e. the 13-*cis* configuration. At each step, we performed 1 ps of constrained dynamics using the constraining potential function:

$$U(\omega) = k\,(1 + \cos(\omega - \omega_{\text{ref}}))\,, \tag{2.20}$$

where ω_{ref} is the minimum of the constraint, ω is the current value of the dihedral C_{12}−C_{13}=C_{14}−C_{15}, and k is the force constant.

We generated two types of 13-*cis* setups: One, where retinal had been isomerized in one monomer only; and one, where retinal had been isomerized in both monomers. For each of these systems, we generated one setup each containing unprotonated and protonated E162, respectively. After the isomerization, we prolonged each 13-*cis* simulation to 100 ns.

QM/MM Computations of Channelrhodopsin with All-*trans* Retinal

We used the SCCDFTB module inside CHARMM to perform QM/MM calculations[112]. To prepare the all-*trans* QM/MM systems, we took one snapshot each from the end of each of the MM trajectories. In case of the wild type all-*trans* simulations, we saved the coordinates at the end of the 250 ns production run. The same protocol

was used for QM/MM simulations of the C1C2 mutants, which were started from the end of their respective MM production run at 110 ns. During the first 50 ps of QM/MM dynamics, we placed on all heavy atoms a mass-scaled harmonic positional constraint given by the constraining potential:

$$U(x) = p \sum_i \left\{ k_i m_i \left(x_i - x_{i,\mathrm{ref}} \right)^2 \right\}, \tag{2.21}$$

where, for an atom i, k_i is the force constant of the constraint in kcal/mol/Å^2, m_i is the atomic weight, x_i is the current position in Å, $x_{i,\mathrm{ref}}$ is the position of a reference set of coordinates in Å, and p is a pre-factor used to scale the constraint. The pre-factor started at a value of 1.0, and every 10 ps it was reduced by 0.25. This equilibration was followed by a production run of 1 ns.

Because on the 1 ns timescale water molecules inside the Schiff base region could be replaced by water molecules initially located farther away from the Schiff base, we constrained MM water molecules at the entrance of the Schiff base region by 2.5 kcal/mol/Å^2. This keeps QM water molecules inside the Schiff base region without having to directly apply a constraining potential to them. A similar approach has been successfully used before[148].

The QM/MM dynamics calculations were run using the Leapfrog Verlet algorithm and a canonical ensemble (NVT) with a damping constant of 5 ps^{-1} at a temperature of 300 K. We used an integration time step of 1 fs and saved coordinates every 1 ps.

QM/MM Computations of Channelrhodopsin with 13-*cis* Retinal

We performed QM/MM MD calculations of C1C2 with 13-*cis* retinal. We used two different approaches to generate 13-*cis* QM/MM systems. In the first approach, we started QM/MM MD simulations using coordinate snapshots from the end of MM simulations of 13-*cis* C1C2 and performed QM/MM MD simulations using the same setup method as for the all-*trans* QM/MM MD. In the other approach, we used the all-*trans* QM/MM simulations as starting point to isomerize the retinal and continued the MD simulations as follows: We applied a force constant k of 100 kcal/mol/rad^2 and harmonically constrained the dihedral C_{12}–C_{13}=C_{14}–C_{15} using:

$$U(\omega) = k \left(\omega - \omega_0 \right)^2, \tag{2.22}$$

where ω_0 is the minimum of the constraint, ω is the current value of the dihedral and k is the force constant. We started with a value of $-180°$ for ω_0, increased ω_0 in steps of 10° until reaching 0°. At every step, we performed 0.3 ps of

equilibration, adding up to a total of 5.7 ns. Once the 13-*cis* retinal configuration had been obtained, QM/MM MD simulations were continued using the same QM/MM protocol as described above for all-*trans* C1C2.

The Choice of Quantum Mechanical Region

Because of channelrhodopsin's dimeric structure, the QM/MM treatment could be applied to either one of the active sites in the two C1C2 monomers. We only treated with QM the active site of one of the protein monomers at a time. This was done because there is no direct interaction between the two Schiff base regions and including both sites in the QM region would have been an unnecessary increase of computational cost.

We performed independent QM/MM computations in which the QM treatment was applied either to monomer A or to monomer B of C1C2. The QM region consisted of the side chains of E129, K132, E162, D292 and K296 with the covalently bound retinal molecule (Fig. 2.9). Water molecules that were in the proximity of the Schiff base nitrogen were also included in the QM region, and depending on the setup, the amount of QM waters was 2–4. QM/MM calculations involving 13-*cis* retinal additionally included T166 and S295 as QM residues (Fig. 2.9).

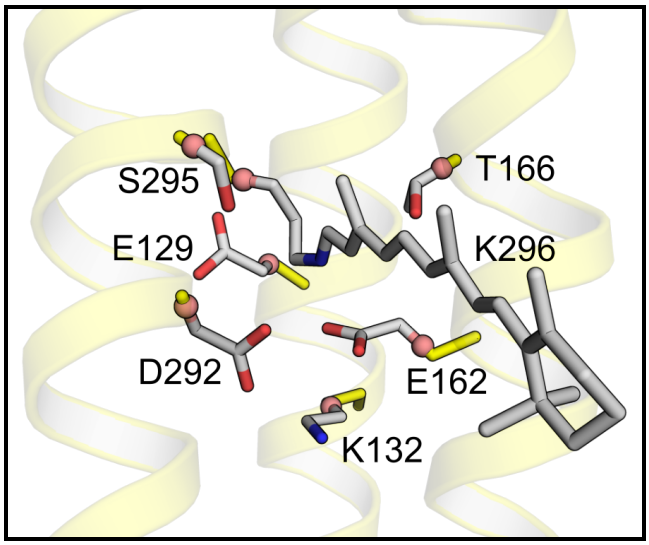


Figure 2.9: The QM Region for QM/MM analysis of Channelrhodopsin C1C2. Amino acid residues that have been part of the QM region of the QM/MM computations are shown in stick representation, and link atoms are shown as salmon-coloured spheres. Residues threonine-166 (T166) and serine-295 (S295) were only included in QM/MM calculations of channelrhodopsin that contained 13-*cis* retinal.

Link atoms were placed on C_β for aspartate (D292, E162D), on C_δ for single lysine (K132) and on C_γ for glutamate (E129, E162) and lysine–retinal (K296);

mutated alanine was not treated with QM (Fig. 2.9). We used the divided frontier link atom scheme[113]. The QM/MM computations were initiated from the end of each MM trajectory.

Proton Transfer Calculations

Proton transfer calculations were performed starting from the last coordinate snapshot of the QM/MM simulations using Alan Grossfield's implementation[160] of the Weighted Histogram Approach Method (WHAM)[161], a generalization of the histogram method[162] and an extension of the Umbrella sampling method[163,164]. WHAM has proved itself to be a reliable approach[165], and has been used in the past for retinal proton transfer calculations[166]. By using WHAM, the PMF[167] of a reaction coordinate can be derived by combining bins with different constraints placed on the reaction coordinate.

The reaction coordinate D of the proton transfer was defined as the difference between donor–donor hydrogen distance and acceptor–donor hydrogen distance:

$$D = d_{\mathrm{DH}} - d_{\mathrm{AH}}, \tag{2.23}$$

where d_{DH} is the distance between donor and donor hydrogen, and d_{AH} is the distance between donor hydrogen and acceptor. We used the restrained distances (RESD) command[168] inside CHARMM to constrain D to generate bins for sampling the deprotonation potential of mean force. The constraint is given by:

$$U(\omega) = k\,(\omega - \omega_0)^2\,, \tag{2.24}$$

where k is the constraining force constant in kcal/mol/Å^2 , and D_{ref} is the minimum of the constraint in Å. We used a force constant k of 150 kcal/mol/Å^2.

To generate starting points to begin sampling the bins for the WHAM analysis, we performed a short initial sampling as follows: Using the initial coordinates and starting from a value of $D_{\mathrm{ref},0} = -1.2$Å for an amino acid as acceptors or $D_{\mathrm{ref},0} = -1.0$Å for a water molecule as acceptor, the RESD constraint was applied; the system was equilibrated for 2.5 ps and a restart file was saved; D_{ref} was decreased by 0.1 Å and another equilibration followed. This was done until $D_{\mathrm{ref}} = -1.7$Å was reached. A second run was started from the initial coordinates; this time $D_{\mathrm{ref},0}$ was increased by 0.1 Å, followed by a 2.5 ps equilibration, until $D_{\mathrm{ref}} = 1.7$Å was reached. Once all 35 restart files had been generated, a 100 ps equilibration was started for each bin, using the respective D_{ref} value. The last 50 ps of each bin were employed for the WHAM calculations, with the values of

the reaction coordinate D being saved at every time step.

To extract factors influencing the proton transfer energetics, we ran additional WHAM analyses with reduced systems as follows:

In the first test system, denoted as the PROT setup, we started from the end of the QM/MM simulation of C1C2 and deleted all lipid molecules and all water molecules more than 5 Å away from the protein. The PMF computation on the PROT setup was then performed without additional geometry optimization. We performed 10 ps of equilibration, and then performed PMF computations using the same protocol as described above for the complete system.

In the second test system, denoted as DRY, starting from the end of the QM/MM simulation of C1C2 as well, we deleted, except for the water molecule closest to the Schiff base nitrogen, all water molecules within 4 Å of the Schiff base nitrogen or the carboxyl(ate) oxygen atoms of E162 or D292. In the DRIER setup, we deleted all water molecules within 4 Å of the Schiff base nitrogen or carboxyl(ate) oxygens without any exception. We found that deletion of QM waters in the DRY test setup led to the need for a brief energy minimization before the initial 10 ps equilibration and the PMF computation could be performed.

Finally, the crystal setup consisted of the starting crystal structure of a C1C2 monomer (PDB ID: 3UG9[64]). In the crystal structure, there is only one water molecule (w19) in the Schiff base region; the distance between the oxygen atom of w19 and the Schiff base nitrogen of 4.43 Å is too long for a direct hydrogen bond. Consequently, the PMF computations on the crystal test system considered only proton transfer pathways from the retinal Schiff base to E162 and D292.

To further test the importance of hydration for the energetics of the proton transfer pathways, we used the DOWSER plug-in[169,170] inside VMD[171] to generate possible missing water molecules in the Schiff base region, and then performed PMF computations. These test systems have been labelled as the DOWSER setups.

To preserve the shape of the protein during computations on the PROT, crystal and DOWSER test systems described above, we constrained coordinates of the heavy atoms of the loop regions and of all water molecules within 5 Å of the loop regions by using a harmonic constraint of 10 kcal/mol/Å^2 and 5 kcal/mol/Å^2 for loops and water molecules, respectively.

2.5 | Hardware and Software Used

This section contains a brief overview over the software and hardware used for molecular visualization and for performing MM and QM calculations.

Gaussian

Gaussian 09[172] was the software package used to execute the *ab initio* calculations that were necessary for the parametrization work. For all Gaussian calculations, we used default convergence criteria. Gaussian calculations were run on the soroban cluster of the FU on an Asus Dual Intel node equipped with Intel Xeon Westmere X5650 processors.(12 MB Cache, 2.66 GHz).

CHARMM

The CHARMM software package[26,95,173,174] was used for system setup and performing QM/MM calculations. CHARMM calculations were executed on the HLRN Hannover cluster on MEGWARE compute blades with Intel Xeon SandyBridge E5-4650 processors (20 MB Cache, 2.70 GHz), on the soroban cluster (same hardware as above), on the yoshi cluster of the Physics Department of the FU on HP BL460c G6 blades with Intel Xeon X5570 processors (8 MB Cache, 2.93 GHz), and on the tron cluster of the Physics Department of the FU on a Dell PowerEdge C6145 with AMD Opteron 6128HE processors (12 MB Cache, 2.00 GHz).

NAnoscale Molecular Dynamics

NAnoscale Molecular Dynamics (NAMD)[136,137,175] is a program for running MD calculations in parallel and was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. Simulations using NAMD were run on the HLRN clusters on Cray XC30 racks equipped with Intel Xeon IvyBridge E5-2695 v2 processors (30 MB Cache, 2.40 GHz).

PyMOL

PyMOL[176] is an open source molecular visualization program used to generate part of the molecular graphics in this dissertation.

Visual Molecular Dynamics

Visual Molecular Dynamics (VMD)[171,177] is a computer program for molecular visualization and modelling. Part of the molecular graphics presented in this dissertation were rendered using the Tachyon Ray Tracing system[178] inside VMD. VMD was also used to perform data analysis on simulation trajectories.

Part Two

Parametrization of Iron-Containing Cofactors

Wenn jemand sucht, dann geschieht es leicht, daß sein Auge nur noch das Ding sieht, das er sucht, daß er nichts zu finden, nichts in sich einzulassen vermag, weil er nur an das Gesuchte denkt, weil er ein Ziel hat, weil er vom Ziel besessen ist. Finden aber heißt: frei sein, offen stehen, kein Ziel haben.

Hermann Hesse, *Siddhartha*

3 Results and Discussion

Parts of this chapter and of its results have been published in modified form in the Journal of Computational Chemistry[179] and have been reproduced with permission. Copyright is held by the Journal of Computational Chemistry.

During the first part of this PhD project, new CHARMM-compatible parameters were derived for the HAEMHIS, FEBIHIS and FEHIS complexes.

The final parameter set shows very good agreement between molecular mechanics (MM)-optimized structures using the new parameters and the quantum mechanics (QM)-optimized target structures. For HAEMHIS, MM bonds and angles agreed to within 0.01 Å and 1.8° with the QM target data; for FEBIHIS, the agreement was to within 0.03 Å for bonds and to within 0.8° for angles; for FEHIS, the agreement was to to within 0.01 Å for bonds and to within 1.5° for angles (Tables 3.8, 3.9 and 3.10).

Final MM water interaction energies achieved an agreement with the HF/6-31G* values to within 0.4 kcal/mol, 0.6 kcal/mol and 1.2 kcal/mol for HAEMHIS, FEBIHIS and FEHIS, respectively (Tables 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6).

A full list of the new parameters can be found in Appendix A.

In what follows, the main observations from the derivation of the new parameter set for HAEMHIS, FEBIHIS and FEHIS will be discussed, and the results of the molecular dynamics (MD) simulations of neuroglobin using two new HAEMHIS parameter sets will be addressed.

3.1 | Water Interaction Energies and Distances

This section is based on section 'Water interactions of FeHis and HemeHis' from Adam *et al.* (2018)[179].

To derive new sets of partial charges, Hartree–Fock (HF) water interaction energies had to be computed and compared to water interaction energies derived using MM[179]. The MM atomic partial charges were adjusted until good agreement was reached between the QM and MM interaction energies. Following this procedure led to the final partial charges shown in Fig. 3.1 and Fig. 3.2 for ferrous and ferric HAEMHIS, respectively; in Fig. 3.3 for ferrous and ferric FEBIHIS; and in Fig. 3.4 for ferrous and ferric FEHIS.

For the water interaction sites sampled, the final partial charges yield MM interaction energies that agree on average to within 0.4 kcal/mol with their QM counterparts. The difference between QM and MM in water interaction distances was less than 0.04 Å–0.25 Å(Tables 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6).

We calculated water interactions for donor and acceptor sites that were accessible, and therefore, most likely to interact with water or the protein environment. For the system presented here, these interaction sites were located on histidine side chains (modelled by 4-methylimidazole (4MEI)) or bicarbonate.

For HAEMHIS, N_δ and C_ε interactions (Fig. 2.3) could not be sampled: Because of the proximity to the hæm plane and the strong interaction with it, water molecules placed at these interaction sites would be constrained to an unfavourable geometry (Fig. 3.5) and calculations for these water interactions did not converge. Partial charges of such interaction sites that could not be sampled were kept at the standard CHARMM values for hæm and 4MEI.

In FEBIHIS and FEHIS as well, the N_δ and C_ε sites could not have their water interactions sampled because of the geometry of FEBIHIS (Fig. 1.3). For FEBIHIS, partial charges were kept close to the Merz–Singh–Kollman (MK) value since there were no CHARMM reference structures and values available. Initial partial charge guesses for FEHIS were based on MK charges as well. (See Appendix A Tables A.4, A.5, A.6 and A.7 to find the MK partial charges used for the parametrization of FEBIHIS and FEHIS.)

Tables 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6 show the water interaction energies and distances computed using the final parameter sets. We note that for the C_β hydrogen interaction site, for example, only an average value is given. This was done because CHARMM requires chemically equivalent atoms to share the same partial charges. Even though, there are three interaction sites for each 4MEI

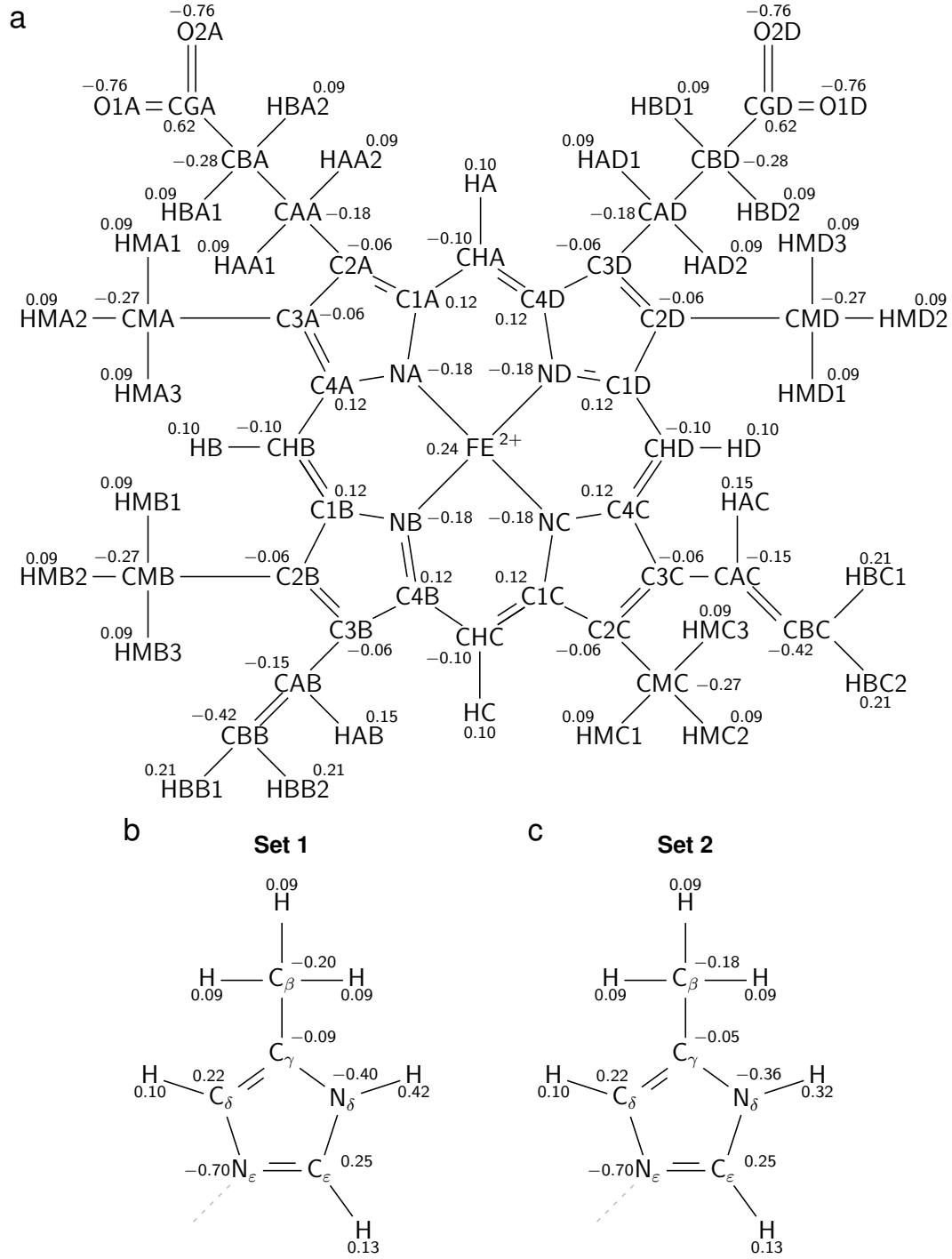


Figure 3.1: Partial charges of ferrous HAEMHIS parameter Set 1 and Set 2[179]. (a) Independent of the parameters set derived, ferrous HAEMHIS retained atomic partial charges for hæm b as found in the hæm extension[118,119] of the CHARMM36 protein force field[97] (July 2016 version). HAEMHIS parameter Set 1 uses optimized partial charges for 4MEI (b), while Set 2 uses standard CHARMM[27] 4MEI charges (c).

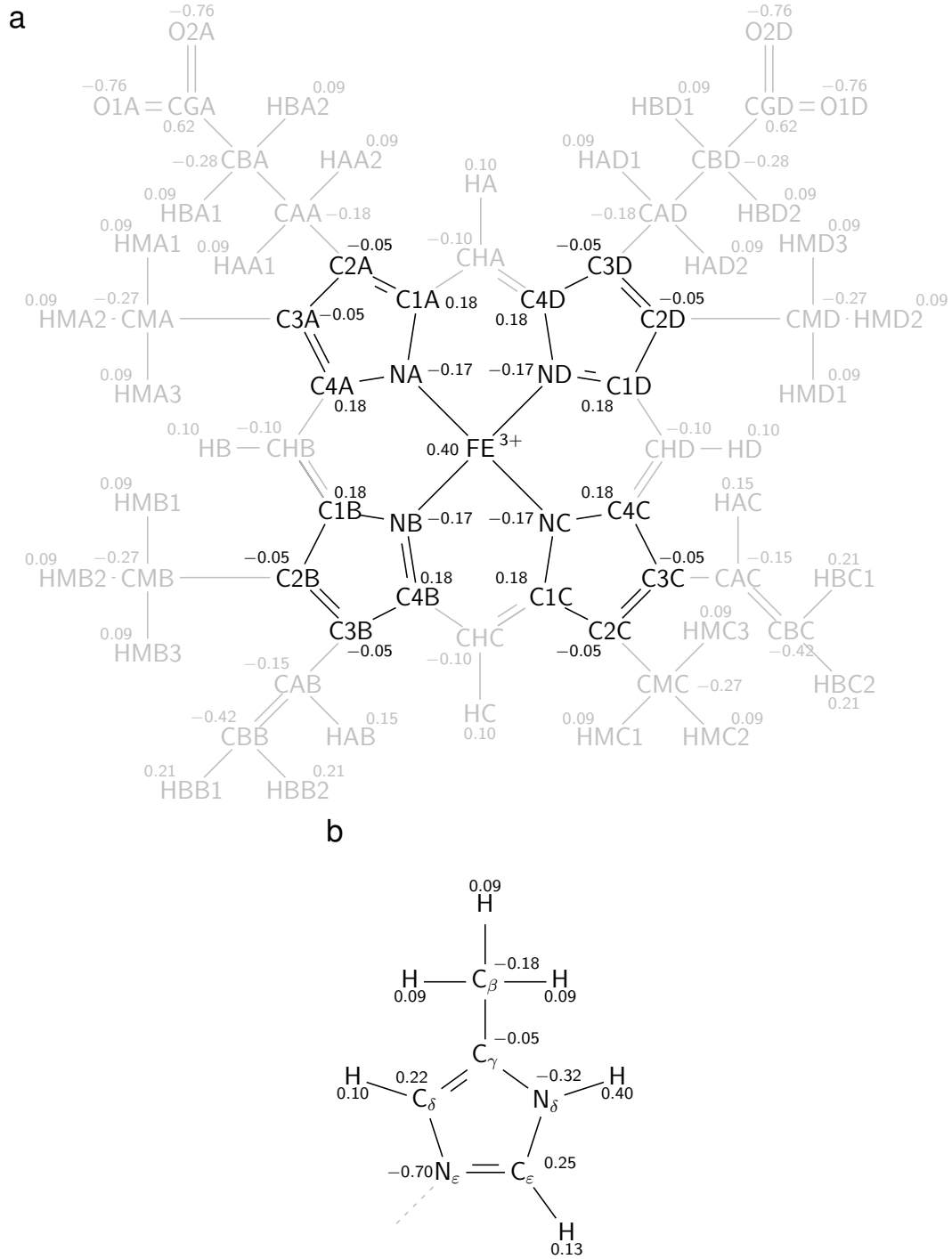


Figure 3.2: Partial charges for the ferric HAEMHIS complex[179]. Optimized atomic partial charges of hæm b (a) and 4MEI (b) are shown in black. Grey indicates partial charges and structures that use standard CHARMM36 parameters[99,118,119].

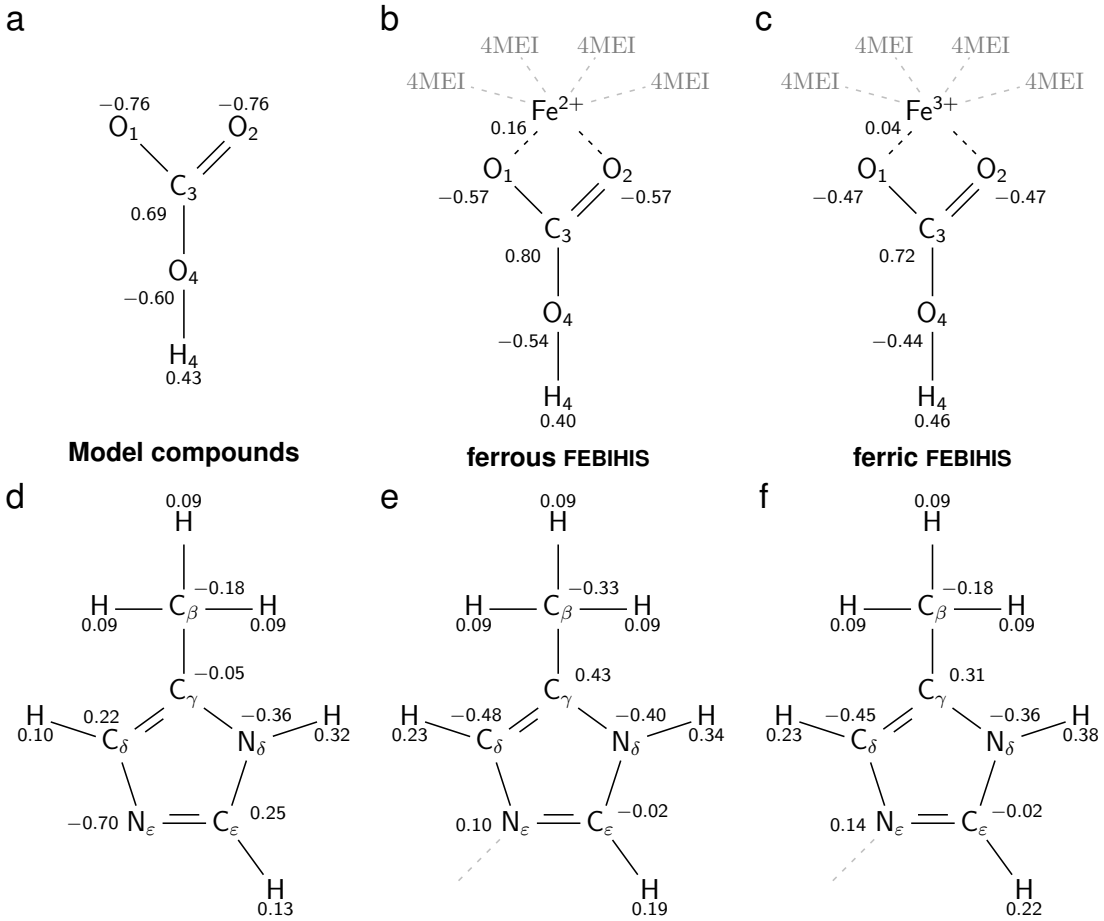


Figure 3.3: Partial charges for the FEBIHIS complex and its model compounds as found in CGenFF[179]. The FEBIHIS model complex consists of a bicarbonate ion (a) and four 4MEI (d) forming coordination bonds with a central ferrous (b) or ferric (c) non-hæm iron. Parameters for single bicarbonate and single 4MEI were already present in CGenFF[27]. On its own, 4MEI has zero total charge, but in the FEBIHIS complex part of the iron charge was spread across all 4MEI, resulting in a net charge of +0.33 on each of the 4MEI compounds in the ferrous complex (e) and of +0.54 in the ferric complex (f). The iron itself only retains an atomic partial charge of +0.16 and +0.04 in the ferrous (b) and ferric complex (c), respectively. This results in the bicarbonate compound having a net charge of −0.48 in ferrous FEBIHIS (b) and of −0.20 in ferric FEBIHIS (c).

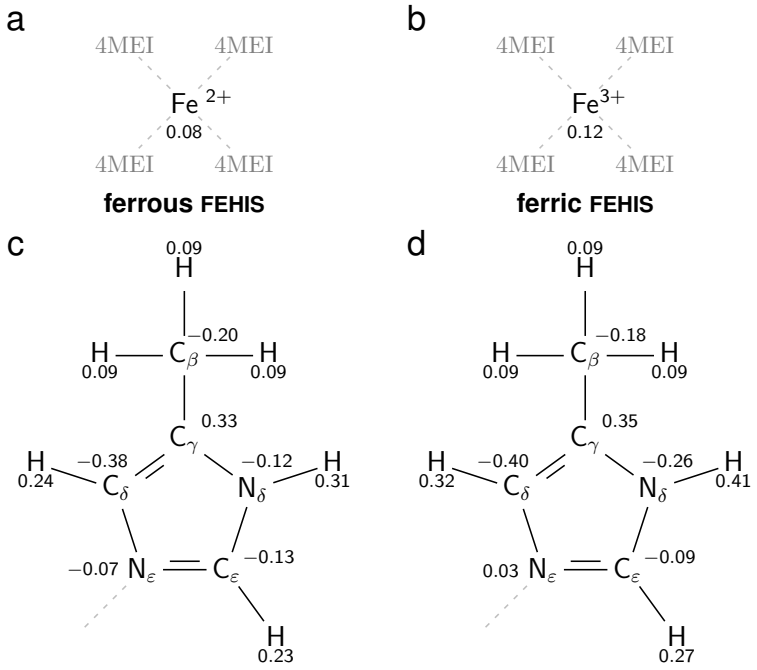


Figure 3.4: Partial charges for the ferrous and ferric FEHIS complex[179]. The final optimized partial charges are shown for the ferrous (a,c) and ferric (b,d) parameter set.

methyl group in a complex, they cannot be parametrized independently of each other, and the parametrization approach was to minimize the average of the differences between QM and MM for the these equivalent interaction sites. The individual interaction energies and distances of such equivalent interaction sites, like 4MEI methyl group water interactions in HAEMHIS, were similar (Table 3.7), further facilitating the parametrization process.

The CHARMM36 protein force field's[97] hæm extension[118,119] contains a patch for connecting a single histidine to hæm[180]. By twice applying this hæm–histidine patch, ferrous HAEMHIS can be modelled. This approach, however, keeps the standard CHARMM partial charges for hæm and histidine. Using original CHARMM partial charges gives a good description of the non-polar interactions, but the H_{N_δ} water interaction energy is underestimated by 1.88 kcal/mol (Table 3.1).

To refine the water interactions for ferrous HAEMHIS, it was unnecessary to change the partial charges of the hæm. We adjusted the partial charges on C_β, C_γ, N_δ and H_{N_δ} of 4MEI to decrease the average absolute deviation between QM and MM water interaction energies from 0.49 kcal/mol using CHARMM partial charges to 0.08 kcal/mol using our optimized partial charges. At the N_δ interaction site, the agreement with the QM values was to within 0.01 Å. This set of ferrous HAEMHIS parameters using refined atomic partial charges and bonded parameters is denoted here as Set 1.

We also derived a parameter set, where only the intramolecular parameters

Table 3.1: Water interaction energies E, in kcal/mol, and distances d, in Å, calculated for the ferrous HAEMHIS complex using optimized CHARMM parameter Set 1 and Set 2[179].

Interaction Site	E_{QM}	$E_{\mathrm{Set\ 1}}$	ΔE	$E_{\mathrm{Set\ 2}}$	ΔE
$H_{C_\beta}-OHH$	−1.35	−1.06	0.29	−1.24	0.11
$H_{N_\delta}-OHH$	−6.91	−6.89	0.01	−5.02	1.88
$H_{\text{Hæm methine}}-OHH$	−1.02	−1.24	−0.22	−1.27	−0.25
$H_{\text{Hæm methyl}}-OHH$	−0.82	−0.59	0.23	−0.61	0.20
Average deviation			0.08		0.49
Av. absolute dev.			0.19		0.61
Interaction Site	d_{QM}	$d_{\mathrm{Set\ 1}}$	Δd	$d_{\mathrm{Set\ 2}}$	Δd
$H_{C_\beta}-OHH$	2.67	2.69	0.02	2.67	0.00
$H_{N_\delta}-OHH$	2.03	2.21	0.18	2.21	0.18
$H_{\text{Hæm methine}}-OHH$	3.02	2.91	−0.11	2.91	−0.11
$H_{\text{Hæm methyl}}-OHH$	2.86	2.97	0.12	2.97	0.11
Average deviation			0.05		0.05
Av. absolute dev.			0.11		0.10

Interaction sites have been labelled according to Fig. 2.3.

were optimized and the charges were kept at their original CHARMM values. This set of ferrous HAEMHIS parameters is denoted here as Set 2.

For the ferric HAEMHIS complex, the initial partial charge guesses for 4MEI were taken from standard CHARMM[27]. As there did not exist any ferric hæm analogue in CHARMM, we calculated MK partial charges for a small model hæm, where the methyl groups had been removed (Fig. 3.6). Compared to ferrous HAEMHIS, the interaction energy at the N_δ interaction site was significantly stronger (Tables 3.1 and 3.2) and the dipole moment of N−H had to be increased

Table 3.2: Water interaction energies E, in kcal/mol, and distances d, in Å, calculated for the ferric HAEMHIS complex using optimized CHARMM parameters[179].

Interaction Site	E_{QM}	E_{MM}	ΔE	d_{QM}	d_{MM}	Δd
$H_{C_\beta}-OHH$	−3.27	−3.24	0.04	2.52	2.63	0.11
$H_{N_\delta}-OHH$	−9.95	−10.03	−0.07	1.95	2.21	0.26
$H_{\text{Hæm methine}}-OHH$	−3.79	−4.13	−0.34	2.90	2.79	−0.11
$H_{\text{Hæm methyl}}-OHH$	−2.76	−2.46	0.30	2.69	2.88	0.20
Average deviation			−0.02			0.11
Av. absolute dev.			0.19			0.17

Interaction sites have been labelled according to Fig. 2.3.

Table 3.3: Water interaction energies E, in kcal/mol, and distances d, in Å, calculated for the ferrous FEBIHIS complex using optimized CHARMM parameters[179].

Interaction Site	E_{QM}	E_{MM}	ΔE	d_{QM}	d_{MM}	Δd
H_{C_β}−OHH	−3.41	−3.01	0.40	2.54	2.68	0.14
H_{N_δ}−OHH	−9.98	−9.90	0.08	1.95	1.85	−0.10
H_{C_ϵ}−OHH	−3.49	−3.45	0.04	2.30	2.45	0.15
$O_{1/2}$−HOH	−7.42	−7.46	−0.04	1.94	1.74	−0.20
O_3−HOH	−2.91	−2.85	0.06	2.23	1.98	−0.25
H_{O_3}−OHH	−8.31	−8.07	0.24	1.86	1.82	−0.04
Average deviation			0.13			−0.05
Av. absolute dev.			0.14			0.15

Interaction sites have been labelled according to Fig. 2.4.

Table 3.4: Water interaction energies E, in kcal/mol, and distances d, in Å, calculated for the ferric FEBIHIS complex using optimized CHARMM parameters[179].

Interaction Site	E_{QM}	E_{MM}	ΔE	d_{QM}	d_{MM}	Δd
H_{C_β}−OHH	−5.88	−5.36	0.51	2.41	2.60	0.19
H_{N_δ}−OHH	−14.64	−14.51	0.14	1.88	1.80	−0.08
H_{C_ϵ}−OHH	−7.22	−7.08	0.14	2.24	2.30	0.06
$O_{1/2}$−HOH	−2.50	−2.92	−0.43	2.06	1.79	−0.28
O_3−HOH	−0.53	−0.52	0.02	2.44	2.10	−0.34
H_{O_3}−OHH	−15.16	−14.75	0.41	1.76	1.74	−0.02
Average deviation			0.13			−0.08
Av. absolute dev.			0.27			0.16

Interaction sites have been labelled according to Fig. 2.4.

to account for this change (Fig. 3.2b). In the final parameters, the hæm iron's partial charge had increased from 0.24 to 0.40, but the bulk of the additional positive charge was spread over the carbons of the pyrrole groups (Fig. 3.2a).

For FEBIHIS, the final atomic partial charges of the bicarbonate and 4MEI atoms had values very distinct from the single molecule charges as found in CHARMM General Force Field (CGenFF)[27]. Due to their proximity to the non-hæm iron, much of the iron's positive charge was distributed across 4MEI and bicarbonate.

The iron itself and the surrounding N_ε of the 4MEI groups were buried deep inside the FEBIHIS complex, preventing direct calculation of their water interaction energies, and making assessment of the MK charges of the inner FEBIHIS complex more difficult. The partial charges of the central iron and four N_ε were, therefore, kept at their starting MK charge value of +0.16 and +0.10, respectively, for the

Table 3.5: Water interaction energies E, in kcal/mol, and distances d, in Å, calculated for the ferrous FEHIS complex using optimized CHARMM parameters[179].

Interaction Site	E_{QM}	E_{MM}	ΔE	d_{QM}	d_{MM}	Δd
H_{C_β}−OHH	−5.85	−5.50	0.36	2.41	2.57	0.16
H_{N_δ}−OHH	−14.51	−14.34	0.17	1.88	1.81	−0.07
H_{C_δ}−OHH	−6.48	−6.10	0.38	2.35	2.52	0.16
H_{C_ϵ}−OHH	−6.20	−6.22	−0.02	2.28	2.21	−0.07
Average deviation			0.22			0.05
Av. absolute dev.			0.23			0.11

Interaction sites have been labelled according to Fig. 2.5.

Table 3.6: Water interaction energies E, in kcal/mol, and distances d, in Å, calculated for the ferric FEHIS complex using optimized CHARMM parameters[179].

Interaction Site	E_{QM}	E_{MM}	ΔE	d_{QM}	d_{MM}	Δd
H_{C_β}−OHH	−8.98	−7.87	1.11	2.29	2.55	0.25
H_{N_δ}−OHH	−20.41	−20.39	0.02	1.79	1.74	−0.06
H_{C_δ}−OHH	−11.71	−11.69	0.02	2.35	2.50	0.16
H_{C_ϵ}−OHH	−11.12	−11.07	0.05	2.15	2.21	0.06
Average deviation			0.30			0.10
Av. absolute dev.			0.30			0.13

Interaction sites have been labelled according to Fig. 2.5.

ferrous complex, and at +0.04 and +0.14, respectively, for the ferric complex (Fig. 3.3).

In both the ferrous and ferric systems, the charge of the iron is smaller than the +0.24 value in six-coordinated hæm in CHARMM[118]. The presence of the negatively charge bicarbonate molecule might be the explanation for this.

The difference in partial charge between standard CHARMM 4MEI and 4MEI in FEBIHIS were even greater.By itself, the unprotonated N_ε in 4MEI has a large negative charge of −0.70. In ferrous and ferric FEBIHIS, the optimized charge has a value of +0.10 and +0.14, respectively. Similar to HAEMHIS (Fig. 3.5), the geometry of FEBIHIS prevented direct sampling of the C_δ interaction site (Fig. 1.3); but by adjusting the C_δ−H_{C_δ} and N_δ−H_{N_δ}dipole moments as well as the charge on the methyl carbon, we acquired an overall good description of the ferrous and ferric water interaction energies (Tables 3.3 and 3.4). The carboxylate oxygens and the carbon of the bicarbonate have CGenFF charges of −0.76 and +0.69, respectively (Fig. 3.3a). In the final ferrous FEBIHIS parameter set, the oxygens have a charge of −0.57 and the carbon has a charge of +0.80; for

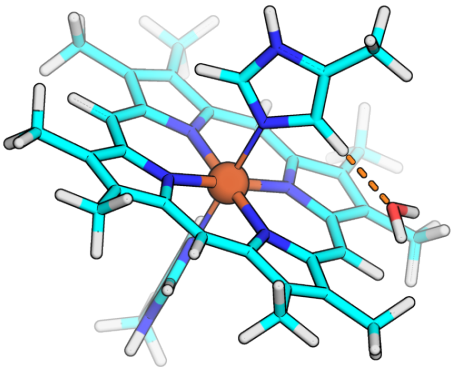


Figure 3.5: N_δ water interaction site of HAEMHIS that was not sampled. To calculate the water interaction energy at the N_δ site of HAEMHIS, the standard CGenFF approach was used and a TIP3P water molecule was constrained to a linear interaction geometry. The only degrees of freedom were the distance between water oxygen and H_{N_δ} and the orientation of the water molecule. Because of these constraints, the water molecule remained trapped in an unfavourable position, and the QM calculation was prevented from reaching convergence.

Table 3.7: Water interaction energies E, in kcal/mol, and distances d, in Å, calculated for the methyl groups of 4MEI in HAEMHIS using parameter Set 1.

Interaction Site	E_{QM}	E_{MM}	ΔE	d_{QM}	d_{MM}	Δd
$H_{C_\beta}(1)$−OHH	−1.43	−1.12	0.31	2.70	2.70	0.00
$H_{C_\beta}(2)$−OHH	−1.21	−0.96	0.24	2.59	2.67	0.08
$H_{C_\beta}(3)$−OHH	−1.42	−1.10	0.31	2.70	2.70	0.00
$H_{C_\beta}(4)$−OHH	−1.21	−0.96	0.24	2.59	2.67	0.08
$H_{C_\beta}(5)$−OHH	−1.43	−1.12	0.32	2.70	2.70	0.00
$H_{C_\beta}(6)$−OHH	−1.41	−1.10	0.31	2.70	2.70	0.00
Average	−1.35	−1.06	0.29	2.67	2.69	0.02

Interaction sites have been labelled according to Fig. 2.3.

the ferric complex, the charges become −0.47 for the oxygens and +0.72 for the carbon. These differences between bound and unbound bicarbonate are explained by the electrostatic interactions of the bicarbonate compound with other groups of the FEBIHIS system, especially with the positively charged non-hæm iron.

For FEHIS, the MK charges predicted a higher positive charge on the non-hæm iron, namely +0.60 and +0.36 for ferrous and ferric FEHIS, respectively. In the case of ferrous FEHIS, one could argue that the positive charge taken up by the bicarbonate moiety remained on iron; whereas in the ferric system the excess bicarbonate charge was spread over the whole complex. The large charge difference between iron and N_ε, and the resulting strong dipole moment caused problems, when optimizing water interactions and intramolecular parameters. Consequently, the charge of the central iron was not kept at the MK value, but

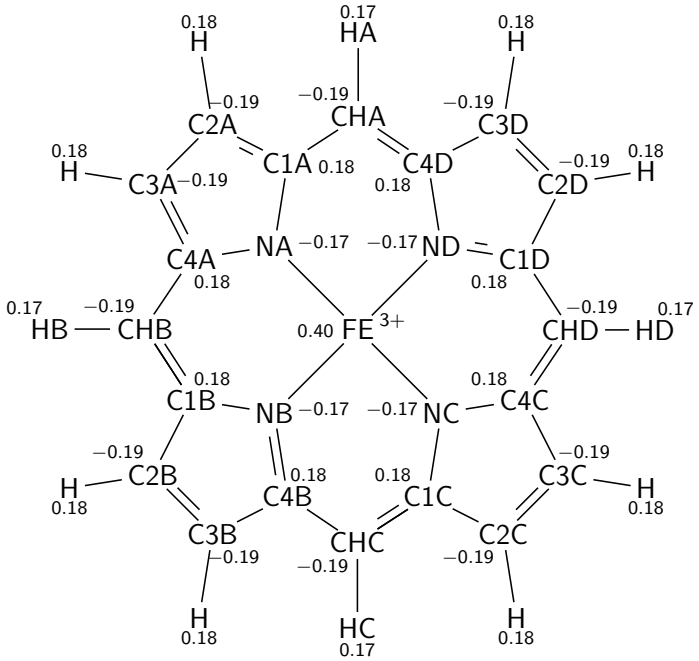


Figure 3.6: Adjusted MK partial charges for model hæm without any methyl groups. MK charges were calculated for a reduced hæm model system consisting of the hæm core without its methyl groups. To be usable as CHARMM guess charges, the MK charges were averaged for atoms that had to share the same CHARMM charge, e.g. NA, NB, NC and ND originally had MK charges of −0.18, −0.16, −0.17 and −0.18, respectively. The charges of the hydrogens of the pyrrolic groups were later summed into their adjacent carbon atom. The iron charge was increase by 0.02 to retain an integer total charge for the molecule.

instead distributed to other atoms of the FEHIS complex, until a satisfactory description of the water interactions was achieved (Tables 3.5 and 3.6).

Note, that CHARMM force fields tend to underestimate water interaction energies for methyl groups[27]. The methyl group interactions presented here for the final partial charge sets agree with this observation.

3.2 | Optimization of Bonds and Angles

This section is based on section 'Bond and angle parameters for HemeHis and FeHis' from Adam *et al.* (2018)[179].

New bond and angle parameters will be presented in this section. For HAEMHIS, the CHARMM36 protein force field with hæm extension[97,118] was the source for the initial guesses for the bonded parameters. For both ferrous and ferric system, the initial bond parameters underestimated the length of the iron–nitrogen bonds in the hæm plane. The bond Fe−$N_{Hæm}$ had a length of 2.01 Å and 2.00 Å in the QM-optimized structures of the ferrous and ferric FEHIS complex, respectively; in

the MM-optimized structures, the initial length was 1.96 Å. When using standard CHARMM parameters, the strength of the interaction between iron and N_ε seemed to be underestimated, as the Fe−N_ε distance, 2.01 Å (ferrous FEHIS) and 1.99 Å (ferric FEHIS) in QM, was 2.07 Å–2.08 Å in MM. After the optimization of the bonds parameters, the MM length of all iron-involving bonds agreed to within 0.01 Å with the respective QM value (Table 3.8).

For FEBIHIS, the initial guess for the parameters for the coordination bond between non-hæm iron and the carboxylate oxygens O_1 and O_2 of the bicarbonate (Fig. 2.4) was taken from the bond between hæm iron and carbon dioxide found in the hæm extension to the CHARMM36 protein force field[180]. For both FEBIHIS and FEHIS, the bond between non-hæm iron and N_ε of the 4MEI groups was taken from the CHARMM patch connecting hæm to a single histidine[180].

When comparing MM geometries of ferrous and ferric structures of FEBIHIS and FEHIS to QM-optimized structures, it became apparent that the length of the coordination bond between iron and the coordinating nitrogen atoms was being overestimated. In ferrous FEBIHIS, the initial MM parameters gave a bond length of 2.23 Å, as compared to the 1.98 Å value calculated with QM. In ferric FEBIHIS, the starting parameters yielded a bond length of 2.24 Å, and not the QM value of 1.99 Å. The Fe−N_ε separation in the crystal structure ranged from 2.10 Å–2.29 Å (PDB ID: 3WU2[9]). Applying the standard CHARMM parameters to FEHIS showed the same behaviour as in FEBIHIS, with the Fe−N_ε distance for the ferrous and ferric complexes being 2.26 Å and 2.27 Å, respectively, in MM, and 1.94 Å and 1.90 Å, respectively, in QM.

To optimize the force constant for the coordination bond between iron and

Table 3.8: Comparison of QM and MM geometries of the ferrous and ferric HAEMHIS complex, using the initial guess and final optimized force field parameters[179].

	HAEMHIS (Fe^{2+})			HAEMHIS (Fe^{3+})		
Parameter	QM	$MM_{initial}$	MM_{final}	QM	$MM_{initial}$	MM_{final}
	Bonds (Å)			Bonds (Å)		
Fe−$N_{Hæm}$	2.01	1.96	2.01	2.00	1.96	2.00
Fe−N_ε	2.01	2.08	2.01	1.99	2.07	1.99
	Angles (°)			Angles (°)		
Fe−N_ε−C_δ	128	131	128	127	132	128
Fe−N_ε−C_ε	126	121	127	126	123	128

Average bonds and angles are given for: 4MEI; and the hæm plane nitrogens NA, NB, NC, and ND (denoted as $N_{Hæm}$). Atoms have been labelled according to Fig. 2.3.

Table 3.9: Comparison of QM and MM geometries of the ferrous and ferric FEBIHIS complex, using the initial guess and final optimized force field parameters[179].

	FEBIHIS (Fe^{2+})			FEBIHIS (Fe^{3+})		
Parameter	QM	$MM_{initial}$	MM_{final}	QM	$MM_{initial}$	MM_{final}
	Bonds (Å)			Bonds (Å)		
$Fe-N_{\varepsilon}$	1.98	2.23	1.98	1.99	2.24	1.99
$Fe-O_{1/2}$	2.04	1.72	2.04	1.96	1.72	1.97
$C_3-O_{1/2}$	1.27	1.30	1.26	1.28	1.30	1.25
C_3-O_3	1.35	1.33	1.35	1.31	1.33	1.34
	Angles (°)			Angles (°)		
$Fe-N_{\varepsilon}-C_{\delta}$	129	136	129	128	134	128
$Fe-N_{\varepsilon}-C_{\varepsilon}$	125	119	125	125	121	125
$Fe-O_{1/2}-C_3$	86	77	86	88	78	88
O_1-Fe-O_2	66	83	66	67	83	66
$C_3-O_3-H_{O_3}$	106	106	106	109	110	108

Average bonds and angles are given for: 4MEI, and the O_1- and O_2-involving bonds and angles (denoted as $O_1/2$). Atoms have been labelled according to Fig. 2.4.

N_{ε} in all systems, we employed potential energy scan (PES) computations (Fig. 3.7a). The initial CHARMM force constant of 65 kcal/mol/Å^2 was found to be too weak to correctly reproduce the profile of the PES, and it had to be raised to 80 kcal/mol/Å^2 for HAEMHIS and to 140 kcal/mol/Å^2 for FEBIHIS and FEHIS (Table 3.11).

Force constants optimized for the ferrous systems gave an adequate description of the ferric systems, and the ferric force constants were kept at the value optimized for the ferrous structures. (See Appendix A.6 for more information.)

Table 3.10: Comparison of QM and MM geometries of the ferrous and ferric FEHIS complex, using the initial guess and final optimized force field parameters[179].

	FEHIS (Fe^{2+})			FEHIS (Fe^{3+})		
Parameter	QM	$MM_{initial}$	MM_{final}	QM	$MM_{initial}$	MM_{final}
	Bonds (Å)			Bonds (Å)		
$Fe-N_{\varepsilon}$	1.94	2.26	1.95	1.90	2.27	1.90
	Angles (°)			Angles (°)		
$Fe-N_{\varepsilon}-C_{\delta}$	126	133	126	128	132	129
$Fe-N_{\varepsilon}-C_{\varepsilon}$	127	123	127	125	124	127

Average bonds and angles are given for 4MEI. Atoms have been labelled according to Fig. 2.5.

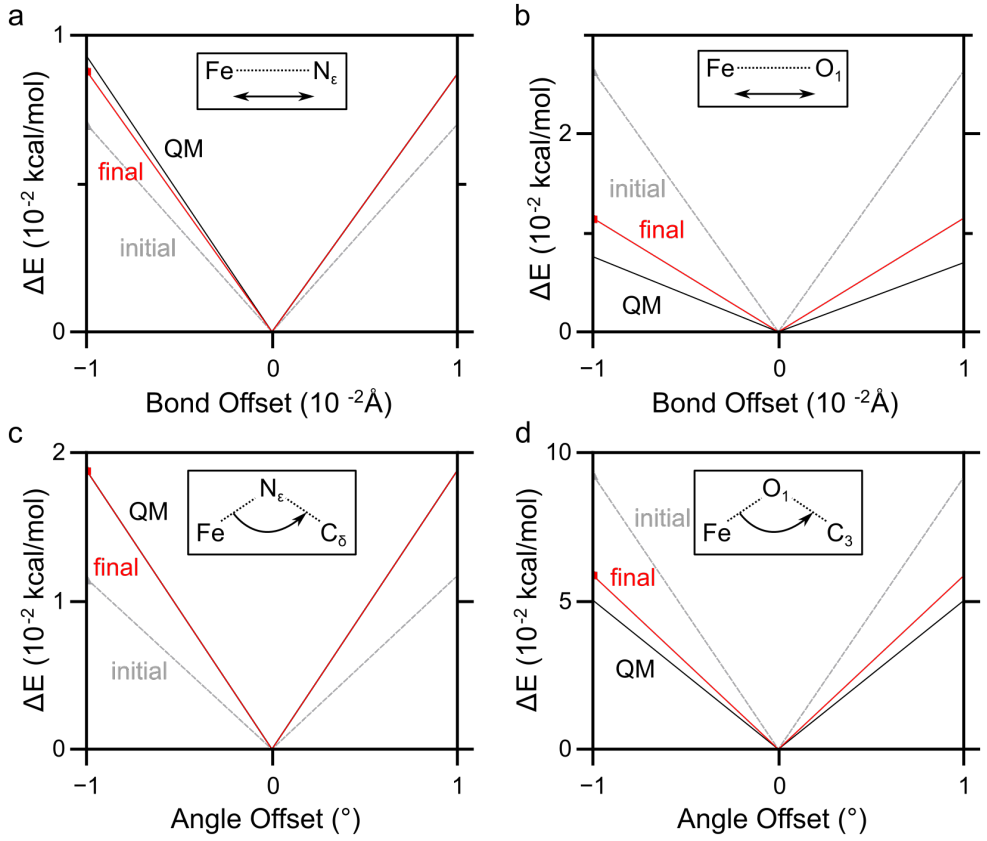


Figure 3.7: Three-point PES to refine force constants for bonds and valence angles[179]. (a) Optimization of the force constant of the iron–4MEI bond in ferrous HAEMHIS. Initial MM results are shown as gray dashes. The final optimized MM energy differences and the target QM data are shown as black and red lines, respectively. (b) PES for the bond between bicarbonate and non-heme iron in ferrous FEBIHIS. (c) PES for the iron–4MEI angle in ferrous HAEMHIS. (d) PES for the angle involving the non-heme iron and bicarbonate in ferrous FEBIHIS.

For the FEBIHIS complex, we additionally optimized the bond force constant of the coordination bond between non-hæm iron and bicarbonate carboxylate oxygens. Using the starting parameters in ferrous FEBIHIS, the PES analysis of the Fe−$O_{1/2}$ bond indicated that the energy required for stretching the bond by −0.01 Å and 0.01 Å was overestimated by ∼165 % and ∼260 %, respectively. Reducing the force constant for the Fe−$O_{1/2}$ bond from 250 kcal/mol/Å^2 to 100 kcal/mol/Å^2 (Table 3.11) improved the agreement between QM and MM bond stretching energies to within 3×10^{-3} kcal/mol (Fig. 3.7b). In the same way, the PES, for bending by 1° the valence angle involving non-hæm iron and the bicarbonate's carboxylate oxygen atoms, indicated an overestimation by ∼180 % (Fig. 3.7d), and the associated force constants was reduced by 5 kcal/mol/rad^2 (Table 3.11).

The final parameter sets for HAEMHIS, FEBIHIS and FEHIS gave good agreement

Table 3.11: Comparison of the initial guess, taken from standard CHARMM, for force constants of specific bonded parameters with values optimized here[179].

Parameter	Initial	HAEMHIS	FEBIHIS	FEHIS
		Bonds (kcal/mol/Å^2)		
Fe−N_ε	65	80	140	140
Fe−$O_{1/2}$	250	—	100	100
		Angles (kcal/mol/rad^2)		
Fe−N_ε−C_δ	30	25	25	25
Fe−N_ε−C_ε	30	25	25	25

Atoms have been labelled according to Fig. 2.5.

between structures optimized with MM and QM-optimized structures. For ferrous HAEMHIS, the QM optimization was performed with Becke3,Lee–Yang–Parr functional (B3LYP); for ferrous FEBIHIS and FEHIS, it was done with Møller–Plesset perturbation theory (MP2); and for the ferric complexes, UB3LYP was used (Tables 3.9,3.10 and 3.8). For ferrous and ferric HAEMHIS, the Fe−N_ε and Fe−$N_{\text{Hæm}}$ bonds ultimately had the same length in QM and MM, and MM angles containing the N_ε coordination bond agreed to within 1° with the QM target data. Ferrous HAEMHIS parameter Set 1 and Set 2 produce the same geometry and therefore have not been depicted separately (Table 3.8). In ferrous and ferric FEBIHIS, MM valence angles containing the Fe−N_ε coordination bond fully agree with their QM counterparts (Table 3.9). We note, however, that the Fe−N_ε coordination bond is shorter in the QM-optimized model systems than it is in the crystal structure (PDB ID: 3WU2[9]). In FEHIS, the iron–histidine coordination bond was shorter than in FEBIHIS, and after the optimization of the FEHIS parameters, the bonds and valence angles of both ferrous and ferric system could be correctly reproduced (Table 3.10).

3.3 | Improvement of Dihedral Angles

This section is based on section 'Optimization of selected dihedral angles for HemeHis and FeHis' from Adam *et al.* (2018)[179].

Dihedral PES were performed to optimize newly introduced dihedral parameters. In HAEMHIS, the dihedral angle N_A−Fe−N_ε−C_ε describing the hæm–histidine torsion had to be optimized (Fig. 3.8a–b). In FEBIHIS, two new dihedrals were introduced: $O_{1/2}$−C_3−O_4−H_{O_4} and Fe−$O_{1/2}$−C_3−O_4 (Fig. 3.8c–f). No new

dihedral parameters had to be introduced for FEHIS.

To derive parameters for these dihedral angles, we used QM to calculate the energy required to twist each one of these three dihedral angles. As reference value, we used the energy of the geometry-optimized FEHIS structure without any constraints. Starting from this reference structure, for the flexible dihedrals N_A−Fe−N_ε−C_ε and $O_{1/2}$−C_3−O_4−H_{O_4}, we performed PES using a step size of 5°, with maximum twists of 90° (clockwise twist) and −90° (anticlockwise twist). For the rigid dihedral Fe−$O_{1/2}$−C_3−O_4, the PES was performed up to a maximum twist of ±30°. We then used the QM PES of each dihedral angle to manually adjust the CHARMM dihedral angle parameters until the MM PES could not be further refined. This procedure was applied to the ferrous and ferric structures.

To compute starting MM PES for the hæm–histidine torsion in ferrous and ferric HAEMHIS, we used the parameters from the CHARMM hæm patch[180]. Comparing the starting MM profile with the target QM data showed that the energy barrier present at the 45°-twisted geometry was underestimated by CHARMM. In the ferrous complex, the QM and MM barriers were 1.21 kcal/mol and 0.77 kcal/mol, respectively (Fig. 3.8a). Furthermore, the local minimum of the twisted MM geometry was situated at 80°, as compared to 90° in the QM profile, and, with values of 0.49 kcal/mol and 0.35 kcal/mol for MM and QM, respectively, it was energetically less favourable than QM predicted. In Set 1, we increased the force constant of the dihedral parameter term with multiplicity 4 from 0.05 kcal/mol/rad^2 to 0.14 kcal/mol/rad^2 and introduced two additional multiplicities (force constant 0.07 kcal/mol/rad^2, multiplicity 2; force constant 0.04 kcal/mol/rad^2, multiplicity 3) to improve the agreement between QM and MM and to be able to correctly reproduce the QM barrier height and the energy difference between the 0° and ±90° structures (Fig. 3.8a). For Set 2, increasing the force constant of the multiplicity 4 term to 0.19 kcal/mol/rad^2 was sufficient to reproduce the QM profile.

For N_A−Fe−N_ε−C_ε in the ferric complex, QM target data could only be obtained for a PES in one direction, i.e. from 0°–−90°. The stronger attraction between ferric hæm and 4MEI caused the QM barrier and the energy difference between the 0° and the ±90°-twisted geometry to increase to 1.92 kcal/mol and 0.58 kcal/mol, respectively. The optimized parameters show very good agreement between QM and MM barrier height and energy difference, with the MM values being 1.93 kcal/mol and 0.54 kcal/mol, respectively (Fig. 3.8b).

In FEBIHIS, the dihedral angle $O_{1/2}$−C_3−O_4−H_{O_4} describes the orientation of the hydroxyl group of the bicarbonate and the rigid dihedral Fe−$O_{1/2}$−C_3−O_4 describes the rotation about the coordination bond between carbon and carboxylate

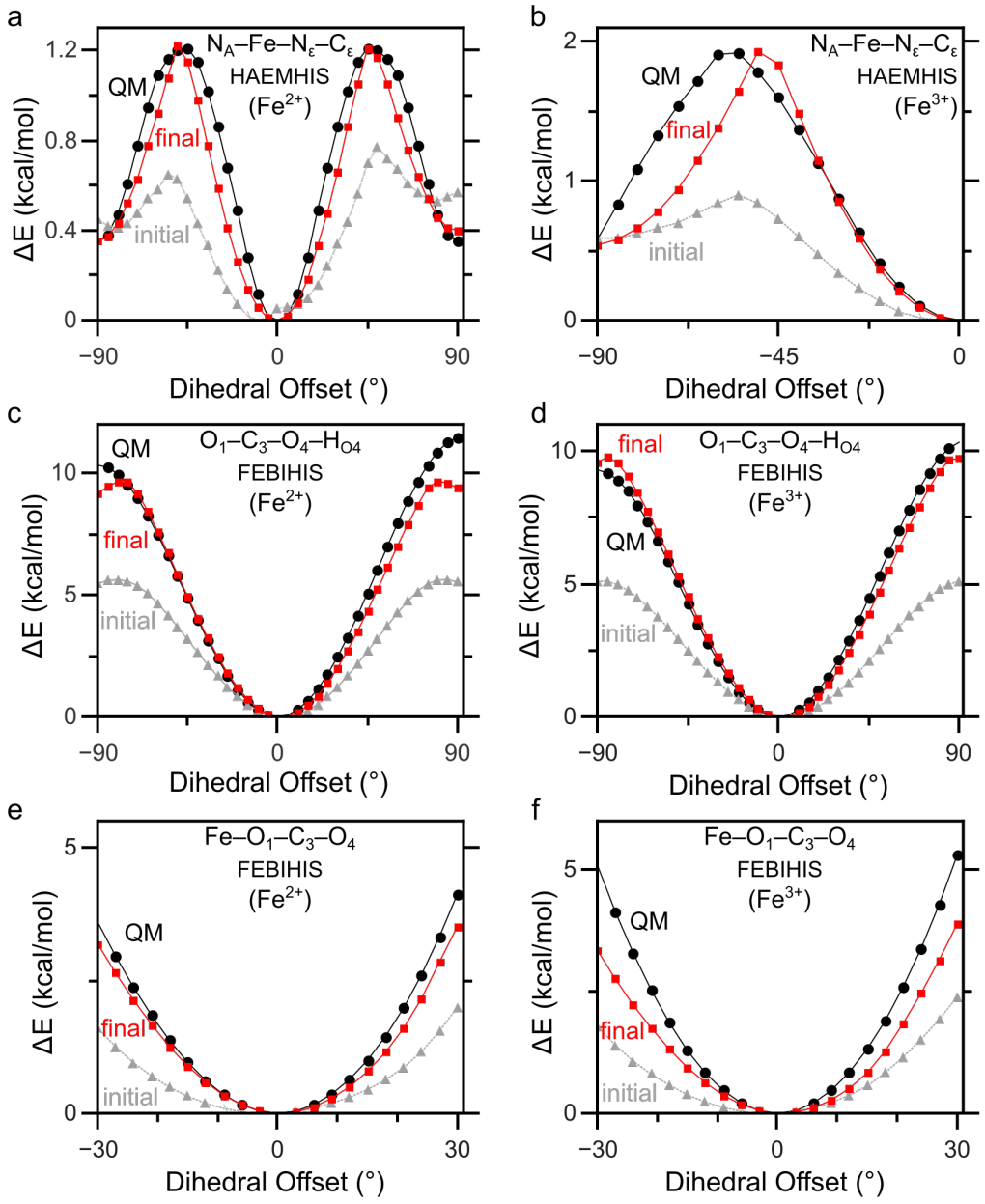


Figure 3.8: Energy profiles for PES computed with QM (black, circles), initial CHARMM parameters (grey, triangles) and with the final optimized parameters sets (red, squares)[179]. The dihedral offset gives the dihedral angle relative to its value in the QM-minimized structure. For all dihedrals, initial CHARMM parameters produced an MM PES profile that was too shallow, while the final parameters shows a marked improvement. (a) For $N_A-Fe-N_\varepsilon-C_\varepsilon$ in ferrous HAEMHIS, the location and height of the barriers could be correctly reproduced. (b) In ferric HAEMHIS, a PES of $N_A-Fe-N_\varepsilon-C_\varepsilon$ was only possible from 0°–−90°. (c, d) Torsion of the hydroxyl group of bicarbonate in ferrous (c) and ferric (d) FEBIHIS. The barrier heights of QM and MM agreed to within 0.6 kcal/mol in the ferric system. (e, f) Because of constraints imposed by the geometry of the FEBIHIS cluster, a full 180°-torsion about the O_1-C_3 bond was not possible. A dihedral parameter with a multiplicity of 2 was used to describe this rigid dihedral.

oxygens of the bicarbonate.

For $O_{1/2}-C_3-O_4-H_{O_4}$, the description of the energetics close to the minimum position of the bicarbonate as well as the overall overall PES profile could be greatly improved for both ferrous and ferric FEBIHIS. In the ferrous system, for relatively minor twists of $\leq 10°$, which cost $\sim$1 kcal/mol and can be easily sampled at room temperature, the initial guess based on standard CHARMM parameters gave a good description of the QM profile (Fig. 6a). Larger twists of $\sim 45°$ required $\sim$3 kcal/mol with the guess parameters, as compared to $\sim$5 kcal/mol with QM or the final parameter set presented here (Fig. 3.8c). Likewise, compared to the QM PES, torsional freedom of the dihedral $Fe-O_{1/2}-C_3-O_4$ in ferrous FEBIHIS was overestimated, when using standard CHARMM parameters. Employing the revised bonded and non-bonded parameters derived here markedly improved the torsional energy profile (Fig. 3.8e).

For the dihedral $O_{1/2}-C_3-O_4-H_{O_4}$ in ferric FEBIHIS, the torsional MM PES initially underestimated the QM barrier height of 9.29 kcal/mol and 10.33 kcal/mol at $-90°$ and $90°$ by $\sim$4 kcal/mol–5 kcal/mol. The refined parameters significantly enhanced the agreement with QM and yielded final barrier heights of 9.78 kcal/mol and 9.78 kcal/mol for $-90°$ and $90°$, respectively (Fig. 3.8d). The PES profile of ferric $Fe-O_{1/2}-C_3-O_4$ was markedly improved as well (Fig. 3.8f).

We note that using the same chemical types for all four 4MEI groups of FEBIHIS and FEHIS limits the ability of the force field to account for asymmetry in the orientation of histidine side chains in the ferrous and ferric FEBIHIS complexes as observed in the QM-optimized structures. More precisely, because of the usage of chemical types, the same dihedral parameter would be applied to all dihedrals $N_\varepsilon-Fe-N_\varepsilon-C_\varepsilon$ independent of whether they described 4MEI groups on opposite side of the central non-hæm iron or 4MEI groups next to each other. For this reason, the geometry of the FEBIHIS and FEHIS complexes (Figs. 1.3, 2.4 and 2.5) makes it not feasible to define the dihedral $N_\varepsilon-Fe-N_\varepsilon-C_\varepsilon$ using a single dihedral parameter, and we set the force constant of this parameter to 0. The same holds true for $N_\varepsilon-Fe-N_\varepsilon-C_\delta$.

3.4 | Reliability of the Parametrization

Assessing the quality of parameters is a non-trivial task[181]. The target data used for the parametrization has to be derived for model systems in ideal configurations, and water interactions are only calculated for ideal interaction geometries at specific sites. Even dihedral energy scans can be a not ideal solution because they

use only a single reaction coordinate to describe dihedral torsion. Neither can aiming for the most expensive method available guarantee perfect target data as, for example, lack of dispersion in B3LYP or large basis set superposition errors in MP2 can yield dihedral PES profiles that miss local minima[182]. Automated approaches that seek to derive parameters and assess their robustness have been examined for CHARMM[183,184,185], but in the end, one always has to ensure that the QM method chosen was sensible and that good care was taken during the fitting process.

To further evaluate the quality of our parameters and the choice of QM methods, we chose to perform additional tests of the water interactions and dihedral properties in FEBIHIS, and we performed an MD simulation of neuroglobin using the new HAEMHIS parameters.

Additional Water Scans for FEBIHIS

As already mentioned, no van der Waals, i.e. non-bonded, parameters were parametrized because no experimental data were available and the systems were too large for high-level QM calculations. The quality of the intermolecular interactions was improved during the optimization of the water interactions; however, away from the minimum position of a specific water interaction site, the quality of the energies had a great dependence on the choice of CHARMM chemical type, and thus the non-bonded parameters applied.

For the H_{N_δ} interaction site in ferrous FEBIHIS, we performed both QM and

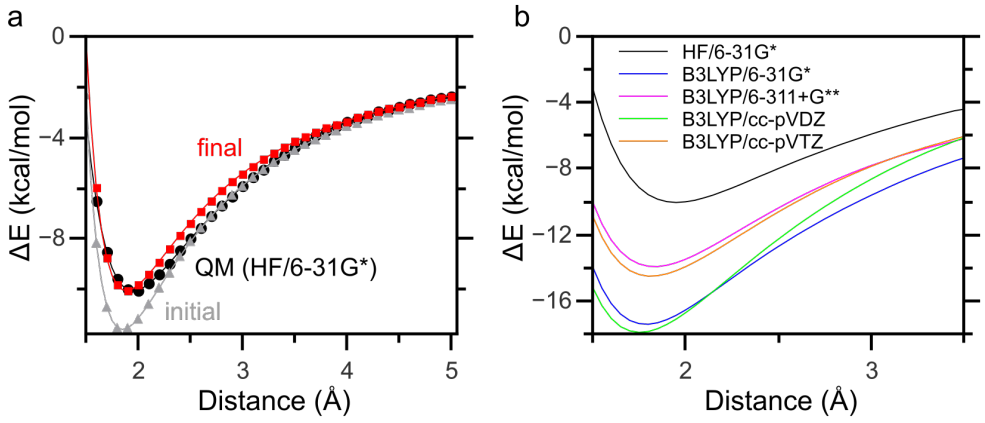


Figure 3.9: Water interaction energies at the H_{N_δ} interaction site in ferrous FEBIHIS[179]. (a) Comparison of QM and MM water interaction energies. The QM target data used during the parametrization was computed using HF/6-31G* (black, circles). For MM, the results of the initial guess parameters (grey, triangles) and the optimized parameters (red, squares) are shown. (b) Comparison of the HF/6-31G* energy profile to B3LYP water interactions calculated with different basis sets.

MM water interaction energy scans in the range 1.5 Å–5.0 Å (Fig. 3.9a)[179]. When moving the TIP3P water molecule away from the interaction site, then, no matter whether the unparametrized CHARMM guesses or the optimized parameters were used, the MM energies always ended up approaching the HF energies. This highlighted the importance of the non-bonded parameters, while at the same time confirming that (i) non-bonded parameters of chemically similar atoms are more transferable than other force field parameters and that (ii) our choice of chemical types was the correct one.

We furthermore calculated water interaction energies for the B3LYP method using 6-31G* and the larger basis sets 6-311+G**[186,187,188], cc-pVDZ[189] and cc-pVTZ[189] (Fig. 3.9b). In general, the B3LYP energies give a stronger interaction energy with the minimum energy being −17.42 kcal/mol, −13.93 kcal/mol, −17.91 kcal/mol and −14.51 kcal/mol for 6-31G*, 6-311+G**, cc-pVDZ and cc-pVTZ, respectively; as opposed to the HF/6-31G* interaction energy of −10.05 kcal/mol. The water interaction distance, as well, varies between 1.75 Å and 1.85 Å. The reasons for the differences in behaviour can be various (e.g. the added correlation in going from cc-pVDZ to cc-pVTZ) and will not be further discussed as the point of the calculations was to show that the CHARMM force field had been optimized for HF water interactions and that using any other target QM data for the refinement of the water interaction energies might result in compatibility issues with already existing parameters. In the case of the H_{N_δ} interaction site in ferrous FEBIHIS, usage of B3LYP with the basis sets shown in Fig. 3.9b would have created a much more polarized molecule, which would have biased the interactions of the FEBIHIS complex with its surrounding molecules in CHARMM.

Testing B3LYP for Dihedral Potential Energy Scans in FEBIHIS

When it comes to the QM method used for the HAEMHIS system, B3LYP had been used in the past to derive parameters for hæm coordinating histidine and cysteine[190]. The choice of this method was therefore deemed reliable. For FEBIHIS and FEHIS, no similar parametrization could be found and MP2 was chosen for the ferrous systems.

At the MP2/6-31G* level, performing the initial energy minimization of ferrous FEBIHIS to derive the QM target structure took 1.2 d using 12 CPUs and required 27 SCF cycles, i.e. ∼3800 s/CPU/cycle (Table 3.12). The energy minimization done for the ferric FEBIHIS complex started from QM-optimized ferrous FEBIHIS structure, thereby reducing the number of SCF cycles necessary

Table 3.12: Benchmarks of calculations for minimizing the ferrous and ferric FEBIHIS structures with different QM methods using the 6-31G* basis set.

System	Method	CPUs*	Cycles†	Wall‡ (h)	Efficiency (s/CPU/cycle)
Ferrous FEBIHIS	MP2	12	27	28.76	3834.27
Ferric FEBIHIS	B3LYP	16	15	1.72	412.18
	MP2	32	21	213.75	36 643.01

* Number of CPUs used for the calculation

† SCF cycles until convergence was achieved

‡ Wall time of the whole energy minimization calculation

to reach convergence. Using the B3LYP method and 16 CPUs, the minimization was accomplished within 1 h 45 min and after 15 SCF cycles (∼400 s/CPU/cycle, Table 3.12). Using unrestricted MP2 and 32 CPUs, 21 SCF were performed before convergence was reached and the whole minimization took ∼9 d—i.e. with ∼36 600 s/CPU/cycle the calculations were more than a whole order of magnitude slower (Table 3.12)! Because of the results of this benchmark, the unrestricted MP2 method was deemed computationally too expensive to be used to generate the QM target data necessary for the parametrization of ferric FEBIHIS.

As using MP2 was not an option for ferric FEBIHIS, we performed a B3LYP PES calculation for $O_{1/2}$–C_3–O_4–H_{O_4} in ferrous FEBIHIS to compare the B3LYP result to the MP2 profile used during the parametrization (Fig. 3.10). Both the

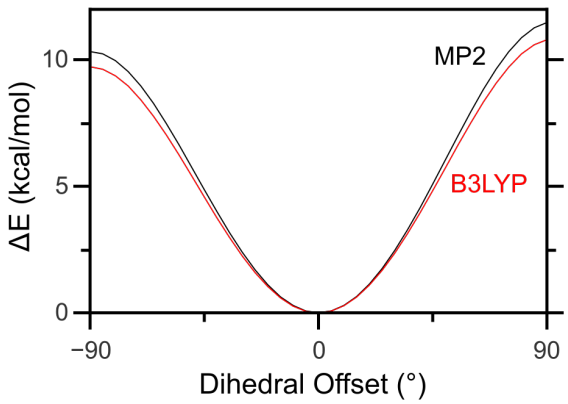


Figure 3.10: Torsional PES profiles for $O_{1/2}$–C_3–O_4–H_{O_4} in ferrous FEBIHIS performed with with MP2 (black) vs B3LYP (red). The MP2 profile was used during the parametrization of ferrous FEBIHIS. The B3LYP PES was computed to test the reliability of the B3LYP method when calculating dihedral PES profiles for the ferric FEBIHIS system. The barriers of the B3LYP profile are ∼0.7 kcal/mol than for MP2, but otherwise both profiles show good agreement, especially in the lower energy regime important for sampling in room temperature simulations.

MP2 and B3LYP profile showed good agreement with each other. Up to a 45° twist of the dihedral in either direction the difference between MP2 and B3LYP was less than 0.3 kcal/mol. In MP2, the barrier height was 10.3 kcal/mol and 11.5 kcal/mol at −90° and 90°, respectively. The B3LYP profile's energy always remained beneath the corresponding MP2 value, and the B3LYP energy barrier at −90° had a value of 9.7 kcal/mol and the one at 90° of 10.8 kcal/mol.

The B3LYP description of the $O_{1/2}-C_3-O_4-H_{O_4}$ dihedral did not perfectly agree with MP2, but the agreement, in particular for energies lower than 5 kcal/mol was high enough as not to affect the quality of derived parameters. Based on this calculation, we assumed a similar quality for the B3LYP profile, when applied to ferric FEBIHIS, especially bearing in mind that unrestricted MPn could give an inferior description if spin-contamination plays a role[191].

Testing the HAEMHIS Parameters with MD Simulations of Neuroglobin

This subsection is based on section 'Testing the revised HemeHis parameters with molecular dynamics of neuroglobin' from Adam *et al.* (2018)[179].

An additional way to test the quality of a specific parameter set is to conduct an MD simulation with said set and verify its robustness under true dynamic conditions, when many different conformations and interaction are sampled. To test the ferrous HAEMHIS parameters, we ran four independent all-atom MD simulations of neuroglobin in a TIP3P water box[179]. The simulations using HAEMHIS parameter Set 1 are denoted as SIM11 and SIM12, the simulations using HAEMHIS parameter Set 2 are denoted as SIM21 and SIM22. Hæm b in neuroglobin forms coordination bonds with histidine-64 (H64) and histidine-96 (H96) (Fig. 3.11).

In the neuroglobin crystal structure (PDB ID: 1OJ6[131]), both hæm-coordinating histidines have hydrogen-bonding partners: water w22 for H64 and the backbone oxygen of leucine-92 (L92) for H96. The hæm propionate groups hydrogen-bond to water molecules, with one of the propionate groups forming an additional hydrogen bond with tyrosine-44 (Y44) (Fig. 3.11).

Analysis of the root-mean-square deviation (RMSD) of the backbone C_α atoms with respect to the starting crystal structure revealed the α-helical segments to stay very stable throughout all simulation trajectories. The RMSD of the α-helices was (1.09 ± 0.11) Å, (0.80 ± 0.07) Å, (0.78 ± 0.08) Å and (0.79 ± 0.07) Å for SIM11, SIM12, SIM21 and SIM22, respectively; with average full-protein RMSDs of less than 2.3 Å (Fig. 3.12). The larger RMSD values for the loop regions might signify a lack of convergence for the dynamics of the protein loops.

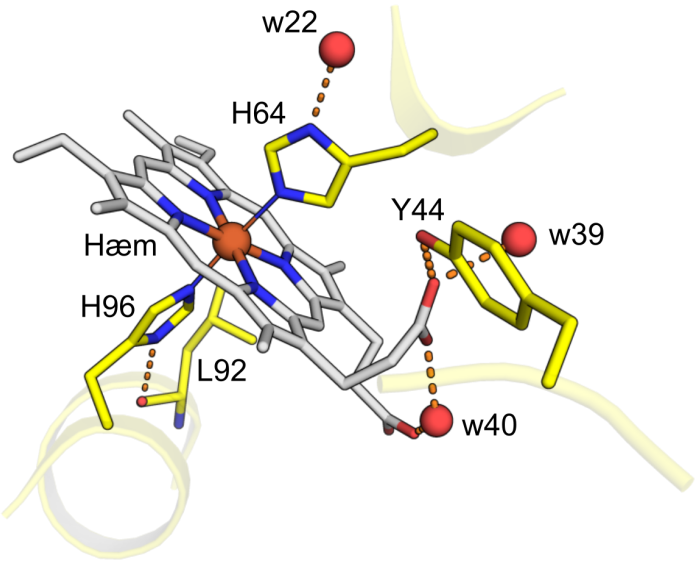


Figure 3.11: Close view of the hæm–histidine complex as found in neuroglobin[179]. The molecular graphics shows hæm b coordinating two histidine side chains and nearby hydrogen-bonding groups in the starting crystal structure (PDB ID: 1OJ6, chain B[131]). Water molecule w40 bridges the hæm propionate groups via hydrogen bonds with distances of 2.5 Å and 2.8 Å. The hydrogen bond between the backbone oxygen of L92 and N_δ of H96 was conserved throughout all simulations and had a value of (2.8 ± 0.1) Å when using HAEMHIS parameter Set 1 and of (2.9 ± 0.2) Å when using HAEMHIS parameter Set 2.

The average distance between the C_α atoms of H64 and H96 was 12.5 Å–12.7 Å with a standard deviation of 0.2 Å, and by staying close to (12.2 ± 0.1) Å, the distance computed for the four neuroglobin chains in the starting crystal structure[131], this was a first indicator for the quality of the HAEMHIS parameters (Fig. 3.13).

In addition to the hæm planarity, we also looked at structural parameters that described the geometry of the histidine groups relative to the hæm plane, namely: the hæm–histidine distance d, measured between the hæm iron and N_ε of H64 and H96; the angles $\angle Fe{-}N_\varepsilon{-}C_\varepsilon$ (α) and $\angle Fe{-}N_\varepsilon{-}C_\delta$, describing the bend of H64/H96 with respect to the hæm plane; and the dihedral angle ω, defined as $N_A{-}Fe{-}N_\varepsilon{-}C_\varepsilon$, that gave the twist of the histidine imidazole group relative to the hæm (Fig. 3.15 and Table 3.14). We used data from the 1OJ6 crystal structure[131] as reference values for these structural parameters (Table 3.14).

To further investigate the dynamics of the hæm, we analysed the planarity of the hæm by measuring the the out-of-plane bending of the hæm iron, characterized by $\angle CH{-}Fe{-}CH$ (Fig. 3.14a). For simulations using HAEMHIS Set 1 (SIM11 and SIM12), $\angle CH{-}Fe{-}CH$ had a value of $(177 \pm 2)°$, and for simulations using Set 2 (SIM21 and SIM22), it was $(176 \pm 2)°$, thereby indicating a slightly bent hæm

structure for both parameter sets (Fig. 3.14b and Table 3.13). This observation was backed by several crystal structures of neuroglobin that contained hæm b coordinated by two histidine residues (Table 3.13) and was compatible with B3LYP computations on hæm models and inspections of hæm-containing proteins conducted by Autenrieth *et al.*[190].

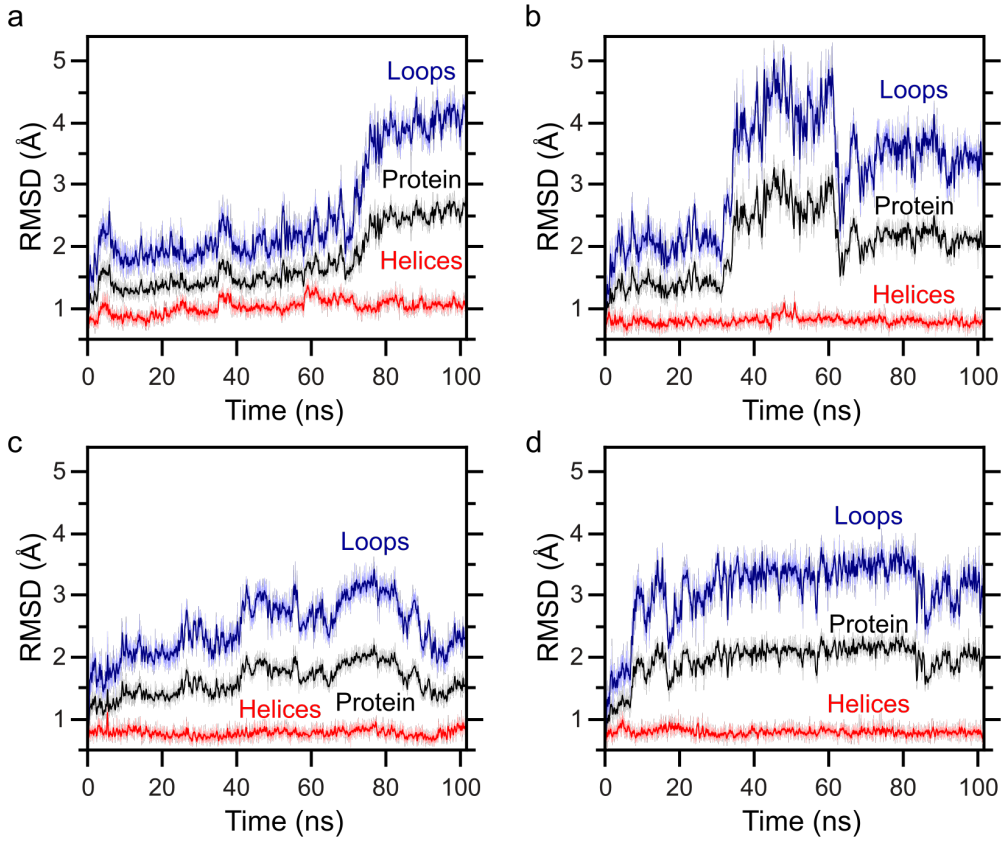


Figure 3.12: RMSD for the C_α atoms of neuroglobin during MD simulations using revised HAEMHIS parameters[179]. The loop segments have large RMSD values, their dynamics probably not being converged (blue). In contrast, helical segments have relatively small RMSD values (red). The RMSD of the protein as a whole is shown in black. Simulations SIM11 (a) and SIM12 (b) used HAEMHIS parameter Set 1 with optimized partial charges. Simulations SIM21 (c) and SIM22 (d) used the revised parameter Set 2 with CHARMM partial charges for hæm and histidine. All RMSD plots have been overlayed with a smoothed fit for clarity.

Inspection of the values sampled for d, α and ω (Fig. 3.15 and Table 3.14) indicated that, overall, the interaction of H64 and H96 with the hæm plane was well described by the revised HAEMHIS force field parameters presented here. For both H64 and H96, the lengths of the Fe−N_ε coordination bond computed from the simulations were in good agreement with the crystal structure (Table 3.14) and with QM (Table 3.8). Despite the agreement with the crystal data, the

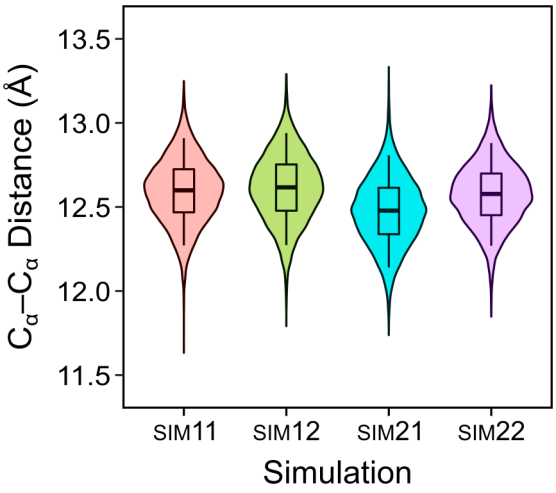


Figure 3.13: C_α–C_α distance for H64 and H96 in neuroglobin[179]. The violin plots show the distribution for the C_α–C_α distance sampled during the last 50 ns of simulation time. The inlayed box-and-whiskers plots indicate the 5th, 25th, 50th, 75th and 95th percentile. With (12.6 ± 0.2) Å, (12.7 ± 0.2) Å, (12.5 ± 0.2) Å and (12.6 ± 0.2) Å for SIM11, SIM12, SIM21 and SIM22, respectively, the values of the C_α–C_α distance of the different simulations are very close to the (12.2 ± 0.1) Å reference distance computed from distances found in the 1OJ6 crystal structure[131].

average Fe−N_ε distance in our simulations was slightly lower than the neuroglobin crystal structure average. Seeing as other bis-histidine-ligated hæm proteins can have shorter Fe−N_ε distances (Table 3.15), this further suggests applicability of the HAEMHIS parameters to such systems.

On average, ∠Fe−N_ε−C_ε was $(127 \pm 4)°$ in our simulations, as compared to $(126 \pm 1)°$ for the four protein chains of the crystal structure[131] (Table 3.14). In the crystal structure, there were differences in the value of ∠Fe−N_ε−C_ε and ∠Fe−N_ε−C_δ depending on the histidine residue, e.g. ∠Fe−N_ε−C_δ was $(125 \pm 1)°$ and $(119 \pm 2)°$ for H64 and H96, respectively. The symmetric nature of CHARMM parameters was not able to reproduce these differences in hæm–histidine interaction; however, the simulation value of $(128 \pm 3)°$ agreed with the H64 crystal value and was close to ∠Fe−N_ε−C_δ measured for H96 (Table 3.14).

The dihedral angle ω that characterized the orientation of the histidine with respect to the the hæm plane was well described for both H64 and H96 (Fig. 3.15 and Table 3.14).

As illustrated by the simulation results (Fig. 3.14 and 3.15 and Table 3.13 and 3.14), both Set 1 and Set 1 of the optimized HAEMHIS parameters presented here gave a reliable description of the hæm geometry and of the interactions between hæm and histidine.

In the crystal structure, a hydrogen bond could be observed between N_δ of

Table 3.13: Hæm planarity ∠CH−Fe−CH in our simulations and in various crystal structures of neuroglobin[179].

System	Planarity (°)	Resolution (Å)
SIM11	177 ± 2	—
SIM12	177 ± 2	—
SIM21	176 ± 2	—
SIM22	176 ± 2	—
1OJ6[131]	176 ± 1	1.95
1Q1F[192]	173 ± 5	1.50
2VRY[193]	175 ± 4	1.87
4B4Y[194]	170 ± 2	2.30
4MPM[195]	176 ± 2	1.74

H96 and the backbone oxygen of L92 (Fig. 3.11). This L92–H96 interaction was persistently sampled throughout all simulations. In SIM11 and SIM12, using HAEMHIS Set 1, the hydrogen bond was present ∼90 % of the time and the hydrogen bond distance was (2.8 ± 0.1) Å whereas, in SIM21 and SIM22, using HAEMHIS Set 2, the frequency was ∼80 % with a hydrogen bond distance of (2.9 ± 0.2) Å. These values agreed with the (2.7 ± 0.0) Å distance computed for the four chains of the neuroglobin crystal structure[131], with parameter Set 1 yielding a slightly better agreement with the crystal data. The increased frequency and strength of the L92–H96 hydrogen bond when using HAEMHIS Set 1 was due to the refinement of the histidine water interaction energies. When utilizing unrefined

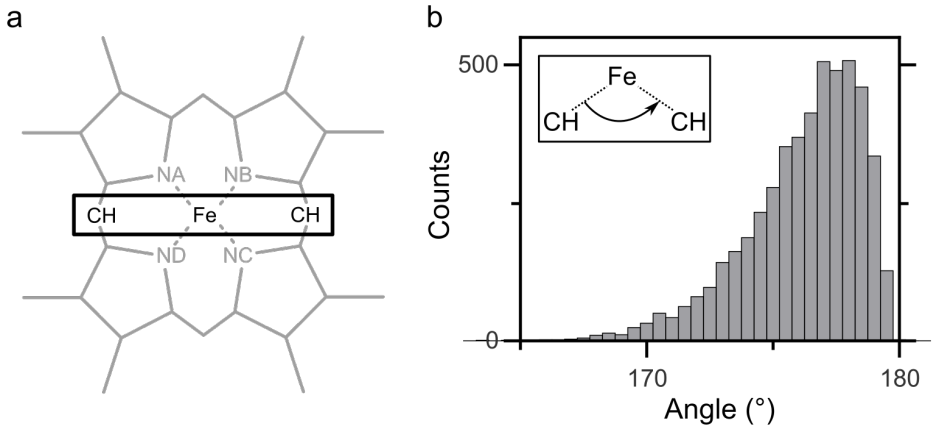


Figure 3.14: Planarity of the hæm plane during simulations of neuroglobin[179]. (a) The planarity of hæm is measured by the angle ∠CH−Fe−CH with two opposing methine carbons at the ends and the hæm iron in its centre. (b) The histogram shows the out-of-plane bending of the hæm iron in neuroglobin in SIM11. Throughout most of the simulation, the hæm stayed close to planarity, with an average angle ∠CH−Fe−CH of $(177 \pm 2)°$ during the last 50 ns of the simulation.

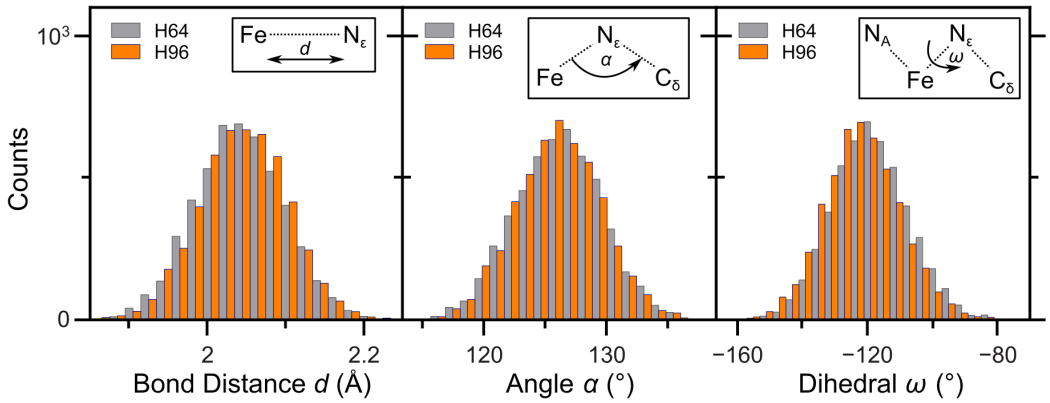


Figure 3.15: Geometry of the HAEMHIS complex during classical MD simulations of neuroglobin[179]. The bond lengths, valence angles and dihedral angles, measured for SIM11, are illustrated together with their corresponding histogram. Histograms show the distribution for the length of the coordination bond Fe−N_ε (left), the angle ∠Fe−N_ε−C_ε (centre) and the dihedral N_A−Fe−N_ε−C_ε (right) sampled during the last 50 ns of the simulation. The grey and orange bars correspond to data sampled from H64 and H96, respectively.

CHARMM charges (Set 2), the histidine N_δ interaction site underestimated water interactions by 1.88 kcal/mol—only the partial charge optimization performed for Set 1 made the N_δ interaction energy agree well with QM (Table 3.1).

The effect of the improved N_δ site water interactions was even more apparent when analysing H64. The crystal structure showed a hydrogen bond between H64 and the nearby water molecule w22 (Fig. 3.11). In simulations using the charge-optimized HAEMHIS Set 1 (SIM11 and SIM12), the hydrogen bond between H64 and w22 was present more than 97 % of the time. When using Set 2, w22 initially hydrogen-bonded with H64, but left the interaction site after 27 ns and 78 ns in SIM21 and SIM22, respectively, with no new water molecule entering and no new hydrogen bond forming at the N_δ site of H64.

As anticipated, based on the crystal structure (Fig. 3.11), both hæm propionate groups were consistently forming hydrogen bonds with water throughout all simulations. We note that the crystal structure contained a hydrogen bond between Y44 and one of the hæm propionate groups (Fig. 3.11). Because of the exposure of the propionate groups to water, a bond between hæm propionate group and Y44 was sampled only 0 %–29 % in our simulations, with this hydrogen bond breaking and reforming throughout the simulations.

Table 3.14: Geometry of the HAEMHIS complex during classical MD simulations of neuroglobin compared with the crystal structure[179].

Parameter	Residue	Crystal*	SIM11	SIM21
			Bond distance d (Å)	
Fe−N_ε	H64	2.11 ± 0.05	2.05 ± 0.06	2.06 ± 0.06
	H96	2.05 ± 0.04	2.05 ± 0.06	2.05 ± 0.06
			Angles (°)	
Fe−N_ε−C_ε (α)	H64	126 ± 1	127 ± 4	128 ± 4
	H96	132 ± 2	126 ± 3	127 ± 3
Fe−N_ε−C_δ	H64	125 ± 1	128 ± 3	127 ± 4
	H96	119 ± 2	128 ± 3	128 ± 3
			Dihedral ω (°)	
N_A−Fe−N_ε−C_ε	H64	-124 ± 4	-117 ± 12	-124 ± 14
	H96	-116 ± 4	-121 ± 12	-127 ± 11

*Crystal refers to values measured in the crystal structure (PDB ID: 1OJ6[131]). For H64 and H96, we report the geometry parameters illustrated in Fig. 3.15. Results for SIM12 and SIM22 can be found in Appendix A.7.

Table 3.15: Distances between hæm iron and histidine nitrogen N_ε for crystal structures of different proteins and for the neuroglobin simulation[179].

System	Distance (Å)
Neuroglobin (1OJ6[131])	2.08 ± 0.05
Photosystem II (3WU2[9])	2.04 ± 0.04
Myoglobin (2NRL[196])	2.03
Neuroglobin SIM11	2.05 ± 0.06
Neuroglobin SIM12	2.05 ± 0.06
Neuroglobin SIM21	2.05 ± 0.06
Neuroglobin SIM22	2.05 ± 0.06

Get your facts first, and then you can distort them as much as you please.

Mark Twain

4 Conclusions

CHARMM ALL-ATOM FORCE FIELD PARAMETERS were derived for the iron-containing cofactor complexes of photosystem II: the bis-histidine–hæm complex (Fig. 1.2) and the non-hæm iron–bicarbonate complex (Fig. 1.3). We optimized partial charges and intramolecular parameters for ferrous and ferric HAEMHIS and FEBIHIS, and additionally derived ferrous and ferric parameters for the non-hæm iron complex in the absence of bicarbonate (FEHIS). The parameters presented here facilitate reliable all-atom simulations of proteins that contain hæm and non-hæm iron complexes.

The HAEMHIS complex, which consists of hæm b forming a coordination bond with two histidine residues (Fig. 2.3), can be found in various proteins, such as photosystem II[9] (Fig. 1.2), neuroglobin[131,192,193,194,195] (Fig. 3.11), myoglobin[196], mitochondrial respiratory complex II[197] or in cytochromes[17,18,198]. To describe the HAEMHIS complex with CHARMM, we started from the parameters for single-histidine-ligated hæm[118,119,180]. We then optimized intramolecular parameters that involved the iron coordination bonds, i.e. force constants and equilibrium values for the bond lengths and valence angles (Figs. 3.7a and 3.7c and Tables 3.11 and 3.8) and the dihedral angle for torsion of the histidine imidazole groups with respect to the hæm plane (Fig. 3.8a–b). The agreement between target quantum mechanics (QM) data and molecular mechanics (MM)-obtained results was improved for all of these parameters.

For ferrous HAEMHIS, we derived two sets of parameters: one that used optimized partial charges (Set 1) and another one that used the standard CHARMM charges for hæm and histidine (Set 2). We tested both parameter sets by performing classical molecular dynamics (MD) simulations of neuroglobin and

could show that—independent of the charge set used—reliable description of the hæm–histidine geometry and behaviour was obtained (Figs. 3.13, 3.14 and 3.15, and Table 3.13 and 3.14). Water hydrogen-bonding interaction at the N_δ interaction cite of H64, however, were improved when employing Set 1 with optimized partial charges.

In addition to the CHARMM parametrization protocol, data from protein MD simulations might be used to improve a parameter set by adapting an optimized complex to specific proteins. The addition of an effective force to improve the description of conformational flexibility of the hæm site has been explored in myoglobin and cytochrome c[199]. Such approaches make a parameter set less transferable, but can give better results by accounting for the influence a specific protein environment has.

The non-hæm iron complex found in photosystem II[14,15,19,200] consists of a negatively charged bicarbonate and four histidines connected via coordination bonds to a positively charged iron ion (Figs. 1.3 and 2.4). To our knowledge, no standard CHARMM parameters exist for this complex, and we derived new force field parameters to describe intra- and intermolecular interactions of the non-hæm iron complex with (FEBIHIS) and without bicarbonate (FEHIS).

For the intramolecular parameters of FEBIHIS and FEHIS, we optimized bonded interactions for bonds and valence angles involving coordination bonds between the non-hæm iron and imidazole groups or bicarbonate, and obtained excellent agreement with QM data used as reference (Figs. 3.7b and 3.7d and Tables 3.9, 3.10 and 3.11). For FEBIHIS, we additionally optimized the dihedral parameters $O_{1/2}{-}C_3{-}O_4{-}H_{O_4}$, defining the orientation of the bicarbonate's hydroxyl group, and $Fe{-}O_{1/2}{-}C_3{-}O_4$, describing the torsion about the $O_{1/2}{-}C_3$ bond. Using the optimized FEBIHIS parameters, we achieved a good description of these dihedrals, correcting the shallow energy profiles obtained with the starting CHARMM parameters (Fig. 3.8c–f).

Because an Møller–Plesset perturbation theory (MP2) description of ferric FEBIHIS was not feasible (Table 3.12), we performed supplementary computations for ferrous FEBIHIS to assess the reliability of the Becke3,Lee–Yang–Parr functional (B3LYP) method, and obtained satisfactory agreement between B3LYP and MP2 for describing the $O_{1/2}{-}C_3{-}O_4{-}H_{O_4}$ dihedral potential (Fig. 3.10). We used the unrestricted B3LYP method for optimizing ferric FEBIHIS and FEHIS.

Atomic partial charges are used to define Coulombic interactions in classical force fields and are very important for all-atom MD simulations. In accordance with the CHARMM protocol for optimizing partial charges[27,99], we used HF/6-31G* to generate QM target data for water interaction energies and distances

at water accessible interaction sites of HAEMHIS, FEBIHIS and FEHIS(Fig. 2.7). Subsequently, the MM partial charges were manually adjusted until CHARMM was able to correctly reproduce the QM values. The final charge sets provided MM water interaction energies to within 0.4 kcal/mol, 0.6 kcal/mol and 1.2 kcal/mol of the QM values for HAEMHIS, FEBIHIS and FEHIS, respectively (Tables 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6).

For FEBIHIS, we additionally tested the newly derived parameters, by performing additional analyses of selected water interaction sites (Fig. 3.9). The calculations showed good agreement between QM and MM water interaction energies even away from the location of the minimum (Fig. 3.9a). Moreover, the results confirmed, for the computation of water interactions, how the choice of basis sets and methods other than HF/6-31G* might compromise the compatibility with other CHARMM force fields (Fig. 3.9b).

In photosystem II, the histidine side chains connect the non-hæm iron and the bicarbonate to the quinones Q_A and Q_B. The quinone interaction happens via hydrogen-bonding to N_δ of H214 and H215. In room temperature simulations, such interactions in a flexible protein environment will be dynamic. The good agreement between QM and MM water interaction energies of the histidines, particularly at the N_δ site (Fig. 3.9a), suggested a good description for protein interactions during MD simulations, as was the case for the improved water hydrogen-bonding interaction of H64 in the neuroglobin simulations SIM11 and SIM12.

The FEBIHIS parameters have been derived for a complex consisting of a non-hæm iron coordinating a bicarbonate and four histidine side chains. In nature, similar complexes can be found: In the bacterial photosynthetic reaction center[201], glutamate takes the place of the FEBIHIS bicarbonate; superoxide reductase has cysteine instead of bicarbonate[202]; in the factor inhibiting hypoxia-inducible factor, an iron ion coordinates bicarbonate, two histidines and α-ketoglutaric acid[203]; and in human lactoferrin and transferrin, iron coordinates bicarbonate, one histidine, one aspartate and two tyrosines[204,205]. The FEBIHIS and FEHIS parameters might be a good starting point to derive parameters for these complexes.

The topology and parameter files for HAEMHIS, FEBIHIS and FEHIS presented here have been released recently[179] and we anticipate that these parameters will be useful for performing classical CHARMM all-atom simulations of photosystem II and other proteins that contain hæm or non-hæm iron complexes. Analysis of the geometry and the interactions of the hæm and non-hæm iron complexes in such simulations can yield new insights into the functioning of proteins and could be used to further improve the quality of force field parameters.

Part Three

Proton Transfer in Channelrhodopsin

GUARD: *A swallow carrying a coconut?*
KING ARTHUR: *It could grip it by the husk!*
GUARD: *It's not a question of where he grips it! It's a simple question of weight ratios! A five ounce bird could not carry a one pound coconut.*

Monty Python and the Holy Grail

5 Results and Discussion

Parts of the results presented in this chapter have been submitted to PLOS ONE and are currently under revision.

The second part of this PhD project dealt with the role the environment of the retinal Schiff base in channelrhodopsin has in stabilizing the Schiff base proton in the closed/dark state. To this effect, we performed molecular mechanics (MM) and combined quantum and classical mechanics (QM/MM) calculations on the channelrhodopsin chimæra C1C2 (PDB ID: 3UG9[64]).

While the crystal structure contained only a single water molecule in the Schiff base region (w19, Fig. 1.5), the protein interior became quickly hydrated within a few nanoseconds during MM equilibration (Fig. 5.4), and our simulations confirmed the existence of a complex hydrogen-bonding network that helped to mediate interactions in the vicinity of the active site. The proton on the retinal Schiff base nitrogen was stabilized by the presence of the water molecules that had entered the retinal pocket and by the presence of the positive amino acid residue K132 (Figs. 1.5, 5.13, 5.14, 5.15 and 5.16). Likewise, protonation of E162 had a strong stabilizing effect on the Schiff base proton (Figs. 5.13f and 5.15). Inside hydrated wild type channelrhodopsin, however, the protonation state of E162 did not have a detectable impact on the reaction energy or barrier height of Schiff base deprotonation (Fig. 5.12).

Supplementary data regarding the analysis of the MM and QM/MM trajectories can be found in Appendix B.

Unless specified otherwise, all MM average values were computed from the last 50 ns of the respective simulation.

5.1 | Simulation Overview

We ran independent all-atom simulations of wild type C1C2 (PDB ID: 3UG9[64]) containing either all-*trans* or 13-*cis* retinal, and of the mutants K132 (mutant 1), E129A (mutant 2) and E162D (mutant 3) containing all-*trans* retinal. The nomenclature of the different simulation setups is introduced in Table 5.1.

The wild type simulations containing all-*trans* retinal with either unprotonated E162 (simWTu) or protonated E162 (simWTp) were prolonged to 250 ns. The wild type simulations with monomer 1 containing 13-*cis* retinal and monomer 2 containing all-*trans* retinal were denoted as simW1Cu and simW1Cp for negative and neutral E162, respectively, while the wild type simulations with retinal isomerized in both monomers were denoted as simW2Cu and simW2Cp for negative and neutral E162, respectively. All 13-*cis* simulations were prolonged to 100 ns.

We generated all-*trans* trajectories of 110 ns length for the mutants K132A (mutant 1, simM1), E129A (mutant 2, simM2) and E162D (mutant 3, simM3), with small caps u indicating unprotonated and small caps p indicating protonated E162.

Repeat simulations were performed for all simulations to improve statistical sampling and to ensure reproducibility of observations made.

For each simulation trajectory, convergence of the C_α RMSD had been achieved

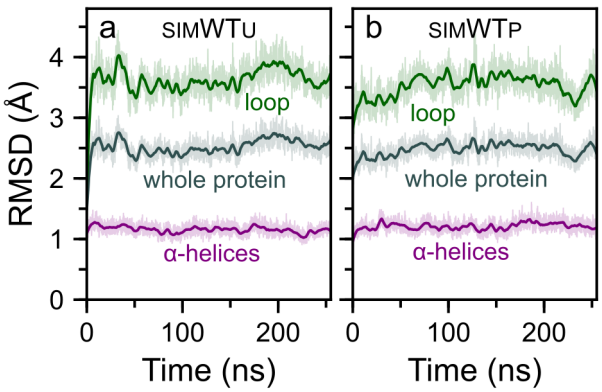


Figure 5.1: C_α RMSD profiles for simWTu and simWTp. (a) During the last 50 ns of simWTu, the RMSD had a value of (2.6 ± 0.1) Å for the whole protein (dark green), and a value of (1.1 ± 0.1) Å and (3.7 ± 0.2) Å for the α-helical regions (purple) and loops (dark slate grey), respectively. (b) With RMSD values of (2.5 ± 0.1) Å for the whole protein, (1.2 ± 0.1) Å for α-helices and (3.5 ± 0.2) Å for loops, simWTp showed the same good agreement with the crystal structure (PDB ID: 3UG9[64]) as simWTu. RMSD profiles of the remaining simulations can be found in Appendix B.1.

Table 5.1: Overview over the different simulations performed using wild type C1C2 (PDB ID: 3UG9[64]) and its mutants K132A (mutant 1), E129A (mutant 2) and E162D (mutant 3).

Label	Protein	Retinal	E162 Protonation	Length (ns)
simWTu	wild type	all-*trans*	unprotonated	250
simWTp	wild type	all-*trans*	protonated	250
simW1Cu	wild type	13-*cis*/all-*trans*	unprotonated	100
simW1Cp	wild type	13-*cis*/all-*trans*	protonated	100
simW2Cu	wild type	13-*cis*	unprotonated	100
simW2Cp	wild type	13-*cis*	protonated	100
simM1u	K132A	all-*trans*	unprotonated	110
simM1p	K132A	all-*trans*	protonated	110
simM2u	E129A	all-*trans*	unprotonated	110
simM2p	E129A	all-*trans*	protonated	110
simM3u	E162D	all-*trans*	unprotonated	110
simM3p	E162D	all-*trans*	protonated	110

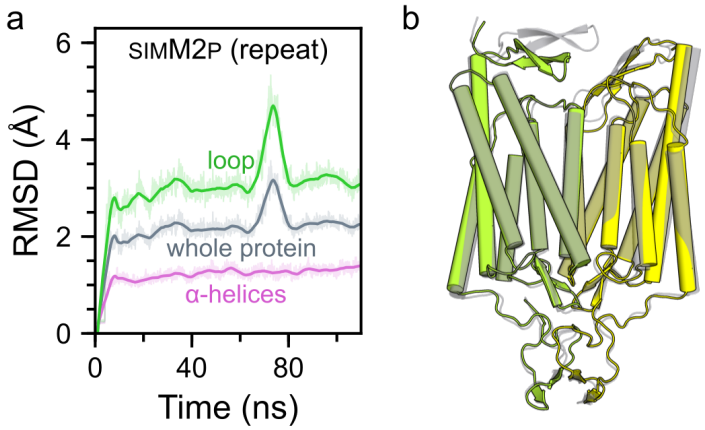


Figure 5.2: High RMSD value caused by sampling of a different loop conformation. (a) In the repeat simulation of simM2p, the C_α RMSD had a sudden peak of ∼3.4 Å for the whole protein (slate grey) and of ∼4.8 Å for the loop region selection (lime green). Analysis of the RMSD profile showed that this large peak was caused almost exclusively by the C-terminal loop. (b) A simulation snapshot taken at 70 ns is superimposed in grey on top of a snapshot taken at 80 ns of simM2p. The C-terminal loop can be seen to sample a different conformation, while the remainder of the protein shows good structural agreement between the snapshots.

for the last 50 ns that we used here for data analysis. The convergence had been indicated by the RMSD values reaching a stable plateau (Fig. 5.1, see Appendix B.1 for more information). We note that during the last 50 ns of simWTu the RMSD for the entire protein was (2.6 ± 0.1) Å and that the RMSD profile had reached its plateau within less than 10 ns (Figure 5.1a). The α-helical regions of the

protein stayed even closer to the crystal structure with an RMSD of (1.1 ± 0.1) Å, while the loop RMSD remained stable at (3.7 ± 0.2) Å. This observation of a very low α-helical RMSD was shared across all simulations, with average RMSDs of 1.1 Å–1.5 Å. Average RMSD values for the whole protein were between ∼2 Å–3 Å, which agreed with observations made for molecular dynamics (MD) simulations of channelrhodopsin[206,207] and similar complexes[208,209,210].

For the loop regions, a larger RMSD was measured—up to (4.0 ± 0.3) Å in the repeat simulation of simM2p (Fig. 5.2a). The large RMSD contributions from loop regions could be expected because crystal structures are derived in dehydrated media under cryogenic conditions. Once a protein is solvated and heated to room temperature, more conformations can be sampled, resulting in higher RMSD values (Fig. 5.2b). Moreover, on the ∼100 ns time scale of MD simulations, the motions of loops cannot be fully sampled[211], and including loop regions might make an otherwise well behaved protein appear not converged.

5.2 | Hydration and Hydrogen Bonding

Analysis of RMSD profiles indicated that, for all C1C2 simulations, the protein structure was overall well preserved (Figs. 5.1). Nevertheless, the number of water molecules inside the intrahelical region of C1C2 was significantly higher in our simulations as compared to the starting crystal structure (Fig. 5.3). The increase in the number of water molecules was more pronounced in the extracellular half of the protein, where water molecules entered rapidly close to the Schiff base region without causing large conformational changes (Fig. 5.3).

The presence of water molecules inside the protein enabled the formation of extended hydrogen-bonding networks within the Schiff base region, connecting its various residues to each other. We analysed the hydrogen-bonding behaviour of the key amino acid residues E162, D292, K132 and E129 and of the retinal Schiff base (Fig. 2.9) to examine if channel opening did occur and to identify possible proton transfer pathways for the first step of the Schiff base deprotonation. Specifically, we examined opening and closure of the central gate at E129, the role of K132 and immediate hydrogen-bonding partners of the Schiff base nitrogen.

Water in the Schiff Base Region

In the C1C2 crystal structures (PDB ID: 3UG9[64] and 4YZI[31]), only a single water molecule was resolved inside the Schiff base region. In contrast, crystal structures of bacteriorhodopsin and channelrhodopsin-2 exhibited a larger amount

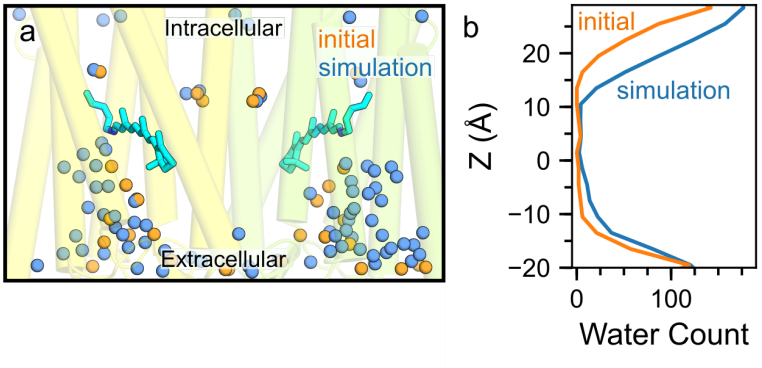


Figure 5.3: Hydration of the intrahelical region in C1C2 in simWTu. (a) Close view of C1C2 overlaying starting crystal water oxygens (orange spheres) with waters taken from a coordinate snapshot from simWTu (blue). (b) Distribution of water molecules along the z-axis in simWTu for water molecules within 6 Å of the protein. The initial setup (orange) placed the C1C2 crystal structure (PDB ID: 3UG9[64]) inside a hydrated POPC membrane. Comparing the starting coordinates with average numbers computed for the last 50 ns of simWTu (blue) showed an increase in the hydration of the inside of the protein, in particular in the extracellular half.

of internal waters close to their active site, with 3–4 crystal waters resolved in the Schiff base region[67,159,212,213,214]. Moreover, water molecules close to the retinal chromophore can contribute to stabilizing the retinal configuration and conformation[215,216,217]. Studies using infrared spectroscopy have shown recently that the channelrhodopsin chimæra contains more waters inside its active site than other microbial rhodopsins[218]. Because of all the afore-mentioned reasons, more inner water molecules might have been expected in the C1C2 crystal structures.

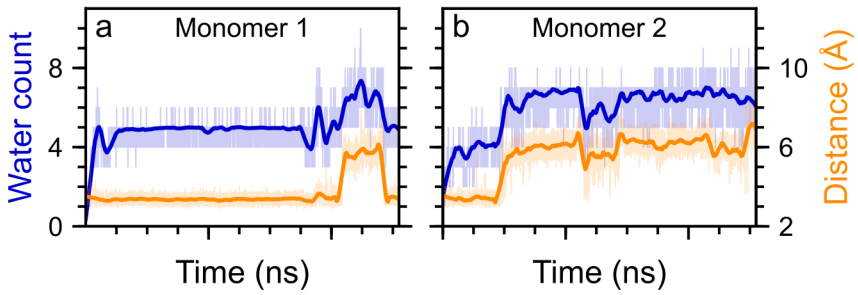


Figure 5.4: Water inside of the Schiff base region in the wild type setup with unprotonated E162 (simWTu) monomer 1 (a) and monomer 2 (b). The number of water molecules forming hydrogen bonds with the Schiff base nitrogen or the carboxyl(ate) oxygens of E162/D292 is shown and overlaid with a smoothed fit (blue). We compared the water number with the distance between the ammonium group nitrogen of K132 and C_γ of D292 (orange). The Pearson correlation coefficient for distance and hydration was 0.63 and 0.79 for monomer 1 and monomer 2, respectively. Results for the other simulations can be found in Appendix B.2.

Table 5.2: Number of water molecules forming hydrogen bonds with the Schiff base nitrogen or the side chains of E162 or D292.

	Number of Water Molecules		
Simulation	Monomer 1	Monomer 2	Combined*
simWTu	6.0 ± 1.0	6.6 ± 0.6	5.9 ± 0.8
simWTp	5.6 ± 0.7	4.5 ± 0.7	4.7 ± 0.7
simW1Cu	5.8 ± 1.1	5.2 ± 0.9	5.9 ± 1.0
simW1Cp	4.7 ± 1.0	2.7 ± 0.6	4.2 ± 0.8
simW2Cu	7.1 ± 0.9	5.2 ± 0.8	5.9 ± 0.8
simW2Cp	3.9 ± 0.7	4.5 ± 0.8	4.1 ± 0.7
simM1u	5.5 ± 0.7	4.7 ± 0.5	6.0 ± 0.8
simM1p	5.3 ± 1.0	5.7 ± 0.7	5.4 ± 0.8
simM2u	5.2 ± 0.6	7.0 ± 0.7	6.5 ± 0.8
simM2p	3.6 ± 0.6	4.1 ± 1.0	3.7 ± 0.6
simM3u	3.8 ± 0.6	6.4 ± 0.7	4.6 ± 0.5
simM3p	4.4 ± 0.7	4.0 ± 0.8	4.3 ± 0.5

*'Combined' reports the number of water molecules calculated for monomer 1 and 2 as well as for results from repeat simulations. Data for the repeat simulations can be found in Table B.1 in the Appendix.

Crystallography can only resolve ordered water molecules whose positions are repeated throughout the whole crystal[219], which is why MD can complement our knowledge about inner protein water molecules; and indeed, compared to the C1C2 crystal structure, our simulations showed a much more hydrated protein interior. To quantify the hydration inside of the Schiff base region, we measured, for each frame, the number of water molecules that formed hydrogen bonds with either the Schiff base nitrogen or the counterions E162 or D292. We then calculated the trajectory average for the last 50 ns (Table 5.2).

Looking at the number of these Schiff base waters, there were on average 5–7 water molecules in wild type setups with unprotonated E162 (simWTu, Fig. 5.3a) and 4–6 waters for protonated E162 (simWTp, Fig. 5.3b). In all simulations, hydration of the Schiff base regions in both monomers occurred quickly within the first few nanoseconds. Overall, the observed water numbers showed good agreement between the monomers in a particular setup. We note, however, that for some simulations the two monomers displayed a different hydration: In simW2Cu, for instance, Schiff base water numbers were 7.1 ± 0.9 and 5.2 ± 0.8 in monomer 1 and monomer 2, respectively. For all simulation setups, hydration of the Schiff base region was on average larger for systems with unprotonated E162 (Table 5.2).

In our simulations, the amount of water in the vicinity of the retinal Schiff base was closer to the amount found in crystal structures of bacteriorhodopsin or channelrhodopsin-2 than in the C1C2 crystal structure with its single water molecule. The quick and extensive hydration of the protein interior, observed by us, agreed well with other computational studies of channelrhodopsin C1C2[158,220]. MD simulations performed for homology models of channelrhodopsin-2 also predicted a strong hydration of the protein inside[148,221,222].

Supplementary illustrations of the water density inside the channelrhodopsin simulations can be found in Appendix B.4.

The E129–N297 Hydrogen Bond in the Central Gate Region

In the closed/dark state of channelrhodopsin, the central gate forms the first barrier that prevents water from entering from the extracellular side via the Schiff base region and contains E129, N297, serine-102 (S102) and the Schiff base counterions E162 and D292[64,158] (Fig. 1.5). All of our setups contained protonated E129, the protonation state of E129 in the dark state of channelrhodopsin[65,150,151].

In most of our simulations, we could see E129 forming a strong hydrogen bond with N297 that persisted throughout the entire trajectory. (See Appendix Tables B.2, B.3 and B.4 for hydrogen-bonding data of E129.) The notable exceptions were simulations of E129A (simM2), the protonated E162 setup with 13-*cis* retinal in one monomer only (simW1Cp) and the repeats of the wild type setup with unprotonated E162 (simWTu) and the E162D setup with protonated D162

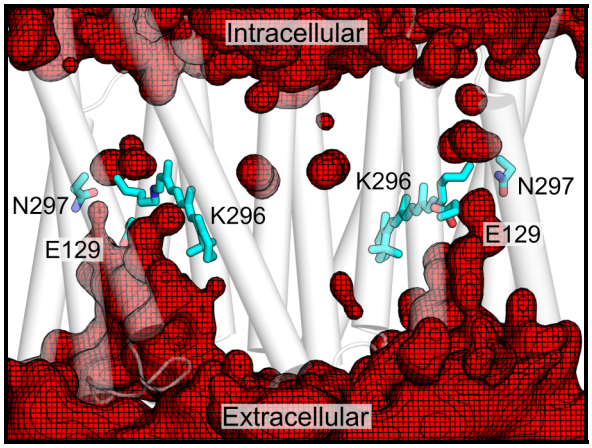


Figure 5.5: Water density inside of channelrhodopsin in simW1Cp. In the 13-*cis*-retinal-containing monomer (monomer 1, left), the hydrogen bond between E129 and N297 was broken after 40 ns and stayed broken for ∼30 ns. In monomer 2, with all-*trans* retinal, the bond broke within 5 ns of starting the production run and never reformed. The water density displayed has been computed for 25 ns of trajectory time sampled while the E129–N297 hydrogen-bond did not exist.

(simM3p).

In simW1Cp, monomer 1 contained 13-*cis* retinal, while monomer 2 contained all-*trans* retinal. Nevertheless, complete breakage of E129–N297 occurred only in monomer 2 with all-*trans* retinal, where the hydrogen bond broke within 5 ns without reforming. For monomer 1, only between 40 ns and 70 ns of trajectory time was E129 not hydrogen-bonded to N297. Despite the bond breakage, no connected water pathway formed between extracellular and intracellular side in either monomer (Fig. 5.5) and the channel remained closed.

The simulation simWTu constituted the most probable setup for c1c2 (using the wild type and unprotonated E162); however, the initial run and the repeat simulation showed very distinct behaviour. In the initial run, a bond between E129 and N297 was sampled at every frame, and E129 interacted only with N297 and nearby water molecules. No water channel formed in this simulation (Fig. 5.6a). In the repeat simulation, E129–N297 broke after ∼120 ns in monomer 1 and after ∼190 ns in monomer 2, and instead, E129 formed a strong hydrogen bond with E162 that was present 92 %–100 % during the last 50 ns. Despite the bond breakage in monomer 1 of the repeat simulation, water molecules were unable to access the space between E129 and N297 (Fig. 5.6b); yet in monomer 2, water molecules formed a hydrogen-bonding network connecting the extracellular to the intracellular side (Fig. 5.6c).

For E162D (simM3), the E129–N297 hydrogen bond was conserved—the single exception being monomer 2 in simM3p. In monomer 2 of simM3p, the hydrogen bond between E129 and N297 broke after ∼50 ns, and water penetrated the region formerly occupied by E129 (Fig. 5.6d). A connected water pathway formed, and four instances of a water molecule crossing monomer 2 were observed: Three times from the extracellular to the intracellular side and once in the opposite direction. The time, it took a water molecule to cross the membrane, was 33 ns, 40 ns and 49 ns—when entering from the extracellular side—and 37 ns—when entering from the intracellular side.

The absence of E129 from the E129A mutant simulations (simM2u and simM2p) led to the hydration of the E129–N297 gate region, and connected water pathways formed in all E129A mutant simulations. During the initial run of simM2u, water molecules connected the extracellular and intracellular side of both monomers. While water molecules were able to freely enter all parts of the channel, no water passage could be observed. Notably, monomer 1 of simM2u had three very stable water molecules in the vicinity of N297 that connected both sides of the gate and were present there during the last 50 ns of the simulation. In the repeat simulation of simM2u, we detected four water molecules per monomer

fully crossing the protein via the central gate. One water molecule traversed monomer 2 from the intracellular to the extracellular side, while all other seven waters crossed the membrane in the opposite direction. The passage times were 17 ns–83 ns. The E129A simulations with protonated E162 showed similar results with 3–5 water molecules passing through the protein. In total, eight waters went from the intracellular to the extracellular side and four went in the opposite direction, taking 19 ns–78 ns for passing the protein. Only in monomer 2 of the initial run of simM2p did no water molecule cross the membrane.

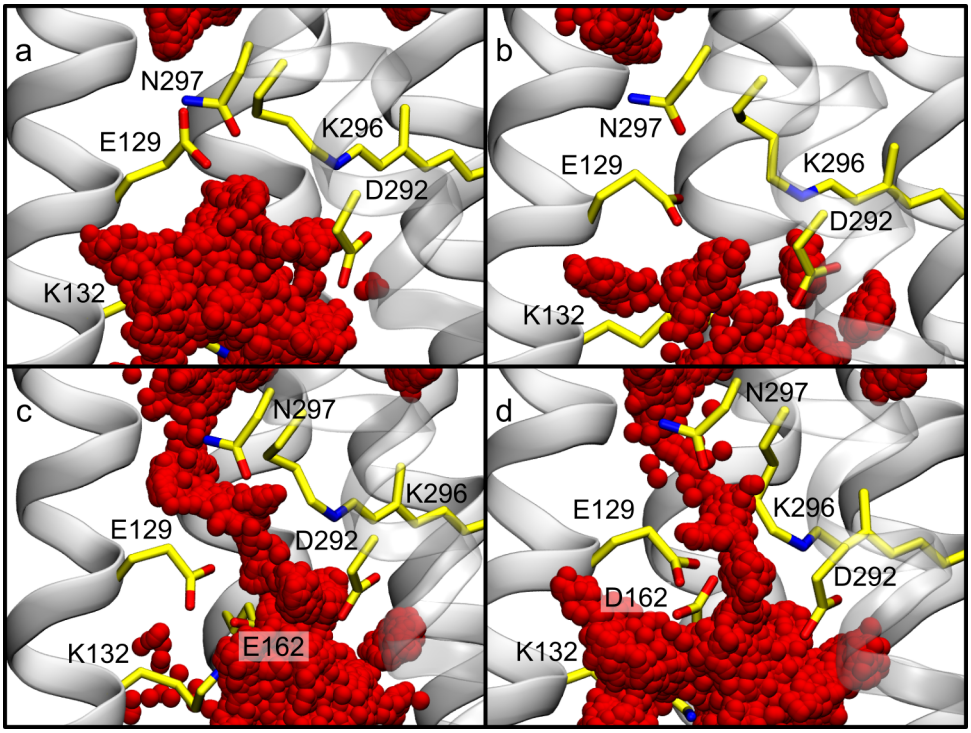


Figure 5.6: Water dynamics close to E129–N297 in the central gate. Water oxygen positions are shown in red for the last 10 ns of the respective simulations. Their position has been averaged over 5 frames to more easily visualize clusters and water pathways. (a) In monomer 1 of simWTu, E129–N297 remained stable throughout the simulation. (b) In monomer 1 of simWTu (repeat), E129–N297 broke, but water molecules were kept away from the gate. (c) In monomer 2 of simWTu (repeat), E129–N297 broke, but unlike the situation in monomer 1, water molecules entered and formed a connected water pathway between extracellular and intracellular side of the membrane. (d) In monomer 2 of simM3p (repeat), water molecules were present in the space between E129 and N297 after breakage of their bond. In this simulation, water passage from the extracellular to the intracellular side, or vice versa, was observed.

We did not make any notable observations for the K132A simulations with unprotonated (simM1u) and protonated E162 (simM1p). E129–N297 remained stable, and mutating K132 to alanine did not appear to affect the functioning of the central gate in the dark state in terms of channel opening.

In the remaining MM simulations, the hydrogen bond between E129 and N297 lasted throughout the whole trajectory. In the QM/MM simulations as well, we sampled E129–N297 for all setups, independent of whether the quantum mechanics (QM) region had been placed in monomer 1 or monomer 2. The analysis of the hydrogen bond in the QM/MM trajectories was performed for the last 750 ps and yielded 100 % occupation for all trajectories, except for simM3u, which had E129–N297 present 43 % of the time. Thorough investigation of the heavy atom distance between the ammonium nitrogen of N297 and the nearest carboxyl oxygen of E129 revealed that the interaction between both residues, though mostly not within hydrogen-bonding range, was never fully broken (Fig. 5.7).

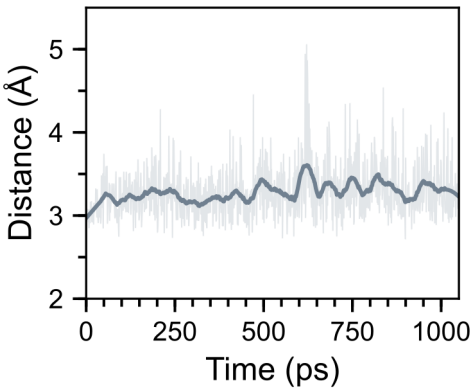


Figure 5.7: Distance between E129 and N297 during QM/MM MD of the E162D setup containing unprotonated E162 (simM3u) with the QM region located in monomer 2. The hydrogen bond E129–N297 was present during 43 % of the QM/MM simulation. Analysis of the heavy atom distance d_{ON} between the ammonium nitrogen of N297 and the closest carboxyl oxygen of E129 showed that interaction between both residues was never fully lost. The carboxyl–ammonium distance d_{ON} was (3.3 ± 0.3) Å.

Several studies have been made about the interaction between E129 and N297 and its importance for gating the ion channel in wild type channelrhodopsin[41,206], and a stable E129–N297 hydrogen bond has been proposed to be necessary for channel closure[151,156]. In our simulations, we observed a high hydrogen-bonding rate between E129 and N297, which appeared to confirm the significance of this interaction for preventing ion transfer in the closed state. In fact, the hydrogen bond was missing only from simW1Cp, the repeats of simWTu and simM3p and simulations of the mutant E129A.

Monomer 1 of the wild type simulation simW1Cp contained 13-*cis* retinal, and the breakage of E129–N297 could simply have been the first stage of channel opening, initiated by retinal isomerization. This was the only 13-*cis* simulation

that contained a continuous water pathway connecting extracellular and intracellular side. Opening of the central gate might have been expected for all 13-*cis* simulations, but most likely, the length of the simulation was not long enough, for larger scale conformational changes of the protein necessary for channel opening to take place.

In the repeat simulation of simM3p, the E162D mutant, the interaction between the counterions D162 and D292 was broken, possibly because removing the methylene group at position 162 increased the distance from D292. Consequently, E129 was able to form a strong hydrogen bond with the protonated D162, instead of bonding with N297.

The most interesting results, however, were obtained for simWTu, the wild type simulation using the predicted E162 protonation state for C1C2[41,64,223]. During previous studies that used the same protonation states, other groups observed a strong E129–N297 hydrogen bond that was not broken[41,206]. This apparent disagreement with our results might be explained by the differences in simulation setup and time. The preceding MM MD studies in the field used a monomer setup of C1C2 and generated shorter trajectories of less than 120 ns. Even though, the C1C2 monomer has been proposed to be a functional unit[59,156], a single monomer in a lipid bilayer might not behave exactly like the natural dimeric setup. We stress that bond breakage between E129 and N297 occurred more than 120 ns after the end of equilibration, and therefore, would have been unlikely to be observed in simulations of shorter time scales.

In the simulations, where E129–N297 broke, the formation of a continuous pathway for water became possible. When water pathways formed, we could observe free water passage in either direction of the channel. The wide range of water passage times of up to ∼80 ns suggested that not all water passage events could be sampled. More sampling—either by prolonging the simulations or by adding further repeat simulations—is necessary to properly evaluate water flux in dark state simulations of C1C2.

Because our simulation setups either contained no ions or only two chloride ions for systems with unprotonated or protonated E162, respectively, we could not investigate ion flux. Free water passage, however, does not imply free flow of ions across the channel. The mutant E129A, for example, lacks the hydrogen bond between E129 and N297 in the central gate and still has reduced proton and, for low pH, sodium conductance[224,225]. In channelrhodopsin-2, E129 deprotonates during the photocycle[150,156], and the presence of a negative E129 is important for ion selectivity[151,226]. New MD studies even suggest that N297 and K132 form a barrier for ion translocation[207]. The significance of the E129 charge for

ion selectivity can be further highlighted by the fact that mutating E129 to a positive residue changed channelrhodopsin to a chloride pump[158]. Yet in C1C2, deprotonation of E129 during the photocycle does not occur[65], and breakage of the hydrogen bond between E129 and N297 can, therefore, not depend on the change of the protonation state of E129.

As was implied by the results from our simulations of C1C2, the hydrogen bond between E129 and N297 can be at most one of the components that prevents ion passage through channelrhodopsin in the closed/dark state. The E129–N297 hydrogen-bonding interaction, though present in most of our simulations, can be broken even for the wild type, and its continued presence is not required inside C1C2. Future MD studies of channelrhodopsin should use trajectory lengths >200 ns to ensure sufficient sampling of the hydrogen-bonding states in the central gate region.

Hydrogen-Bonding Networks in the Schiff Base Region

The water molecules that entered the Schiff base region could engage in hydrogen-bonding networks with the Schiff base and the nearby charged amino acid residues D292, E162, K132 and E129[64] (Fig. 2.9). To identify possible acceptors for the Schiff base proton, we performed a thorough analysis of hydrogen-bonding in the proximity of the Schiff base. (Additional data for the hydrogen-bonding analysis of the Schiff base region can be found in Appendix B.3.)

In channelrhodopsin, the positively charged Schiff base is faced by the negative amino acid residues E162 and D292. Mutations of E162 preserve the functionality of C1C2, and it is assumed that D292 is the primary proton acceptor[64,150]. In the crystal structure (PDB ID: 3UG9[64]), the counterions E162 and D292 interact with the Schiff base and a single nearby water molecule. In our simulations, both residues were constantly exposed to water molecules that entered the Schiff base region. The hydration of the Schiff base region at times mediated the interaction between E162 and D292, which communicated with each other either via direct hydrogen bond or via one- to two-water bridges in both MM and QM/MM.

We used the distance between C_γ of D292 and C_δ of E162 (C_γ in E162D) to characterize the separation of these two residues (Fig. 5.8). In wild type all-*trans* simulations, the average E162–D292 distance was 5.2 Å and 5.1 Å for unprotonated and protonated E162, respectively. Isomerization of the retinal increased the difference between the average distances obtained for the two protonation states of E162: For unprotonated E162, the separation became 5.5 Å, and for protonated E162, it became 4.7 Å. For the mutants K132A and E162D,

distances for unprotonated and protonated E162 were similar, but longer than for the wild type simulations. In K132A, the unprotonated and protonated distances were 5.5 Å and 5.6 Å, respectively; and for E162D, they were 5.6 Å and 5.8 Å. For setups with protonated E162, mutating E129 to alanine resulted in water molecules not entering the Schiff base region and E162 and D292 permanently engaging in strong hydrogen-bonding with each other, with an average distance of 4.2 Å. The unprotonated setup had a separation of 5.3 Å, similar to the wild type values.

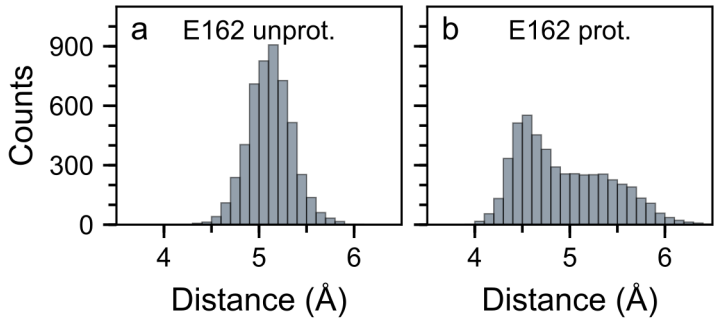


Figure 5.8: Histogram of the distance between E162 and D292 in monomer 1 of the wild type setups with unprotonated (simWTu) and protonated E162 (simWTp). (a) In simWTu, interactions between E162 and D292 were mediated by a one-water bridge, and the separation between the carboxyl carbons of both residues in monomer 1 was (5.1 ± 0.2) Å during the last 50 ns. (b) In simWTp, direct hydrogen-bonding of E162 and D292 became possible, which resulted in a wider spread of the distance data and a value of (4.9 ± 0.5) Å.

K132, which closes off the Schiff base pocket on the extracellular side, is another important charged amino acid residue in the Schiff base region. K132 might serve as a barrier for ions[207], and the presence of complex hydrogen-bonding networks that involve K132 has been suggested[218] and recently been observed[206].

For MM setups with unprotonated E162, we saw K132 strongly interacting via hydrogen-bonding with either E162 or D292 or with both residues at the same time. In QM/MM, the hydrogen-bonding percentages of the K132–E162 and K132–D292 interactions were higher, and even setups with protonated E162 contained hydrogen-bonding between K132 and E162/D292—though at a lower level than observed for unprotonated E162. To understand K132's possible function as an ion gate, we analysed the linear correlation between the K132–D292 separation and the solvent exposure of the Schiff base region. As separation, we defined the distance between the ammonium nitrogen of K132 and the carboxyl group carbon of D292. The solvent exposure was the number of water molecules that formed hydrogen bonds with either the Schiff base nitrogen or the carboxyl(ate)

groups of E162/D292.

Our analysis revealed that these two quantities were strongly linearly correlated for systems containing unprotonated E162 (Table 5.3). The correlation coefficient had values of ∼60 %–80 %, with smaller K132–D292 separation meaning fewer water contacts were present (Fig. 5.4). The correlation persisted even without direct hydrogen-bonding between K132 and E162 or D292. No correlation was apparent for setups using the E162D mutant or protonated E162.

Table 5.3: Linear correlation between K132–D292 distance and the number of water molecules forming a hydrogen bond with either the Schiff base or the counterions E162/D292 in MM MD of wild type C1C2 and its mutants. The Pearson correlation coefficient is shown.

	First Run		Repeats	
Simulation	Monomer 1	Monomer 2	Monomer 1	Monomer 2
simWTu	0.6	0.8	0.7	0.6
simWTp	0.1	0.3	0.4	0.2
simW1Cu	0.7	0.6	0.8	0.3
simW1Cp	0.3	0.2	−0.1	0.2
simW2Cu	0.3	0.3	0.7	0.4
simW2Cp	0.4	0.1	−0.5	0.0
simM2u	0.3	0.6	0.7	0.7
simM2p	0.1	0.3	0.3	0.2
simM3u	−0.4	−0.1	−0.2	0.0
simM3p	0.2	0.1	0.3	0.1

In the final step of the hydrogen-bonding analysis of the Schiff base region, we looked at the immediate hydrogen-bonding partners of the retinal Schiff base nitrogen (Table 5.4). (See Appendix B.3 for hydrogen-bonding percentages of the repeats and the QM/MM simulations.)

E162 and D292 were the residues closest to the Schiff base in the crystal structure (Fig. 1.5), and for all but setups containing 13-*cis* retinal, they were the only hydrogen-bonding partners the Schiff base had with amino acid residues. For all-*trans* systems, it was not possible to distinguish wild type and mutant simulations based on the hydrogen-bonding behaviour of the Schiff base, as the Schiff base interaction analysis yielded similar results (Table 5.4). We could identify three distinct Schiff base hydrogen-bonding partners: D292, E162 or water (Table 5.4), where water that hydrogen-bonded to the Schiff base always formed a one-water bridge connecting the Schiff base nitrogen to either E162 or D292. Hydrogen-bonding between Schiff base and D292/water was sampled more often, because, for the MM systems, protonated E162 did not form stable

hydrogen bonds with the positively charged Schiff base nitrogen. We note that in some of the QM/MM trajectories, e.g. simWTu monomer 1, protonated E162 was hydrogen-bonded to the Schiff base nitrogen up to ∼40 % of the time.

Table 5.4: Hydrogen-bonding partners of the Schiff base nitrogen during the last 50 ns of MM MD.

Simulation	Monomer	Hydrogen-Bonding Percentage (%)				
		Water	E162	D292	S295	T166
simWTu	1	19	92	—	—	—
	2	68	—	51	—	—
simWTp	1	99	—	3	—	—
	2	57	—	68	—	—
simW1Cu	1	—	—	—	81	—
	2	8	10	89	—	—
simW1Cp	1	—	—	—	80	—
	2	—	12	96	—	—
simW2Cu	1	48	—	—	23	—
	2	9	—	—	—	92
simW2Cp	1	20	—	94	—	—
	2	—	—	—	93	—
simM1u	1	82	—	38	—	—
	2	—	—	100	—	—
simM1p	1	37	—	85	—	—
	2	86	—	24	—	—
simM2u	1	—	100	—	—	—
	2	92	—	22	—	—
simM2p	1	—	16	98	—	—
	2	18	34	72	—	—
simM3u	1	—	100	—	—	—
	2	100	—	8	—	—
simM3p	1	16	—	87	—	—
	2	12	—	90	—	—

To increase clarity, only percentages >3 % are shown.

In C1C2 monomers containing 13-*cis* retinal, the retinal isomerization to 13-*cis*,15-*anti* caused the Schiff base nitrogen to face away from the counterions E162/D292, and T166 and S295 (Fig. 2.9) became the new interaction partners of the Schiff base (Table 5.4). For simulations, with high hydrogen-bonding percentage between Schiff base and S295, the Schiff base nitrogen was interacting with the backbone oxygen of S295 (Fig. 5.9a). Hydrogen bonds with water remained possible and could bridge the interaction with S295 or T166 (Fig. 5.9b). In monomer 1 of simW2Cp and in monomer 2 of the repeats of simW2Cu and

simW2Cp, the retinal conformation changed from 13-*cis*,15-*anti* to 13-*cis*,15-*syn* during the MM MD (Fig. 5.9c), and the Schiff base nitrogen strongly interacted with D292 (Table 5.4). We note that in one 13-*cis* QM/MM trajectory (simW1Cu monomer 1), the Schiff base nitrogen did not have any hydrogen-bonding partner 76 % of the time (Fig. 5.9d).

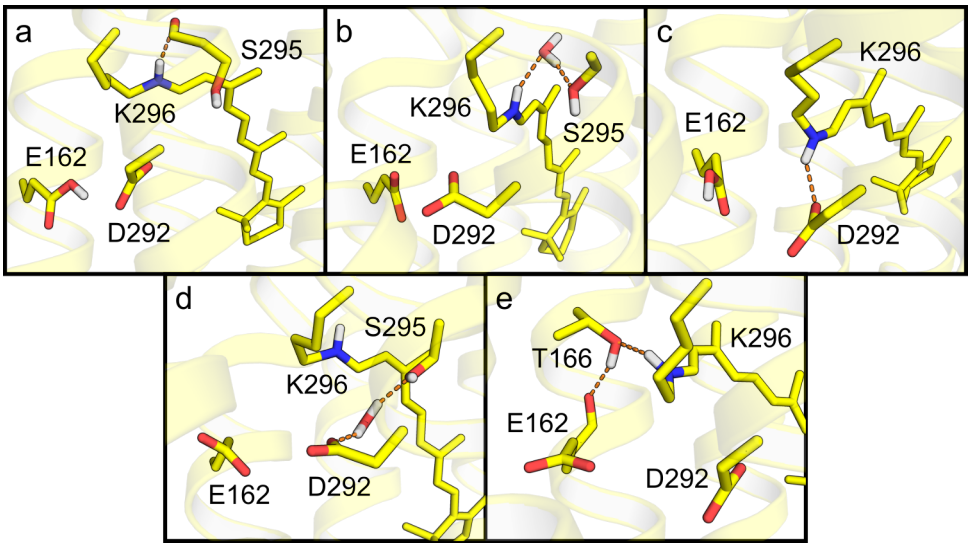


Figure 5.9: Hydrogen-bonding partners of the Schiff base nitrogen for simulations containing 13-*cis* retinal. (a) In most of the MM setups, 13-*cis* retinal formed a hydrogen bond with the backbone oxygen of S295. (b) The Schiff base nitrogen interacts with S295 via a one-water bridge (simW2Cu). (c) In simW2Cu, the retinal changed from 13-*cis*,15-*anti* to 13-*cis*,15-*syn*, and formed a direct hydrogen bond with D292. (d) In simW1Cu with the QM region in monomer 1, the Schiff base rarely had a hydrogen-bonding partner. In this snapshot, S295 is too far from the Schiff base nitrogen to engage in hydrogen-bonding. Interaction with S295 would offer the possibility to connect the Schiff base to D292 via an extended hydrogen-bonding network. (e) The Schiff base hydrogen-bonds to the hydroxyl group of T166 which in return interacts with the backbone of E162. This hydrogen-bonding network did not offer the possibility for a proton transfer pathway that connects to the primary proton acceptor D292.

In what follows, we will discuss the results obtained for the analysis of the hydrogen-bonding networks in the Schiff base region.

The evaluation of the distance and hydrogen-bonding between E162 and D292 revealed that both residues strongly interact with each other. In the crystal structure, E162 and D292 are located at 4.78 Å distance from each other[64]. Our wild type simulations had average E162–D292 distances that tended to be higher than the crystal value, with setups containing protonated E162 showing good agreement with the separation found in the crystal structure due to larger standard deviations (Fig. 5.8). We stress, however, that some of the unprotonated E162 simulations agreed with the crystal structure value—e.g. in monomer 1 of the repeat of the wild type simulation with unprotonated E162 (simWTu),

E162 and D292 had a separation of (4.8 ± 0.2) Å. Simulations, where water-mediated hydrogen-bonding or both direct hydrogen-bonding and water-mediated hydrogen-bonding between E162 and D292 could be observed, gave E162–D292 distances closest to the crystal observations. The results from the E162–D292 distance analysis, although not providing conclusive evidence for the preference of either protonated E162 or unprotonated E162, hinted at the necessity for water-bridging and for absence of direct hydrogen-bonding between E162 and D292 to dynamically reproduce the E162–D292 crystal separation.

We further investigated the interactions between K132 and E162/D292. In the crystal structure (PDB ID: 3UG9[64]), E162 is hydrogen-bonded to K132 with a carboxyl oxygen–ammonium nitrogen distance d_{ON} of 3.36 Å; for D292, d_{ON} is 3.76 Å. Recent pK_{a} studies suggested that the protonation of E162 should not significantly affect the hydrogen-bonding networks in the Schiff base region[206], but our simulations showed a clear difference in the K132–E162/D292 interaction depending on the protonation state of E162.

For unprotonated E162, K132 would engage in strong hydrogen-bonding behaviour with either E162 or both E162 and D292. Using protonated E162 in a setup broke this direct hydrogen-bonding network. In the QM/MM simulations, the K132–E162 hydrogen bond could still be sampled even for protonated E162, and the disappearance of the interaction in MM might have been an effect of the non-polarizable nature of the CHARMM force field. For both MM and QM/MM, however, protonated E162 interfered with K132 interacting with the counterions E162 and D292, and the K132–E162 salt bridge, found in the crystal structure, needed unprotonated E162 for its conservation. In that, our analysis supported studies that favour unprotonated E162[149,206]. The correlation between K132–D292 distance and hydration of the Schiff base was also broken by mutating E162 to aspartate. This might indicated that the flexibility of the longer glutamate side chain might play a role in the communication between unprotonated E162 and K132.

When K132 and E162/D292 hydrogen-bonded, the Schiff base pocket was more compact than was the case in the absence of such a hydrogen bond or for the mutant K132A. Analysis of the number of water contacts for Schiff base nitrogen and E162/D292 showed a clear correlation with the K132–D292 distance (Fig. 5.4 and Table 5.3). Water entering the Schiff base pocket could bridge the interaction between K132 and E162/D292 without breaking their communication. Likewise, cations trying to enter the Schiff base region would have to pass closely by K132, but unlike water molecules, cations would be repelled by the positively charged K132. An unprotonated E162 and the strong interaction between E162

and K132 could, therefore, be a necessary requirement for K132 to perform its gating function in the closed/dark state of channelrhodopsin and prevent cations from entering the Schiff base region.

Finally, we examined the direct hydrogen-bonding partners of the Schiff base nitrogen. We found that, for the all-*trans* simulations, the Schiff base hydrogen-bonded only to E162, D292 or water molecules that had entered the Schiff base region and were themselves hydrogen-bonded to either E162 or D292 (Table 5.4). These observations were compatible with the crystal structure, where the Schiff base nitrogen is located 3.01 Å and 3.44 Å from E162 and D292, respectively, and with suggestions from experiments that D292 is the primary proton acceptor[64,150]. The varied hydrogen-bonding patterns we found in our simulations—e.g. the Schiff base only hydrogen-bonding to E162 or the Schiff base alternating between hydrogen-bonding to D292 and a water molecule (SIMWTU, Table 5.4)—reflected observations published in the literature, where studies indicated different hydrogen-bonding behaviour for the Schiff base nitrogen: only interacting with E162[218]; interacting with water[221,227]; interacting with either D292 or water[148]; or forming hydrogen bonds with E162, D292 or water[206,222].

The analysis of the hydrogen-bonding network of the Schiff base showed that for proton transfer calculations of all-*trans* retinal in the closed/dark state of C1C2 E162, D292 or a water molecule are the only possible acceptors of the proton of the Schiff base nitrogen. We could further confirm the importance of E162, D292, K132 and E129 for the hydrogen-bonding network of the Schiff base region and that these residues should be included in the QM region of proton transfer calculations.

For systems containing 13-*cis*,15-*syn* retinal, S295 and T166 or water were the only possible hydrogen-bonding partners of the retinal Schiff base (Table 5.4). In the MM simulations, the Schiff base often hydrogen-bonded to the backbone of S295 (Fig. 5.9a) with visible pathway to connect to the extracellular side in case of Schiff base deprotonation. Consequently, simulation snapshots with the Schiff base interacting with backbone oxygens were not considered as starting snapshots to perform proton transfer calculations. Other interaction geometries with the Schiff base binding to the hydroxyl group of S295/T166 or water (Fig. 5.9) offered the option of concerted proton transfer via S295/T166 with D292 remaining the primary proton acceptor.

We note that, during MD simulations, we observed the isomerization of 13-*cis*,15-*anti* retinal to 13-*cis*,15-*syn* retinal (Fig. 5.9c). The isomerization of the C_{15}=N_{16} bond might have been an artifact of dihedral barrier that might be too low or might actually imply the presence of two states of 13-*cis* retinal, similar

dark-adaptation in bacteriorhodopsin[60,228], and needs further investigation.

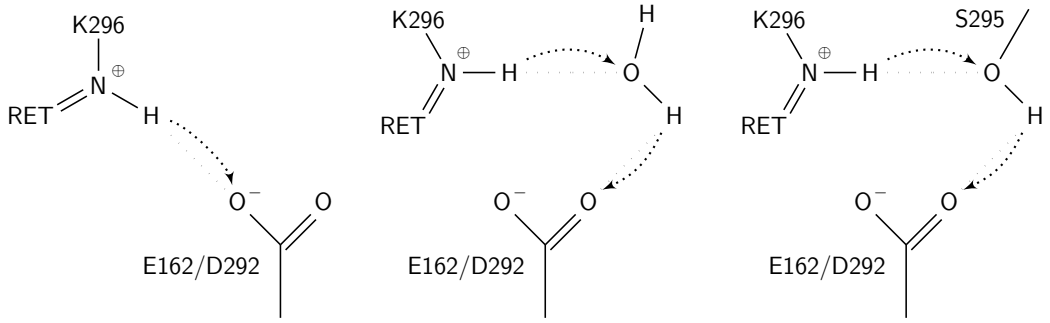
5.3 | Proton Transfer Pathways

We performed proton transfer calculations for wild type C1C2 and the mutant K132A in a hydrated POPC lipid membrane, and for the crystal structure in gas-phase (PDB ID: 3UG9[64]). All calculations were conducted for setups containing unprotonated and protonated E162. After MM equilibration, analysis of the MM trajectories indicated a complex hydrogen-bonding network that connected Schiff base, E162, D292 and K132 to each other (Fig. 2.9, Table 5.2 and Appendix Tables B.6, B.7, B.9 and B.10). The Schiff base of all-*trans* retinal was shown to interact with the negative counterion pair E162/D292 either directly or via water-bridging, and in setups with 13-*cis* retinal, the Schiff base was seen to hydrogen-bond to water or T166 or S295 (Fig. 2.9, Table 5.4 and Appendix Tables B.11 and B.12).

Based on the results of the MM simulations, we considered two potential acceptor groups for the Schiff base proton of all-*trans* retinal: E162 and D292. For these two potential acceptor groups, we computed direct proton transfer to the closest of the two (Pathway 1, Fig. 5.10a) or proton transfer via one water molecule that acted as intermediate carrier for the proton (Pathway 2, Fig. 5.10b). In the case of systems containing 13-*cis* retinal, we computed proton transfer pathways in which the hydroxyl groups of T166 or S295 were involved as intermediate proton carriers, similar to a water molecule (Pathway 3, Fig. 5.10c).

The Weighted Histogram Approach Method (WHAM) uses the partition function to derive the free energy from Umbrella sampling[161]. To avoid noisy potential of mean force (PMF) profiles the sampling bins should have sufficient overlap and sample enough points of the reaction coordinate[161,229]. Moreover, the individual bins should have a smooth shape, i.e. the force constant should be strong enough to produce a Gaussian distribution of the data points of the reaction coordinate. Usage of smaller force constants for the constraining potential can have the advantage that fewer bins might be needed for a reaction, but there is the possibility of not sampling unfavourable regions of the energy landscape.

To ensure the validity of the PMF profiles derived here, we checked the probability density distributions of the bins of all proton transfer calculations performed. We used a force constant k of 150 kcal/mol/Å^2 to constrain the individual bins, as test calculations had shown that this value was sufficient to force retinal deprotonation. For most of our proton transfer computations,



(a) Pathway 1 (b) Pathway 2 (c) Pathway 3

Figure 5.10: Schematic representation of proton transfer pathways. (a) Pathway 1: A direct proton transfer pathway involves a direct jump of the proton from the retinal Schiff base to the carboxylate acceptor group on either E162 or D292. (b) Pathway 2: In an indirect, water-mediated proton transfer pathway, the Schiff base proton is transferred to a water molecule, and a proton from the water molecule is subsequently transferred to the proton acceptor. (c) Pathway 3: For indirect proton transfer via S295, the Schiff base proton deprotonates to the hydroxyl group of S295, which in return loses its original hydrogen to one of the counterions E162/D292.

this force constant generated evenly distributed bins (Fig. 5.11a). Nevertheless, water-mediated proton transfer pathways sometimes failed to properly sample bins close to the transition state (Fig. 5.11b). This might have been caused by the more dynamic nature of a water molecule when compared to an amino acid residue and by the occurrence of concerted proton transfer, where no force constant was constraining the proton that the water released.

At this point, an advantage of WHAM came into play: namely, the method's ability to combine results from multiple simulations more easily than is the case with traditional Umbrella sampling. We performed additional sampling calculations for bins with problematic probability density distributions, and kept gradually increasing the force constant k, until the quality of the sample became adequate (Fig. 5.11c). Using bins with the increased force constants, we managed to improve the distribution close to the transition states (Fig. 5.11d).

Proton Transfer for Channelrhodopsin Containing All-*trans* Retinal

We generated two setups each for wild type C1C2 with unprotonated (simWTu) and protonated E162 (simWTp): one, where the QM region was located in monomer 1, and one, with the QM region in monomer 2. In monomer 1 of simWTu, we calculated direct proton transfer pathways to E162 (Pathway 1e, Table 5.5), and in monomer 2 of simWTu and in both monomers of simWTp, we calculated direct proton transfer pathways to D292 (Pathway 1d, Table 5.5). Indirect proton transfer calculations via water—no matter the system—always

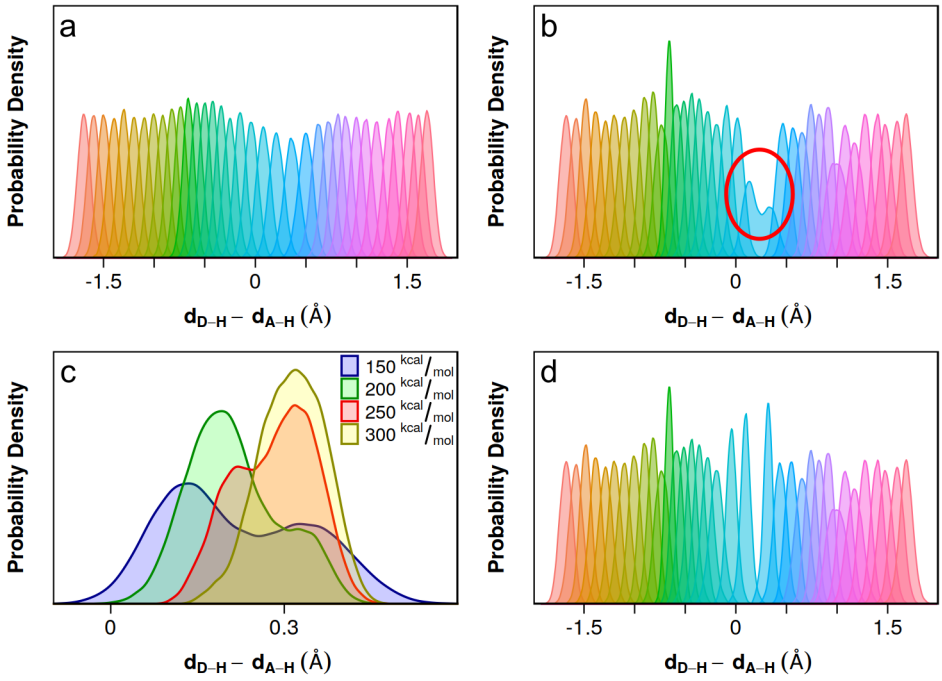


Figure 5.11: Checking the probability density distribution of bins generated for sampling the proton transfer with WHAM. (a) For the direct proton transfer pathway calculation in wild type C1C2 (SIMWTU), the constraining force constant used (150 kcal/mol/Å^2) was sufficient and the reaction coordinate was evenly sampled across the energy landscape. (b) For water-mediated proton transfer in the same setup, however, the constraint was not strong enough to generate a Gaussian distribution for bins at 0.3 Å close to the transition point (highlighted in red). (c) Increasing the force constants to 300 kcal/mol/Å^2 produced a well-behaved probability density distribution. (d) Using a higher force constant close to 0.3 Å allowed for satisfactory sampling of water-mediated proton transfer in wild type C1C2.

resulted in D292 being the final proton acceptor (Pathway 2D, Table 5.5). We found that proton transfer, either direct or concerted via one water molecule, was associated with significant energy barriers of ∼18 kcal/mol–23 kcal/mol and with unfavourable reaction energies of ∼16 kcal/mol–21 kcal/mol (Table 5.6, Fig. 5.12).

We note that, in the case of the direct proton transfer Pathways 1E and 1D, computed for SIMWTU and monomer 2 of SIMWTP, we could not obtain a clear valley for the free energy minimum of the product state (Table 5.6, Fig. 5.12). This could have been an effect of forcing the proton into a specific, potentially unfavourable geometry by directly applying the constraint to the Schiff base and the counterion. Longer sampling might have been necessary to allow for larger structural adaptation and equilibration of these pathways.

Water-mediated proton transfer pathways (Pathway 2D, Table 5.5), on the other hand, displayed very good energetic agreement for the reaction energy,

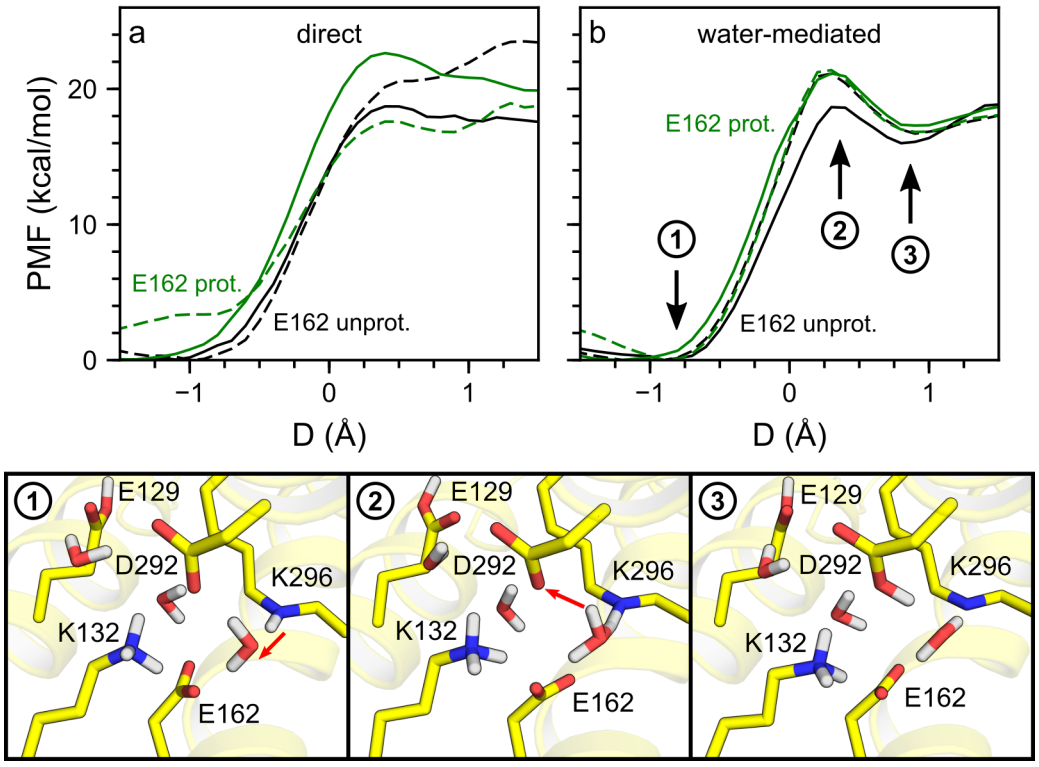


Figure 5.12: Direct (a) and water-mediated (b) proton transfer pathways in wild type C1C2 with unprotonated and protonated E162. PMF profiles computed inside monomer 1 and monomer 2 of the respective setup are plotted as solid and dashed lines, respectively. (a) Direct proton transfer for setups with unprotonated (black, simWTu) and protonated E162 (green, simWTp) had barrier heights of ∼18 kcal/mol–23 kcal/mol, and reaction energies of ∼17 kcal/mol–21 kcal/mol. Some pathways, like Pathway 1e in monomer 1 of simWTu (solid black) or Pathway 1d in monomer 1 of of simWTp (solid green) had only very shallow minima. (b) For the water-mediated proton transfer Pathway 2d, the PMF profiles showed good agreement. Monomer 2 of C1C2 with unprotonated E162 (black dashes, simWTu) and setups using protonated E162 (solid and dashed green, simWTp), in particular, matched very well, with barrier heights of ∼21 kcal/mol and reaction energies of ∼17 kcal/mol. The energy profile of monomer 1 of simWTu showed the same qualitative behaviour, but had both a lower barrier and reaction energy of 18.7 kcal/mol and 16.0 kcal/mol, respectively. Illustrations ①, ② and ③ are coordinate snapshots taken from the Pathway 2d computation with the QM region placed in monomer 1 of simWTu (Panel b, solid black) and indicate reactant, transition and product states, respectively.

and their free energy profiles showed a clear barrier followed by a minimum. In monomer 1 of wild type C1C2 with unprotonated E162 (simWTu), Pathway 2D had a barrier height of 18.7 kcal/mol and a reaction energy of 16.0 kcal/mol (Table 5.6). Pathway 2D in the other wild type setups had higher barriers, but otherwise showed good agreement between setups with barrier heights of 21.1 kcal/mol, 21.1 kcal/mol and 21.4 kcal/mol for monomer 2 of simWTu and monomer 1 and 2 of C1C2 with protonated E162 (simWTp), respectively. The reaction energy of Pathway 2D for monomer 2 of simWTu and monomer 1 and 2 of simWTp had the values 16.7 kcal/mol, 17.3 kcal/mol and 16.8 kcal/mol (Table 5.6, Fig. 5.12).

The high energy barriers and reaction energies involved confirmed the expectation that a protonated Schiff base is the clearly preferred state in the closed/dark state of C1C2.

Impact of the Lipid and Water Environment

After calculating reaction energies and barriers for direct and water-mediated proton transfer pathways with different protonation states of E162, we investigated the contributions the environment inside of the Schiff base region made to the free energy of the proton transfer reaction. We chose monomer 1 of the wild type C1C2 setup with unprotonated E162 (simWTu) and monomer 2 of the setup with protonated E162 (simWTp) to perform these additional computations because PMF profiles of these QM/MM setups had discernible energy minima even for proton transfer calculations of the direct Pathways 1E and 1D (Table 5.5, Fig. 5.12).

For unprotonated and protonated E162, we compared PMF profiles computed for the unmodified QM/MM setup (simWT, Fig. 5.13a), the PROT setup (Fig. 5.13a), the DRY setup (DRY, Fig. 5.13b), the K132A setup (Fig. 5.13c) and the DRY setup applied to K132A (K132A-DRY, Fig. 5.13d). The resulting energy profiles are summarized in Fig. 5.13e–f. (See subsection 'Proton Transfer Calculations' of Section 2.4 in the Methods chapter for more details on the derivation of the model systems.)

The lipid surrounding of the protein did not appear to significantly impact the results of the WHAM calculations using sampling bins of length 100 ps. For the indirect proton transfer Pathway 2D in the PROT setup with unprotonated E162, the barrier height decreased to 17.6 kcal/mol, while the reaction energy retained a value of 16.1 kcal/mol—almost identical to the complete setup (Table 5.6, Fig. 5.13e). For PROT with protonated E162, the barrier height decreased by 0.3 kcal/mol and the reaction energy increased by 0.6 kcal/mol, when compared

Table 5.5: Types of proton transfer pathways analysed.

Pathway	Retinal	Proton Transfer	Mediator	Final Acceptor
1E	all-*trans*	direct	—	E162
1D	all-*trans*	direct	—	D292
1D′	13-*cis*,15-*syn*	direct	—	D292
2E	13-*cis*,15-*anti*	indirect	water	E162
2D	all-*trans*	indirect	water	D292
3S	13-*cis*,15-*anti*	indirect	S295	D292

Table 5.6: List of proton transfer pathways and their respective reaction energies and barrier heights, computed for channelrhodopsin containing all-*trans* retinal.

Run	Pathway*	Protein	E162†	Lipids‡	SB Waters§	Energy (kcal/mol) Reaction	Barrier
1	1E	wild type	U	kept	kept	17.6	18.7
2	1E	wild type	U	deleted	kept	17.7	19.5
3	1E	wild type	U	kept	deleted	12.5	13.5
4	1D	wild type	U	kept	kept	20.6	20.6
5	1D	wild type	P	kept	kept	19.6	22.6
6	1D	wild type	P	kept	kept	16.8	17.6
7	1D	wild type	P	deleted	kept	15.5	16.3
8	1D	wild type	P	kept	deleted	19.8	19.9
9	1D	K132A	P	kept	kept	16.6	18.6
10	1D	K132A	P	kept	deleted	8.8	11.7
11	2D	wild type	U	kept	kept	16.0	18.7
12	2D	wild type	U	deleted	kept	16.1	17.6
13	2D	wild type	U	kept	deleted	11.4	16.5
14	2D	K132A	U	kept	kept	10.0	15.8
15	2D	K132A	U	kept	deleted	6.6	16.1
16	2D	wild type	P	kept	deleted	15.7	19.7
17	2D	wild type	U	kept	kept	16.7	21.1
18	2D	wild type	P	kept	kept	17.3	21.1
19	2D	wild type	P	kept	kept	16.8	21.4
20	2D	wild type	P	deleted	kept	17.4	21.1
21	2D	wild type	P	kept	deleted	19.8	20.8
22	2D	K132A	P	kept	kept	15.2	20.9
23	2D	K132A	P	kept	deleted	8.8	13.9

* See Table 5.5 for an explanation of pathway types.

† Protonation state of E162: unprotonated (U) or protonated (P).

‡ Refers to the POPC lipid bilayer.

§ Refers to water molecules in the Schiff base region.

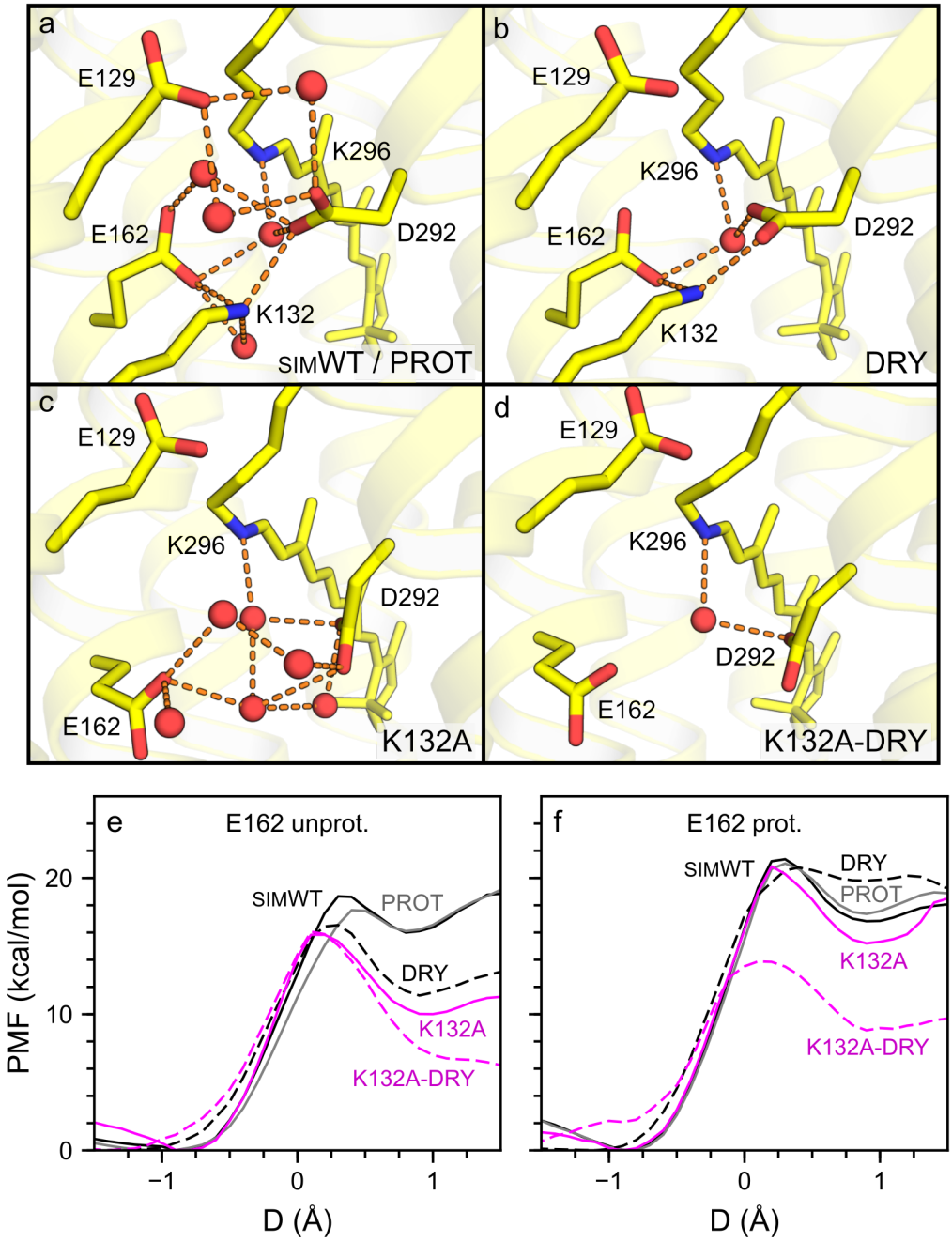


Figure 5.13: Contributions from the lipid, protein and water environment to the proton transfer energetics. All computations were performed for concerted proton transfer via one water molecule in C1C2 containing all-*trans* retinal (Pathway 2D, Table 5.5). To decompose energetic contributions, we generated PMF profiles for different environments, which are illustrated in the four coordinate snapshots: the complete wild type setup (SIMWT, a), the PROT setup (a), the DRY setup (b), the complete K132A mutant setup (c) and the K132A-DRY setup (d). The PMF profiles show results for the complete (solid black), PROT (grey), DRY (black dashes), K132A (solid magenta) and K132A-DRY setups (magenta dashes), for systems using either unprotonated (e) or protonated (f) E162. Note that hydration, E162 protonation and the presence of K132 contribute to the stability of the proton on the retinal Schiff base.

to the complete SIMWTP setup (Table 5.6, Fig. 5.13f). The influence of the lipid bilayer on the QM region might have been limited by the length of the WHAM bins and the usage of cut-offs for SCC-DFTB.

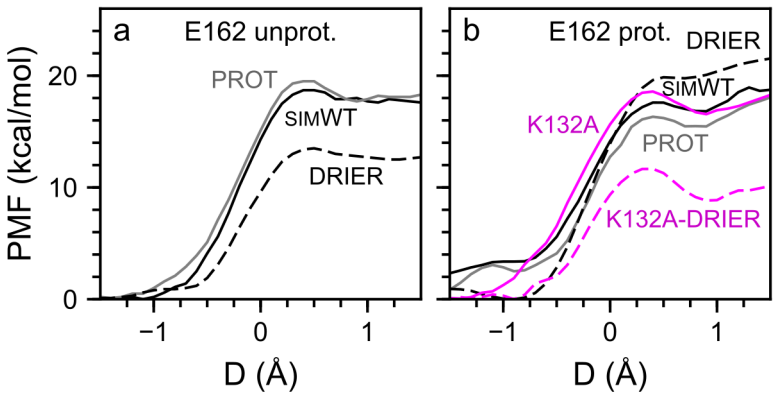


Figure 5.14: Contributions to the free energy for direct proton transfer. A black line indicates PMF profiles derived for the complete wild type system using a hydrated POPC lipid bilayer and periodic boundary conditions (SIMWT setup). A grey line shows PMF profiles after lipids and bulk water had been deleted (PROT setup). Deletion of Schiff base water molecules is indicated by a dashed black line (DRIER setup), the mutation K132A is shown in magenta and deletion of Schiff base water molecules in K132A is indicated by a dashed magenta line (K132A-DRIER setup). Results are shown for direct proton transfer Pathways 1E and 1D (Table 5.5) in setups using unprotonated (a, SIMWTU) and protonated E162 (b, SIMWTP), respectively. We note that for direct proton transfer in K132A with unprotonated E162 no pathway calculations could be completed.

For water-mediated proton transfer, deleting water molecules from inside the Schiff base region more strongly affected the proton transfer energetics than deletion of the POPC lipid bilayer and bulk water in the PROT setup. The barrier of the proton transfer reaction was lowered, to 16.5 kcal/mol for unprotonated E162 and to 20.8 kcal/mol for protonated E162 (Table 5.6, Fig. 5.13e–f). In the DRY setup with unprotonated E162, the reaction energy decreased to 11.4 kcal/mol, whereas for protonated E162 the DRY setup resulted in a less favourable interaction energy of 19.8 kcal/mol.

The PMF profiles for direct proton transfer Pathways 1E and 1D showed a similar contribution pattern (Table 5.6, Fig. 5.14a–b): Removal of the lipid membrane and bulk water had only a small influence; and deletion of Schiff base water decreased the reaction energy for setups using unprotonated E162 (SIMWTU), while increasing it for setups using protonated E162 (SIMWTP). For SIMWTU, the PROT setup had an increased barrier height of 19.5 kcal/mol and almost no change in reaction energy with a value of 17.7 kcal/mol (Table 5.6, Fig.

5.14a); whereas for SIMWTP, both the barrier height and the reaction energy were lowered by 1.3 kcal/mol and 1.4 kcal/mol, respectively (Table 5.6, Fig. 5.14b). The barrier height and reaction energy of the DRIER setups with removed Schiff base water molecules were lowered by 5.2 kcal/mol and 5.1 kcal/mol, respectively, for SIMWTU (Fig. 5.14a) and increased by 2.3 kcal/mol and 3.0 kcal/mol higher for SIMWTP (Table 5.6, Fig. 5.14b).

We found that, for unprotonated E162, hydration of the Schiff base region contributed to the stability of the proton on the Schiff base nitrogen. The stabilizing effect hydration can have on the retinal Schiff base had been described before[49,159,230,231], and in bacteriorhodopsin, the presence of water in the vicinity of the retinal Schiff base can have a large impact on retinal deprotonation[37,48,52]. Additionally, for the protonated state of E162, the PMF calculations implied a very strong stabilizing effect of protonated E162 on the Schiff base that outweighed the stabilizing effect that hydration of the Schiff base region had. These observations showed strong similarity to bovine rhodopsin, where threonine-94 (T94) together with a water molecule interacting with glutamate-113 (E113) stabilizes the Schiff base[232,233,234].

To further test the impact of E162 protonation, we started from the DRY setup with unprotonated E162 for water-mediated Pathway 2D calculations (Table 5.5) and changed the protonation state of E162 by placing on its carboxylate group a proton, which had been added to the system. We then recalculated the PMF

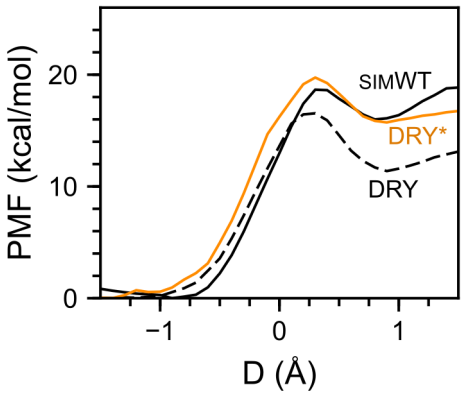


Figure 5.15: Contribution of E162 protonation to proton transfer energetics for the water-mediated Pathway 2D in wild type C1C2. Starting from the wild type setup with unprotonated E162 (SIMWTU, solid black), the DRY setup was derived by deleting water molecules from the Schiff base region (black dashes). A proton was added to E162 of the DRY setup, before repeating the water-mediated proton transfer calculations (DRY*, solid orange). The change of protonation state of E162 negated the effect the deletion of the water molecules had and increased the reaction energy and barrier height to 15.7 kcal/mol and 19.7 kcal/mol, respectively.

profile for Pathway 2D using this new setup (DRY*). Protonating E162 completely negated the effect of the dehydration and increased the barrier height and reaction energy to 19.7 kcal/mol and 15.7 kcal/mol, respectively (Table 5.6, Fig. 5.15). The results of this protonation test were added evidence for the strong stabilizing effect protonation of E162 has on the Schiff base proton.

K132 As an Important Determinant of Proton Transfer Energetics

As a next step, we analysed the effect removal of the positively charged K132 would have on the proton transfer energetics.

For the setup with unprotonated E162, PMF computations of K132A for the water-mediated proton transfer Pathway 2D exhibited a large reduction in both barrier height and reaction energy, which, relative to the unmodified wild type setup (SIMWTU), had decreased by 2.8 kcal/mol and 6.0 kcal/mol, respectively (Fig. 5.13e). Removing the water molecules from inside the Schiff base region of the K132A setup (K132A-DRY, Fig. 5.13d) led to an additional decrease in reaction energy by 3.4 kcal/mol to 6.6 kcal/mol and made the minimum less pronounced.

For Pathway 2D in setups containing protonated E162, the difference in reaction energy between K132A and the wild type—a drop by 1.6 kcal/mol—was lower than had been the case for the setups with unprotonated E162. Moreover, with 20.9 kcal/mol, the barrier height of K132A stayed close to the wild type value (Table 5.6, Fig. 5.13f). The reaction energy and barrier of the K132A-DRY setup had a value of 8.8 kcal/mol and 13.9 kcal/mol, respectively, and mirrored a similar sharp decline in reaction energy as observed for unprotonated E162 setups (Table 5.6, Fig. 5.13e–f).

For direct proton transfer, we only succeeded in computing PMF profiles for Pathway 1D in the K132A mutant setup with protonated E162. Pathway 1E calculations for K132A containing unprotonated E162 did not converge.

Computing direct proton transfer in K132A with protonated E162 increased the reaction barrier by 1.0 kcal/mol, but left the reaction energy unchanged, when compared to the wild type value (Table 5.6, Fig. 5.14b). The dehydrated mutant setup, K132A-DRIER, with no water molecules in the Schiff base region, lowered both reaction energy and barrier to 8.8 kcal/mol and 11.7 kcal/mol, respectively (Table 5.6).

The positive K132, like hydration of the Schiff base pocket, stabilized the proton on the Schiff base nitrogen, and like hydration, protonation of E162 appeared to be the more dominant factor compared to the effects of K132. For

unprotonated E162, the absence of K132 and a dehydrated Schiff base region caused a drop in reaction energy of almost 10 kcal/mol, and these two determinants seemed to account to a large part for the stability of the retinal Schiff base proton. For protonated E162, removal of either water or K132 did not produce the same effect as in the unprotonated system, but once both water and K132 had been eliminated from the setup, we observed a similar big drop in reaction energy. Consequently, to effectively stabilize the Schiff base proton, a protonated E162 seems to need either K132 or water molecules to bridge the interaction between E162 and D292.

Proton Transfer in the Crystal Setup

When deleting Schiff base water molecules to derive the DRY and DRIER setups, we were manipulating the QM region. To verify the effects the hydration had on the proton transfer energetics using an independent approach and to ensure that the observations made were not artefacts, we performed proton transfer calculations for a gas-phase monomer of the C1C2 crystal structure (PDB ID: 3UG9[64]).

We chose D292 as acceptor—conducting only direct proton transfer—and generated PMF profiles for unprotonated vs protonated E162 and for an unhydrated interior vs one with DOWSER-generated water molecules (DOWSER setup). For the unhydrated setup, we removed the single crystal water that was located close to the Schiff base region as it was too far away from the Schiff base nitrogen to be considered as intermediate proton acceptor and to reduce the number of parameters influencing the PMF profile.

In the unhydrated setup with unprotonated E162, the barrier height and reaction energy were 7.2 kcal/mol and 6.1 kcal/mol, respectively (Table 5.7, Fig.

Table 5.7: List of proton transfer pathways and their respective reaction energies and barrier heights, computed for channelrhodopsin crystal setups with all-*trans* retinal.

Run	Pathway*	Protein	E162†	Lipids‡	SB Waters§	Energy (kcal/mol) Reaction	Barrier
24	1D	wild type	U	none	deleted	6.1	7.2
25	1D	wild type	U	none	DOWSER	19.0	19.0
26	1D	wild type	P	none	deleted	17.9	18.1
27	1D	wild type	P	none	DOWSER	17.9	19.4

* See Table 5.5 for an explanation of pathway types.

† Protonation state of E162: unprotonated (U) or protonated (P).

‡ Refers to the POPC lipid bilayer.

§ Refers to water molecules in the Schiff base region.

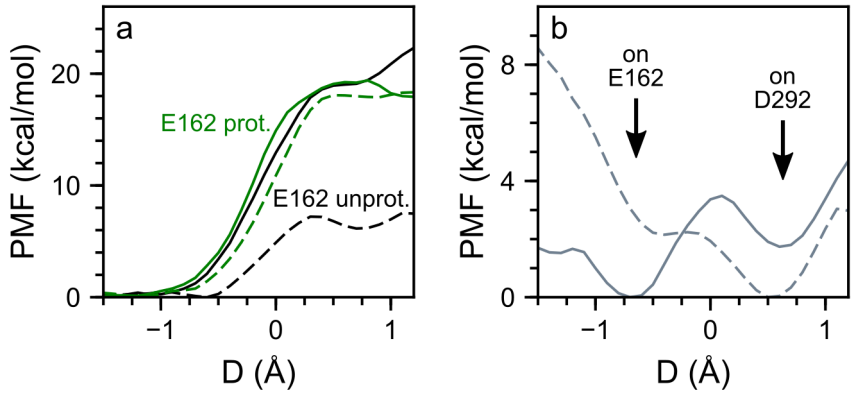


Figure 5.16: Proton transfer calculations for a crystal structure monomer of C1C2 (PDB ID: 3UG9[64]) in the gas-phase. (a) Direct transfer of the Schiff base proton to D292 (Pathway 1D). We computed free energies of retinal deprotonation for the crystal structure with unprotonated (black) and protonated E162 (green). We then compared the energies to PMF profiles obtained for the crystal structure that had been hydrated with DOWSER. Unhydrated and DOWSER setups are indicated by dashes and solid lines, respectively. In the crystal setup, too, protonated E162 had a strong stabilizing effect on the stability of the Schiff base proton. (b) Proton transfer from E162 to D292. To test if protonation of E162 was preferred over D292 protonation, we performed proton transfer calculations from E162 to D292 using a dehydrated system (dashes) and a DOWSER hydrated system (solid line). Hydration changes the protonation preference form D292 to E162. The Schiff base was kept protonated during these computations.

5.16a). For protonated E162, the barrier had a height of 18.1 kcal/mol and was followed by a shallow minimum with a reaction energy of 17.9 kcal/mol (Table 5.7, Fig. 5.16a). The DOWSER setups behaved similarly for both unprotonated and protonated E162 and had a barrier height of ∼19 kcal/mol and a reaction energy of 19.0 kcal/mol and 17.9 kcal/mol for unprotonated and protonated E162, respectively (Table 5.7, Fig. 5.16a).

The results of the crystal computations showed the same hydration effects already observed for wild type setups (SIMWT) and K132A (SIMM1): If E162 was protonated, then hydration of the Schiff base region had no significant effect on the proton transfer energetics. Likewise, for unprotonated E162, an unhydrated Schiff base region had a less stable Schiff base proton than a hydrated Schiff base region. By reproducing the outcomes of the dimer calculations, we confirmed the validity of the conclusions made about the hydration.

We performed an additional test for the crystal structure with protonated E162. We conducted proton transfer calculations from E162 to D292 to verify if there was any preference regarding the protonation of these Schiff base counterions. For the unhydrated setup, the barrier height had a value of 0.1 kcal/mol, and the resulting reaction energy was −2.2 kcal/mol. This meant that placing the

proton on D292 was more favourable than placing it on E162 (Fig. 5.16b). Once missing waters had been added with DOWSER, this was reversed and moving the proton from E162 to D292 had a barrier height of 3.5 kcal/mol and a reaction energy of 1.7 kcal/mol (Fig. 5.16b). These results agreed with the observation that, during our QM/MM simulations with protonated E162, the proton would remain on E162. The function of D292 as primary proton acceptor, however, implies that the presence of the positive Schiff base might play a role in keeping the proton on E162 since in our proton transfer calculations water-mediated Schiff base deprotonation always resulted in D292 protonation.

Proton Transfer for Channelrhodopsin Containing 13-*cis* Retinal

We investigated snapshots of the QM/MM trajectories containing 13-*cis* retinal to identify possible starting geometries to use in proton transfer calculations. This was a non-trivial task as the information provided by the analysis of the hydrogen-bonding network was not conclusive.

For some setups, e.g. monomer 1 of the setup with unprotonated E162 and retinal isomerized in one monomer only (simW1Cu), no hydrogen-bonding partner existed (Fig. 5.9d). In this particular example, we placed a force constant on S295, the closest possible acceptor, and computed proton transfer to S295, despite this residue being located too far away to engage in direct hydrogen-bonding with the Schiff base. For other setups, hydrogen-bonding partners existed, but the resulting hydrogen-bonding network did not connect to the extracellular side of the Schiff base region. In monomer 1 of the setup with a single isomerized retinal and protonated E162 (simW1Cp), for example, the Schiff base interacted with the backbone of S295 (Fig. 5.9a), and in monomer 2 of the setup with both

Table 5.8: List of proton transfer pathways and their respective reaction energies and barrier heights, computed for channelrhodopsin containing 13-*cis* retinal.

Run	Pathway*	Protein	E162†	Lipids‡	SB Waters§	Energy (kcal/mol) Reaction	Barrier
28	3s	wild type	u	kept	kept	23.3	31.3
29	2e	wild type	u	kept	kept	24.6	28.2
30	3s	wild type	u	kept	kept	29.7	34.8
31	1d′	wild type	u	kept	kept	17.5	21.5

* See Table 5.5 for an explanation of pathway types.

† Protonation state of E162: unprotonated (u) or protonated (p).

‡ Refers to the POPC lipid bilayer.

§ Refers to water molecules in the Schiff base region.

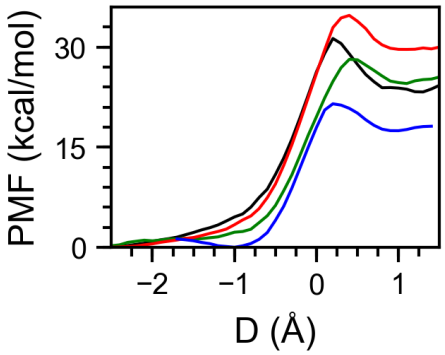


Figure 5.17: Proton transfer in C1C2 with 13-*cis* retinal. We performed proton transfer calculations for C1C2 containing isomerized retinal. For setups containing 13-*cis*,15*anti* retinal, we computed PMF profiles for Schiff base proton transfer to D292 via the hydroxyl group of S295 (Pathway 3S) using unprotonated (black) and protonated E162 (red) and for water-mediated Schiff base proton transfer to E162 (Pathway 2E) using an unprotonated E162 system (green). We additionally computed direct proton transfer from the Schiff base to D292 (Pathway 1D′) for a setup with 13-*cis*,15-*syn* retinal.

retinals isomerized and E162 unprotonated (SIMW2CU), there was a hydrogen bond with T166 that in return was hydrogen-bonding with the backbone of E162 (Fig. 5.9e). We tested proton transfer for the SIMW2CU setup with the T166 interaction, but were unable to achieve T166-mediated proton transfer.

Eventually, we identified four setups where proton transfer was possible: (i) monomer 1 of SIMW1CU (one retinal isomerized, E162 unprotonated) with concerted proton transfer to D292 via the hydroxyl group of S295 (Pathway 3S), (ii) monomer 1 of SIMW1CP (one retinal isomerized, E162 protonated) with the same S295-mediated pathway to D292 (Pathway 3S), (iii) monomer 1 of SIMW2CU (both retinals isomerized, E162 unprotonated) with water-mediated proton transfer and E162 as final proton acceptor (Pathway 2E) and (iv) monomer 1 of SIMW2CP (both retinals isomerized, E162 protonated) that contained 13-*cis*,15-*syn* retinal with direct proton transfer to D292 (Pathway 1D′). With the exception of pathway 2E, D292 was the primary proton acceptor for all 13-*cis* proton transfer calculations (Table 5.8).

Pathway 1D′ had the Schiff base nitrogen of the 13-*cis*,15-*syn* retinal form a direct hydrogen bond with D292, and we calculated a barrier height of 21.5 kcal/mol and a reaction energy of 17.5 kcal/mol—values similar to the results for the all-*trans* setups (Tables 5.6 and 5.8). Pathways 2E and 3S, on the other hand, had much higher barriers and reaction energies. The water-mediated Pathway 2E had a reaction energy of 24.6 kcal/mol and a barrier height of 28.2 kcal/mol, while for the setups SIMW1CU and SIMW1CP, the S295-mediated Pathway 3S yielded

reaction energies and barrier heights between of ∼23 kcal/mol–30 kcal/mol and ∼31 kcal/mol–35 kcal/mol, respectively (Table 5.8, Fig. 5.17).

The very high reaction energies and barriers found in the PMF profiles for the proton transfer calculations could mean that the Pathways 2E and 3S are not the correct proton transfer pathways, or it could indicate a problem with the 13-*cis* model employed during our calculations: We isomerized the retinal by applying a simple harmonic potential without performing any excited state calculations, and the 100 ns time scale of the simulation might not have been long enough for the system to equilibrate and to sample large scale conformational changes. In the case of channelrhodopsin-2, for instance, helix two, which contains the residues E129 and K132, had been shown to reorient after retinal isomerization, thus possibly leading to the opening of the ion channel[235,236,237].

Nevertheless, we discovered the new possible proton transfer Pathway 3S for C1C2 containing 13-*cis*,15*anti* retinal: proton transfer via S295 to D292 (Table 5.5). A second potential proton transfer pathway involved the Schiff base deprotonating to T166. Hydrogen-bonding between Schiff base and T166 upon retinal isomerization has been proposed for channelrhodopsin-2[67]; in our computations, however, we could not identify a T166-mediated pathway that could lead to proton transfer to the counterion pair E162/D292. Water-mediated proton transfer still remained possible for 13-*cis* retinal.

For 13-*cis*,15*anti* retinal, the Schiff base nitrogen was facing away from E162/D292. In some 13-*cis* simulations, we could observe a spontaneous conformational change to 13-*cis*,15*syn* retinal, where interaction with the counterions E162/D292 was reestablished and the proton transfer Pathway 1D′ became possible (Table 5.5).

Quoth he,—These reasons are but strains
Of wanton, over-heated brains
Which ralliers, in their wit, or drink,
Do rather wheedle with than think

Samuel Butler, *Hudibras*

6 Conclusions

We investigated proton transfer pathways for deprotonation of the retinal Schiff base in the dark state of the channelrhodopsin chimæra C1C2 (PDB ID: 3UG9[64]) by using the Weighted Histogram Approach Method (WHAM). For proton transfer to D292/E162, either directly or water-mediated, we showed that the position of the proton on the Schiff base was very stable with barrier heights of ∼18 kcal/mol–23 kcal/mol (Table 5.6, Figs. 5.12, 5.13, 5.14 and 5.15). Furthermore, we found three important determinants that contribute to the stability of the proton of the retinal Schiff base. These determinants were the presence of the positive amino acid residue K132, water molecules in the vicinity of the Schiff base and protonation of E162.

To identify suitable snapshots to generate models for the proton transfer calculations and to better characterize the interactions inside the Schiff base region, we performed extensive analysis of molecular mechanics (MM) and combined quantum and classical mechanics (QM/MM) molecular dynamics (MD) simulations of wild type C1C2 and of the mutants K132A, E129A and E162D. To our knowledge, our work represents the first use of a dimeric C1C2 setup in MD simulations. Throughout our simulations, the alpha-helical regions of the protein stayed close to the crystal structure and we saw a very hydrated protein interior (Fig. 5.3), consistent with observations made in other studies on channelrhodopsin[148,158,220,221,222].

In repeat simulations of wild type C1C2, in the mutant E162D and in setups containing 13-*cis* retinal, we sometimes observed channel opening and water passage between the extracellular and intracellular side of the membrane (Fig. 5.5). Opening of the channel always required the absence of the E129–N297 hydrogen bond (Fig. 5.6). For all other systems, E129–N297 was preserved

throughout the trajectories, in agreement with other studies made[41,206]. Thus, our simulations indicated that a permanent E129–N297 hydrogen bond might not be necessary to prevent ions from crossing during the dark state and further supported the notion that cation passage would be inhibited electrostatically by the presence of the protonated E129[150,151,156,226]—rather than E129–N297.

Ions were absent from our simulations, and we could only analyse the open channel in terms of water passage. Future work on the central gate region and the E129–N297 interaction in particular should include ions to be able to measure ionic flux and to calculate ion distributions.

Analysis of the hydrogen-bonding networks inside of the Schiff base region (Fig. 2.9) revealed strongly interconnected interactions between the retinal Schiff base, E162, D292, E129 and K132, thereby justifying their inclusion in the quantum mechanics (QM) region of our QM/MM computations.

Evaluation of the hydrogen-bonding data appeared to favour negative E162. Both unprotonated and protonated E162 were able to reproduce the E162–D292 separation observed in the crystal structure[64] (Fig. 5.8); however, setups with protonated E162 ran the risk of underestimating the E162–D292 distance when both residues formed a direct hydrogen bond. We discovered a correlation between the K132–D292 distance and the number of water contacts for the Schiff base nitrogen and the counterions E162/D292 (Table 5.3 and Fig. 5.4). This correlation supported a possible controlling/gating function K132 might have[207] and was absent for protonated E162, providing additional support for a negatively charged E162. It might be interesting to test if protonation of D292 would break the correlation in the same manner E162 protonation did.

The retinal Schiff base of all-*trans* retinal only had E162, D292 or water as possible hydrogen-bonding partners (Table 5.4), which matched experimental and theoretical studies of the Schiff base interactions[148,206,218,221,222,227], as well as indications that D292 is the primary proton acceptor in channelrhodopsin[64,150]. For 13-*cis*,15-*anti* retinal, we observed hydrogen bonding between the Schiff base and S295, T166 or water (Table 5.4). For some simulations, we observed 13-*cis*,15-*anti* retinal isomerizing to 13-*cis*,15-*syn*, where the hydrogen-bonding partners were the same as for all-*trans* retinal (Table 5.4).

The acceptor atoms of the proton transfer pathways were decided based on the results of the MM-analysis of the Schiff base hydrogen-bonding network. We generated potential of mean force (PMF) profiles for direct deprotonation to D292/E162 or for the water-mediated Schiff base deprotonation (Table 5.5, Fig. 5.10a–b). For water-mediated proton transfer Pathway 2D, D292 emerged as sole primary proton acceptor, and the resulting reaction energies of the different

setups were in good agreement with each other (Tables 5.5 and 5.6).

We introduced reduced systems to dissect the influence of the lipid, protein and water environment on the Schiff base deprotonation energetics: the PROT setup, without lipids and bulk water; the DRY/DRIER setups with water deleted from the Schiff base region; a setup using the mutant K132A; and the K132-DRY/DRIER setup, with Schiff base water deleted from K132A (Fig. 5.13a–d).

We found that the lipid bilayer had a negligible impact on proton transfer energetics—at least on the timescales of our proton transfer calculations—and that hydration of the Schiff base region, the presence of K132 or a proton on E162 were large stabilizers for the Schiff base proton in the dark state of channelrhodopsin. For systems, containing unprotonated E162, removing water and/or K132 reduced the reaction energy by up to ∼10 kcal/mol (Fig. 5.13e). For protonated E162, as long as either water or K132 remained present, the reaction barrier was preserved; however, removing both water and K132 resulted in a similar drop in reaction energy as observed for systems with unprotonated E162 (Table 5.6, Figs. 5.13f and 5.14a–b). Likewise, placing a proton on E162 in the DRY setup completely restored the proton transfer energetics of the hydrated system (Fig. 5.15).

Complementing computations of proton transfer in crystal structure setups showed the same dependence on hydration as demonstrated for setups based on dimeric C1C2 simulations (Table 5.7, Fig. 5.16a): Protonation of E162 stabilized the Schiff base and dehydration lowered the reaction barrier only for setups with unprotonated E162. When maintaining a protonated Schiff base, hydrating the Schiff base region with DOWSER made placing a proton on E162 more favourable than placing it on D292.

Proton transfer calculations of 13-*cis* retinal failed to discover a pathway with a lower-energy product state. For 13-*cis*,15-*anti*, we detected possible proton transfer via S295 (Pathway 3S, Fig. 5.10c). The high reaction barriers of the 13-*cis* computations (Table 5.8) might have indicated that further equilibration of the systems containing isomerized retinal had been necessary, and prolongation of the 13-*cis* trajectories to at least 250 ns should be considered for future studies.

In the end, we demonstrated that the retinal Schiff base needs a hydrated environment to make continued Schiff base protonation in the dark state of channelrhodopsin more favourable. We also showed that K132 contributes to the stabilization of the Schiff base and that a neutral E162 still needs either K132 or a hydrated Schiff base region to effectively stabilize the Schiff base proton. These results revealed parallels to bacteriorhodopsin[52] and bovine rhodopsin[232,233], where the Schiff base was stabilized by either the Schiff base itself or a glutamate counterion interacting with water molecules or a nearby threonine residue.

Part Four

General Perspectives

遠山に
日の当たりたる
枯野かな

On the distant mountains
Shines the sun:
Withered field.

Takahama Kyoshi

Two distinct theoretical approaches to studying cofactors of proteins have been demonstrated. In Part II, new CHARMM force field parameters were developed to describe iron-containing cofactor complexes found in photosystem II. The new parameters were then tested in classical molecular dynamics (MD) of neuroglobin. In Part III, the retinal cofactor of the channelrhodopsin-1–channelrhodopsin-2 chimæra (C1C2) was studied using molecular mechanics (MM) and quantum mechanics (QM). We first derived MM-equilibrated systems and based the initial analysis of the interactions between retinal and its protein surrounding on MM. We then used combined quantum and classical mechanics (QM/MM) to dissect the effects the retinal environment had on the stability of the Schiff base proton in the dark state of C1C2.

Compared to QM, MM enables much longer simulations, with trajectory timescales in the hundreds of nanoseconds and up to the microsecond range readily accessible[238]. Classical all-atom force fields, however, are no universal cure-all as they are limited to the effect that chemical reactions, like bond breaking, cannot be described and protonation states and partial charges need to be set before the start of a simulation. Moreover, if parameters are not available, a tedious parametrization procedure—as performed in here—becomes necessary.

QM calculations, on the other hand, completely avoid the issue of needing to parameterize specific molecules, but are computationally much more expensive. QM/MM setups are also more difficult to set up, for instance some possible proton transfer pathways generated for Part III did not converge and no potential of mean force (PMF) profiles could be produced. Depending on the type of QM calculation, e.g. for frequency computations, the choice of link atom scheme can play an important role[113] and needs to be taken into consideration.

Presently, MM MD calculations are more commonly encountered than QM MD or QM/MM MD—mainly due to constraints imposed by computational cost. There are efforts to incorporate QM-like properties into classical MD to enable, for instance, explicit proton transfer[239] or polarizability[240,241]. In the end, these MM approaches still require parameters, and even automation of the parameterization process[184,185] will eradicate neither the need for force field optimization nor the issue of the transferability of force field parameters. To be less dependent on rigid choices of protonation states and for sampling real reaction cycles or at least parts of such cycles, the future has to lie in QM/MM MD.

We showed here that large QM/MM MD setups containing more than 100 000 atoms in a periodic cell can be implemented and that for these setups, even within the current limits, simple proton transfer reactions can be computed. To this effect, the channelrhodopsin QM/MM work presented in Part III provides a protocol for analysing proton transfer in setups identical to conventional MM MD setups. Our efforts also demonstrated how to dynamically decompose contributions from the environment surrounding a QM residue of interest by removing suspected influencing factors, like specific amino acid residues or water, without the need for minimum pathway calculations. Our approach can be applied to similar systems, such as the new channelrhodopsin-2 crystal structure[67].

The QM/MM calculations performed here took only twice as long as pure MM calculations using the same hardware would have taken. And exactly herein lies the catch: self-consistent charge tight-binding density functional theory (SCC-DFTB), although fast, is not parallelized. Whereas for MM systems, the number of processors and compute nodes can be increased, thus significantly lowering the computational cost, such an increase would not affect the QM part. Faster CPUs already allow for longer timescales, and even without parallelized SCC-DFTB, we were able to generate 1 ns QM/MM trajectories for our complete system setups in a hydrated lipid membrane. The computation speed has to grow by two orders of magnitude for QM/MM to be competitive with current-day MM, and for this, more efficient parallelization of QM methods needs to be achieved and should be the main focus of current methods development.

الباب الذي يأتيك بالريح
سده واستراح

Part Five

Appendix

Parametrization Work A

A.1 | Topology and Partial Charges

This section contains the atom names, chemical types and partial charges included in the CHARMM patch for HAEMHIS, FEBIHIS and FEHIS[179].

In CHARMM, each atom of a residue needs to have a unique identifier, the so-called atom names. When a residue is loaded in CHARMM, the program assigns each atom a partial charge and a chemical type and constructs the topology of the residue based on the atom names. The file containing all these parameters is called the topology file.

To avoid overwriting parameters already present in other CHARMM force fields, new chemical types were introduced for the irons found in the different complexes. All new iron chemical types use the iron non-bonded parameters of the hæm iron from the hæm extension[118,119] to the CHARMM36 protein force field[97].

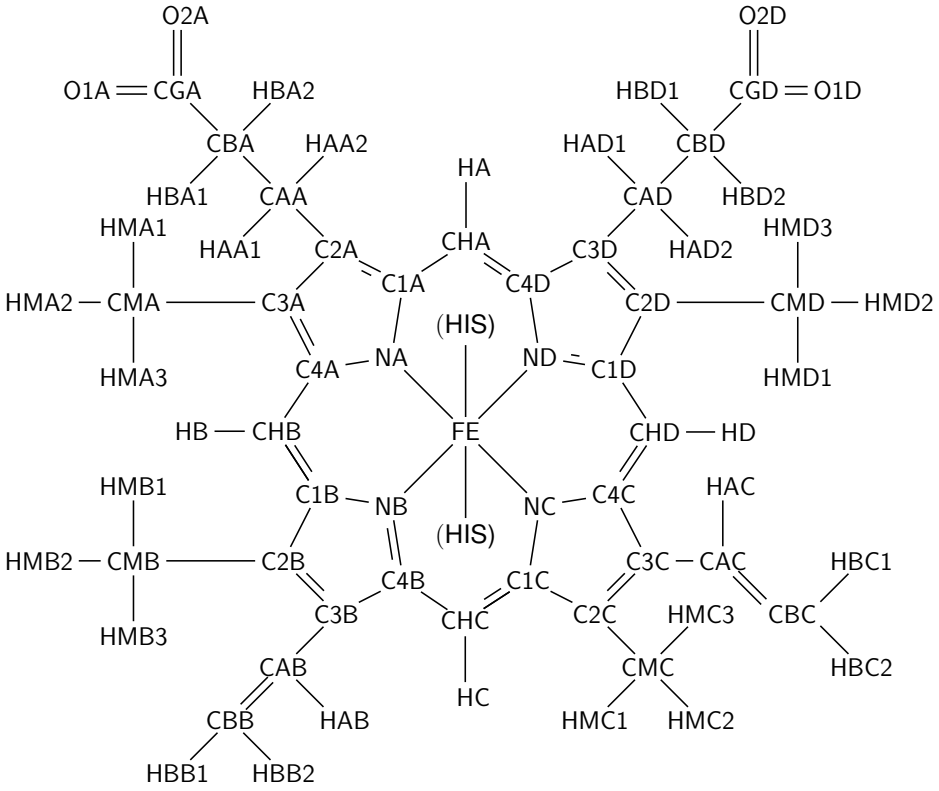


Figure A.1: Structure and atom names of hæm b as part of HAEMHIS. The CHARMM atom names of hæm b are shown[118,119]. When using HAEMHIS in patch form[179], the atom names need to be preceded by the number '1' (e.g. '1NA' instead of 'NA').

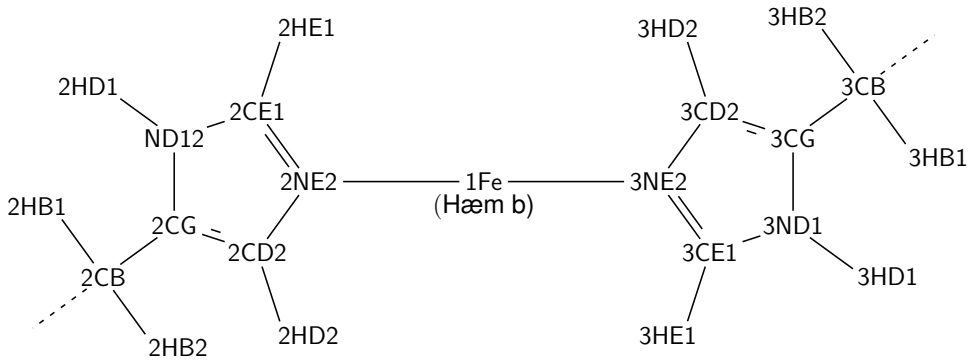


Figure A.2: Structure and atom names of the two histidine side chains that form part of HAEMHIS. The atom names are shown for the two histines binding to the hæm iron in the HAEMHIS complex in its patch form[179].

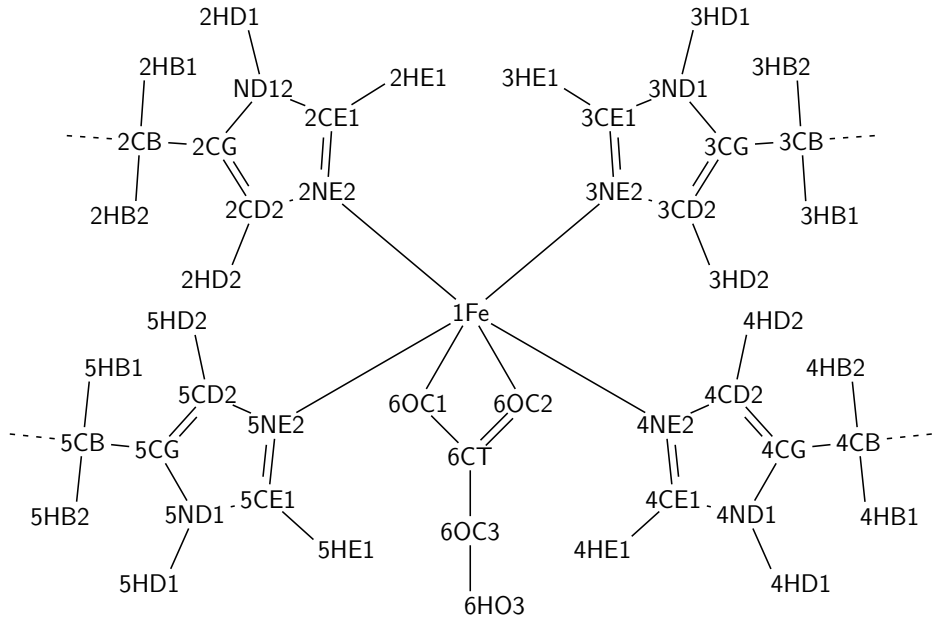


Figure A.3: Structure and atom names of the FEBIHIS Complex. The atom names are shown for FEBIHIS, when applied as a patch[179].

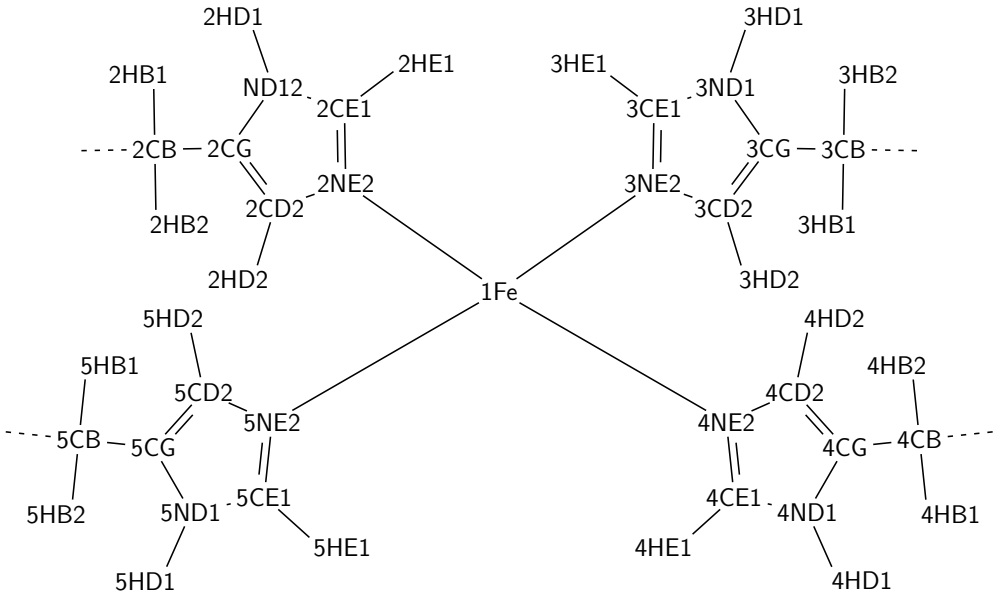


Figure A.4: Structure and atom names of the FEHIS complex. The atom names are shown for FEHIS, when applied as a patch[179].

Table A.1: Overview over the newly introduced CHARMM chemical types used for iron in HAEMHIS, FEBIHIS and FEHIS.

Chemical Type	Complex	Iron Type
FE2H0	HAEMHIS	ferrous hæm iron*
FE2H	HAEMHIS	ferrous hæm iron†
FE3H	HAEMHIS	ferric hæm iron
FE2NH1	FEBIHIS	ferrous non-hæm iron
FE3NH1	FEBIHIS	ferric non-hæm iron
FE2NH2	FEHIS	ferrous non-hæm iron
FE3NH2	FEHIS	ferric non-hæm iron

*HAEMHIS with CHARMM partial charges

†HAEMHIS with optimized partial charges

Table A.2: Chemical types and final CHARMM atomic partial charges for the ferrous HAEMHIS patch.

Atom Name	Chem. Type	Final Charge	Atom Name	Chem. Type	Final Charge	Atom Name	Chem. Type	Final Charge
1FE	FE2H	0.24	1C1D	CPA	0.12	2HB2	HA2	0.09
1NA	NPH	−0.18	1C2D	CPB	−0.06	2NE2	NR2	−0.7
1NB	NPH	−0.18	1C3D	CPB	−0.06	2CD2	CPH1	0.22
1NC	NPH	−0.18	1C4D	CPA	0.12	2HD2	HR3	0.1
1ND	NPH	−0.18	1CHA	CPM	−0.1	2CE1	CPH2	0.25
1C1A	CPA	0.12	1HA	HA	0.1	2HE1	HR1	0.13
1C2A	CPB	−0.06	1CHB	CPM	−0.1	3ND1	NR1	−0.4
1C3A	CPB	−0.06	1HB	HA	0.1	3HD1	H	0.42
1C4A	CPA	0.12	1CHC	CPM	−0.1	3CG	CPH1	−0.09
1C1B	CPA	0.12	1HC	HA	0.1	3CB	CT2	−0.11
1C2B	CPB	−0.06	1CHD	CPM	−0.1	3HB1	HA2	0.09
1C3B	CPB	−0.06	1HD	HA	0.1	3HB2	HA2	0.09
1C4B	CPA	0.12	2ND1	NR1	−0.4	3NE2	NR2	−0.7
1C1C	CPA	0.12	2HD1	H	0.42	3CD2	CPH1	0.22
1C2C	CPB	−0.06	2CG	CPH1	−0.09	3HD2	HR3	0.1
1C3C	CPB	−0.06	2CB	CT2	−0.11	3CE1	CPH2	0.25
1C4C	CPA	0.12	2HB1	HA2	0.09	3HE1	HR1	0.13

Table A.3: Chemical types and final CHARMM atomic partial charges for the ferric HAEMHIS patch.

Atom Name	Chem. Type	Final Charge	Atom Name	Chem. Type	Final Charge	Atom Name	Chem. Type	Final Charge
1FE	FE3H	0.4	1C1D	CPA	0.18	2HB2	HA2	0.09
1NA	NPH	−0.17	1C2D	CPB	−0.05	2NE2	NR2	−0.7
1NB	NPH	−0.17	1C3D	CPB	−0.05	2CD2	CPH1	0.22
1NC	NPH	−0.17	1C4D	CPA	0.18	2HD2	HR3	0.1
1ND	NPH	−0.17	1CHA	CPM	−0.1	2CE1	CPH2	0.25
1C1A	CPA	0.18	1HA	HA	0.1	2HE1	HR1	0.13
1C2A	CPB	−0.05	1CHB	CPM	−0.1	3ND1	NR1	−0.32
1C3A	CPB	−0.05	1HB	HA	0.1	3HD1	H	0.4
1C4A	CPA	0.18	1CHC	CPM	−0.1	3CG	CPH1	−0.05
1C1B	CPA	0.18	1HC	HA	0.1	3CB	CT2	−0.09
1C2B	CPB	−0.05	1CHD	CPM	−0.1	3HB1	HA2	0.09
1C3B	CPB	−0.05	1HD	HA	0.1	3HB2	HA2	0.09
1C4B	CPA	0.18	2ND1	NR1	−0.32	3NE2	NR2	−0.7
1C1C	CPA	0.18	2HD1	H	0.4	3CD2	CPH1	0.22
1C2C	CPB	−0.05	2CG	CPH1	−0.05	3HD2	HR3	0.1
1C3C	CPB	−0.05	2CB	CT2	−0.09	3CE1	CPH2	0.25
1C4C	CPA	0.18	2HB1	HA2	0.09	3HE1	HR1	0.13

Table A.4: Chemical types, final CHARMM atomic partial charges and MK charges for the ferrous FEBIHIS patch.

Atom Name	Chem. Type	Final Charge	MK Charge	Atom Name	Chem. Type	Final Charge	MK Charge
1FE2P	FE2NH1	0.16	0.16	4CG	CPH1	0.43	0.41
2ND1	NR1	−0.40	−0.44	4CB	CT2	−0.24	−0.30
2HD1	H	0.34	0.40	4HB1	HA2	0.09	0.09
2CG	CPH1	0.43	0.41	4HB2	HA2	0.09	0.09
2CB	CT2	−0.24	−0.30	4NE2	NR2	0.10	0.10
2HB1	HA2	0.09	0.09	4CD2	CPH1	−0.48	−0.48
2HB2	HA2	0.09	0.09	4HD2	HR3	0.23	0.23
2NE2	NR2	0.10	0.10	4CE1	CPH2	−0.02	−0.01
2CD2	CPH1	−0.48	−0.48	4HE1	HR1	0.19	0.17
2HD2	HR3	0.23	0.23	5ND1	NR1	−0.40	−0.44
2CE1	CPH2	−0.02	−0.01	5HD1	H	0.34	0.40
2HE1	HR1	0.19	0.17	5CG	CPH1	0.43	0.41
3ND1	NR1	−0.40	−0.44	5CB	CT2	−0.24	−0.30
3HD1	H	0.34	0.40	5HB1	HA2	0.09	0.09
3CG	CPH1	0.43	0.41	5HB2	HA2	0.09	0.09
3CB	CT2	−0.24	−0.30	5NE2	NR2	0.10	0.10
3HB1	HA2	0.09	0.09	5CD2	CPH1	−0.48	−0.48
3HB2	HA2	0.09	0.09	5HD2	HR3	0.23	0.23
3NE2	NR2	0.10	0.10	5CE1	CPH2	−0.02	−0.01
3CD2	CPH1	−0.48	−0.48	5HE1	HR1	0.19	0.17
3HD2	HR3	0.23	0.23	6CT	CC	0.80	0.88
3CE1	CPH2	−0.02	−0.01	6OC1	OC	−0.57	−0.64
3HE1	HR1	0.19	0.17	6OC2	OC	−0.57	−0.64
4ND1	NR1	−0.40	−0.44	6OC3	OH1	−0.54	−0.60
4HD1	H	0.34	0.40	6HO3	H	0.40	0.44

Table A.5: Chemical types, final CHARMM atomic partial charges and MK charges for the ferric FEBIHIS patch.

Atom Name	Chem. Type	Final Charge	MK Charge	Atom Name	Chem. Type	Final Charge	MK Charge
1FE3P	FE3NH1	0.04	0.06	4CG	CPH1	0.31	0.39
2ND1	NR1	−0.36	−0.39	4CB	CT2	−0.09	−0.26
2HD1	H	0.38	0.4	4HB1	HA2	0.09	0.09
2CG	CPH1	0.31	0.39	4HB2	HA2	0.09	0.09
2CB	CT2	−0.09	−0.26	4NE2	NR2	0.14	0.14
2HB1	HA2	0.09	0.09	4CD2	CPH1	−0.45	−0.44
2HB2	HA2	0.09	0.09	4HD2	HR3	0.23	0.23
2NE2	NR2	0.14	0.14	4CE1	CPH2	−0.02	0.03
2CD2	CPH1	−0.45	−0.44	4HE1	HR1	0.22	0.17
2HD2	HR3	0.23	0.23	5ND1	NR1	−0.36	−0.39
2CE1	CPH2	−0.02	0.03	5HD1	H	0.38	0.4
2HE1	HR1	0.22	0.17	5CG	CPH1	0.31	0.39
3ND1	NR1	−0.36	−0.39	5CB	CT2	−0.09	−0.26
3HD1	H	0.38	0.4	5HB1	HA2	0.09	0.09
3CG	CPH1	0.31	0.39	5HB2	HA2	0.09	0.09
3CB	CT2	−0.09	−0.26	5NE2	NR2	0.14	0.14
3HB1	HA2	0.09	0.09	5CD2	CPH1	−0.45	−0.44
3HB2	HA2	0.09	0.09	5HD2	HR3	0.23	0.23
3NE2	NR2	0.14	0.14	5CE1	CPH2	−0.02	0.03
3CD2	CPH1	−0.45	−0.44	5HE1	HR1	0.22	0.17
3HD2	HR3	0.23	0.23	6CT	CC	0.72	0.78
3CE1	CPH2	−0.02	0.03	6OC1	OC	−0.47	−0.49
3HE1	HR1	0.22	0.17	6OC2	OC	−0.47	−0.49
4ND1	NR1	−0.36	−0.39	6OC3	OH1	−0.44	−0.48
4HD1	H	0.38	0.4	6HO3	H	0.46	0.46

Table A.6: Chemical types, final CHARMM atomic partial charges and MK charges for the ferrous FEHIS patch.

Atom Name	Chem. Type	Final Charge	MK Charge
1FE2P	FE2NH2	0.08	0.60
2ND1	NR1	−0.12	−0.28
2HD1	H	0.31	0.39
2CG	CPH1	0.33	0.33
2CB	CT2	−0.11	−0.25
2HB1	HA2	0.09	0.09
2HB2	HA2	0.09	0.09
2NE2	NR2	−0.07	−0.07
2CD2	CPH1	−0.38	−0.35
2HD2	HR3	0.24	0.21
2CE1	CPH2	−0.13	−0.12
2HE1	HR1	0.23	0.22
3ND1	NR1	−0.12	−0.28
3HD1	H	0.31	0.39
3CG	CPH1	0.33	0.33
3CB	CT2	−0.11	−0.25
3HB1	HA2	0.09	0.09
3HB2	HA2	0.09	0.09
3NE2	NR2	−0.07	−0.07
3CD2	CPH1	−0.38	−0.35
3HD2	HR3	0.24	0.21
3CE1	CPH2	−0.13	−0.12
3HE1	HR1	0.23	0.22
4ND1	NR1	−0.12	−0.28
4HD1	H	0.31	0.39
4CG	CPH1	0.33	0.33
4CB	CT2	−0.11	−0.25
4HB1	HA2	0.09	0.09
4HB2	HA2	0.09	0.09
4NE2	NR2	−0.07	−0.07
4CD2	CPH1	−0.38	−0.35
4HD2	HR3	0.24	0.21
4CE1	CPH2	−0.13	−0.12
4HE1	HR1	0.23	0.22
5ND1	NR1	−0.12	−0.28
5HD1	H	0.31	0.39
5CG	CPH1	0.33	0.33
5CB	CT2	−0.11	−0.25
5HB1	HA2	0.09	0.09
5HB2	HA2	0.09	0.09
5NE2	NR2	−0.07	−0.07
5CD2	CPH1	−0.38	−0.35
5HD2	HR3	0.24	0.21
5CE1	CPH2	−0.13	−0.12
5HE1	HR1	0.23	0.22

Table A.7: Chemical types, final CHARMM atomic partial charges and MK charges for the ferric FEHIS patch.

Atom Name	Chem. Type	Final Charge	MK Charge	Atom Name	Chem. Type	Final Charge	MK Charge
1FE3P	FE3NH2	0.12	0.36	4ND1	NR1	−0.26	−0.26
2ND1	NR1	−0.26	−0.26	4HD1	H	0.41	0.40
2HD1	H	0.41	0.40	4CG	CPH1	0.35	0.35
2CG	CPH1	0.35	0.35	4CB	CT2	−0.09	−0.19
2CB	CT2	−0.09	−0.19	4HB1	HA2	0.09	0.09
2HB1	HA2	0.09	0.09	4HB2	HA2	0.09	0.09
2HB2	HA2	0.09	0.09	4NE2	NR2	0.03	0.02
2NE2	NR2	0.03	0.02	4CD2	CPH1	−0.40	−0.33
2CD2	CPH1	−0.40	−0.33	4HD2	HR3	0.32	0.22
2HD2	HR3	0.32	0.22	4CE1	CPH2	−0.09	−0.03
2CE1	CPH2	−0.09	−0.03	4HE1	HR1	0.27	0.21
2HE1	HR1	0.27	0.21	5ND1	NR1	−0.26	−0.26
3ND1	NR1	−0.26	−0.26	5HD1	H	0.41	0.40
3HD1	H	0.41	0.40	5CG	CPH1	0.35	0.35
3CG	CPH1	0.35	0.35	5CB	CT2	−0.09	−0.19
3CB	CT2	−0.09	−0.19	5HB1	HA2	0.09	0.09
3HB1	HA2	0.09	0.09	5HB2	HA2	0.09	0.09
3HB2	HA2	0.09	0.09	5NE2	NR2	0.03	0.02
3NE2	NR2	0.03	0.02	5CD2	CPH1	−0.40	−0.33
3CD2	CPH1	−0.40	−0.33	5HD2	HR3	0.32	0.22
3HD2	HR3	0.32	0.22	5CE1	CPH2	−0.09	−0.03
3CE1	CPH2	−0.09	−0.03	5HE1	HR1	0.27	0.21
3HE1	HR1	0.27	0.21				

A.2 | New CHARMM Bond Parameters

The parameters presented here use the iron chemical types found in Table A.1 and the standard CHARMM chemical types of the CHARMM36 protein force field[97].

Table A.8: The new bond parameters for the ferrous and ferric HAEMHIS complex.

Parameter	$K_b{}^*$ (kcal/mol/Å^2)	$b_0{}^*$ (Å)	Source
NPH–FE2H0	270.20	2.040	NPH–FE[118]
NR2–FE2H0	80.00	2.110	NR2–FE[118]
NPH–FE2H	270.20	2.040	NPH–FE[118]
NR2–FE2H	80.00	2.110	NR2–FE[118]
NPH–FE3H	270.00	2.030	NPH–FE[118]
NR2–FE3H	80.00	2.080	NR2–FE[118]

*see Eq. (2.15) for parameter definitions

Table A.9: The new bond parameters for the ferrous and ferric FEBIHIS complex.

Parameter	$K_b{}^*$ (kcal/mol/Å^2)	$b_0{}^*$ (Å)	Source
NR2–FE2NH1	140.00	1.960	NR2–FE[118]
OC–FE2NH1	100.00	1.995	OM–FE[180]
NR2–FE3NH1	140.00	1.960	NR2–FE[118]
OC–FE3NH1	100.00	1.890	OM–FE[180]
OH1–CC	230.00	1.380	OH1–CD[99]

*see Eq. (2.15) for parameter definitions

Table A.10: The new bond parameters for the ferrous and ferric FEHIS complex.

Parameter	$K_b{}^*$ (kcal/mol/Å^2)	$b_0{}^*$ (Å)	Source
NR2–FE2NH2	140.00	1.940	NR2–FE[118]
NR2–FE3NH2	140.00	1.870	NR2–FE[118]

*see Eq. (2.15) for parameter definitions

A.3 | New CHARMM Angle Parameters

The parameters presented here use the iron chemical types found in Table A.1 and the standard CHARMM chemical types of the CHARMM36 protein force field[97].

Table A.11: The new angle parameters for the ferrous and ferric HAEMHIS complex.

Parameter	K_θ* (kcal/mol/rad^2)	θ_0* (°)	Source
FE2H0–NR2–CPH1	25.00	135.00	FE–NR2–CPH1[118]
FE2H0–NR2–CPH2	20.00	138.00	FE–NR2–CPH2[118]
NPH–FE2H0–NPH	14.39	90.00	NPH–FE–NPH[118]
FE2H0–NPH–CPA	96.15	128.05	FE–NPH–CPA[118]
NR2–FE2H0–NPH	65.00	90.00	NR2–FE–NPH[118]
FE2H–NR2–CPH1	25.00	135.00	FE–NR2–CPH1[118]
FE2H–NR2–CPH2	25.00	137.00	FE–NR2–CPH2[118]
NPH–FE2H–NPH	14.39	90.00	NPH–FE–NPH[118]
FE2H–NPH–CPA	96.15	128.05	FE–NPH–CPA[118]
NR2–FE2H–NPH	65.00	90.00	NR2–FE–NPH[118]
FE3H–NR2–CPH1	25.00	135.00	FE–NR2–CPH1[118]
FE3H–NR2–CPH2	25.00	137.00	FE–NR2–CPH2[118]
NPH–FE3H–NPH	14.39	90.00	NPH–FE–NPH[118]
FE3H–NPH–CPA	96.15	128.05	FE–NPH–CPA[118]
NR2–FE3H–NPH	65.00	90.00	NR2–FE–NPH[118]

*see Eq. (2.15) for parameter definitions

Table A.12: The new angle parameters for the ferrous and ferric FEBIHIS complex.

Parameter	$K_\theta{}^*$ (kcal/mol/rad^2)	$\theta_0{}^*$ (°)	Source
FE2NH1–NR2–CPH1	25.00	123.70	FE–NR2–CPH1[118]
FE2NH1–NR2–CPH2	25.00	127.60	FE–NR2–CPH2[118]
OC–FE2NH1–OC	0.00	101.50	OC–CC–OC[99]
FE2NH1–OC–CC	40.00	85.00	CG2O6–OG302–CG321[27]
FE3NH1–NR2–CPH1	25.00	122.70	FE–NR2–CPH1[118]
FE3NH1–NR2–CPH2	25.00	125.00	FE–NR2–CPH2[118]
OC–FE3NH1–OC	0.00	101.50	OC–CC–OC[99]
FE3NH1–OC–CC	40.00	98.00	CG2O6–OG302–CG321[27]
H–OH1–CC	55.00	113.30	H–OH1–CD[99]
OH1–CC–OC†	50.00	123.00	OH1–CD–OB[99]

*see Eq. (2.15) for parameter definitions

†includes the UB parameters $K_{\mathrm{UB}} = 210.0\,\mathrm{kcal/mol/\AA^2}$ and $S_0 = 2.262\,\mathrm{\AA}$

Table A.13: The new angle parameters for the ferrous and ferric FEHIS complex.

Parameter	$K_\theta{}^*$ (kcal/mol/rad^2)	$\theta_0{}^*$ (°)	Source
FE2NH2–NR2–CPH1	25.00	124.00	FE–NR2–CPH1[118]
FE2NH2–NR2–CPH2	25.00	125.80	FE–NR2–CPH2[118]
FE3NH2–NR2–CPH1	25.00	127.00	FE–NR2–CPH1[118]
FE3NH2–NR2–CPH2	25.00	124.00	FE–NR2–CPH2[118]

*see Eq. (2.15) for parameter definitions

A.4 | New CHARMM Dihedral Parameters

The parameters presented here use the iron chemical types found in Table A.1 and the standard CHARMM chemical types of the CHARMM36 protein force field[97].

Table A.14: The new dihedral parameters for the ferrous and ferric HAEMHIS complex.

Parameter	K_φ* (kcal/mol)	n*	δ_0* (°)	Source
NPH–FE2H0–NR2–CPH1	0.19	4	0	X–FE–NR2–X[118]
NPH–FE2H–NR2–CPH1	0.07	2	180	X–FE–NR2–X[118]
NPH–FE2H–NR2–CPH1	0.04	3	180	X–FE–NR2–X[118]
NPH–FE2H–NR2–CPH1	0.14	4	0	X–FE–NR2–X[118]
NPH–FE3H–NR2–CPH1	0.10	2	180	X–FE–NR2–X[118]
NPH–FE3H–NR2–CPH1	0.07	3	180	X–FE–NR2–X[118]
NPH–FE3H–NR2–CPH1	0.55	4	0	X–FE–NR2–X[118]
X–FE2H0–NPH–X	0.00	2	0	X–FE–NPH–X[118]
X–FE2H–NPH–X	0.00	2	0	X–FE–NPH–X[118]
X–FE3H–NPH–X	0.00	2	0	X–FE–NPH–X[118]

*see Eq. (2.15) for parameter definitions

Table A.15: The new dihedral parameters for the ferrous and ferric FEBIHIS complex.

Parameter	K_φ* (kcal/mol)	n*	δ_0* (°)	Source
OC–CC–OH1–H	2.20	2	180	H–OH1–CA–CA[99]
FE2NH1–OC–CC–OH1	3.00	2	180	X–CD–OH1–X[99]
FE2NH1–OC–CC–OC	0.50	2	180	X–FE–NPH–X[118]
FE3NH1–OC–CC–OH1	3.00	2	180	X–CD–OH1–X[99]
FE3NH1–OC–CC–OC	0.00	2	180	X–FE–NPH–X[118]
CC–OC–FE2NH1–OC	0.00	2	0	X–FE–NPH–X[118]
CC–OC–FE3NH1–OC	0.00	2	0	X–FE–NPH–X[118]

*see Eq. (2.15) for parameter definitions

A.5 | New CHARMM Improper Parameters

The parameters presented here use the iron chemical types found in Table A.1 and the standard CHARMM chemical types of the CHARMM36 protein force field[97].

Because of the changed chemical type of the central hæm iron, HAEMHIS was the only complex that required the addition of new improper parameters. Consequently, the hæm improper parameter involving iron[118,119] was used without modification.

Table A.16: The new improper dihedral angle parameters for the ferrous and ferric HAEMHIS complex.

Parameter	${K_\omega}^*$ (kcal/mol/rad^2)	${\omega_0}^*$ (°)	Source
NPH–CPA–CPA–FE2H0	137.40	0.00	NPH–CPA–CPA–FE[118]
NPH–CPA–CPA–FE2H	137.40	0.00	NPH–CPA–CPA–FE[118]
NPH–CPA–CPA–FE3H	137.40	0.00	NPH–CPA–CPA–FE[118]

*see Eq. (2.15) for parameter definitions

A.6 | Potential Energy Scans of Bonds/Angles

This section contains the results of PES scans performed during the optimization of the force constants of the bonds and angles of HAEMHIS, FEBIHIS and FEHIS.

Table A.17: Energy differences ΔE, in 10^{-2} kcal/mol, for the PES performed during the refinement of the force constants for bond and valence angles of ferrous and ferric HAEMHIS using QM and the optimized MM parameters.

		ΔE (Fe^{2+})			ΔE (Fe^{3+})	
Parameter	Offset	QM	$MM_{Set\ 1}$	$MM_{Set\ 2}$	QM	MM
	(Å)					
Fe–N_ε	−0.01	0.93	0.88	0.88	1.16	0.89
	0.00	0.00	0.00	0.00	0.00	0.00
	+0.01	0.87	0.87	0.87	1.10	0.89
	(°)					
Fe–N_ε–C_δ	−1	1.92	1.91	1.79	1.13	2.07
	0	0.00	0.00	0.00	0.00	0.00
	+1	1.90	1.90	1.78	1.32	2.05
Fe–N_ε–C_ε	−1	1.88	1.88	1.73	1.32	2.02
	0	0.00	0.00	0.00	0.00	0.00
	+1	1.88	1.88	1.74	1.12	2.03

Atoms have been labelled according to Fig. 2.3.

Table A.18: Energy differences ΔE, in 10^{-2} kcal/mol, for the PES performed during the refinement of the force constants for bond and valence angles of ferrous and ferric FEBIHIS using QM and the optimized MM parameters Set 1 and Set 2.

		ΔE (Fe^{2+})		ΔE (Fe^{3+})	
Parameter	Offset	QM	MM	QM	MM
	(Å)				
Fe–N_ε	−0.01	1.11	1.44	0.97	1.40
	0.00	0.00	0.00	0.00	0.00
	+0.01	1.09	1.43	1.29	1.40
Fe–$O_{1/2}$	−0.01	0.76	1.15	1.00	1.15
	0.00	0.00	0.00	0.00	0.00
	+0.01	0.70	1.15	0.83	1.15
	(°)				
Fe–N_ε–C_δ	−1	1.61	0.96	1.68	1.18
	0	0.00	0.00	0.00	0.00
	+1	1.58	1.26	1.70	1.21
Fe–N_ε–C_ε	−1	1.44	0.82	1.54	1.03
	0	0.00	0.00	0.00	0.00
	+1	1.49	1.18	1.66	1.06
Fe–$O_{1/2}$–C_3	−1	5.02	5.86	5.50	5.73
	0	0.00	0.00	0.00	0.00
	+1	5.01	5.85	6.16	5.66

Atoms have been labelled according to Fig. 2.4.

Table A.19: Energy differences ΔE, in 10^{-2} kcal/mol, for the PES performed during the refinement of the force constants for bond and valence angles of ferrous and ferric FEHIS using QM and the optimized MM parameters.

		ΔE (Fe^{2+})		ΔE (Fe^{3+})	
Parameter	Offset	QM	MM	QM	MM
	(Å)				
Fe–N_ε	−0.01	1.47	1.36	1.49	1.38
	0.00	0.00	0.00	0.00	0.00
	+0.01	1.43	1.36	1.96	1.38
	(°)				
Fe–N_ε–C_δ	−1	1.70	0.77	1.59	1.61
	0	0.00	0.00	0.00	0.00
	+1	1.72	0.77	2.00	1.62
Fe–N_ε–C_ε	−1	1.65	0.84	1.95	1.61
	0	0.00	0.00	0.00	0.00
	+1	1.64	0.86	1.59	1.61

Atoms have been labelled according to Fig. 2.5.

A.7 | Repeat Simulations of Neuroglobin

Table A.20: Geometry of the HAEMHIS complex during repeat MD simulations of neuroglobin compared with the crystal structure.

Parameter	Residue	Crystal*	SIM12	SIM22
		Bond distance d (Å)		
Fe–N_ε	H64	2.11 ± 0.05	2.05 ± 0.06	2.05 ± 0.06
	H96	2.05 ± 0.04	2.05 ± 0.05	2.05 ± 0.06
		Angles (°)		
Fe–N_ε–C_ε (α)	H64	126 ± 1	126 ± 4	127 ± 4
	H96	132 ± 2	126 ± 3	127 ± 3
Fe–N_ε–C_δ	H64	125 ± 1	128 ± 4	128 ± 3
	H96	119 ± 2	128 ± 3	128 ± 3
		Dihedral ω (°)		
N_A–Fe–N_ε–C_ε	H64	-124 ± 4	-121 ± 11	-121 ± 12
	H96	-116 ± 4	-122 ± 12	-119 ± 14

*Crystal refers to values measured in the crystal structure (PDB ID: 1OJ6[131]). For H64 and H96, we report the geometry parameters illustrated in Fig. 3.15.

Analysis of Channelrhodopsin B

B.1 | Root-Mean-Square Deviation Profiles

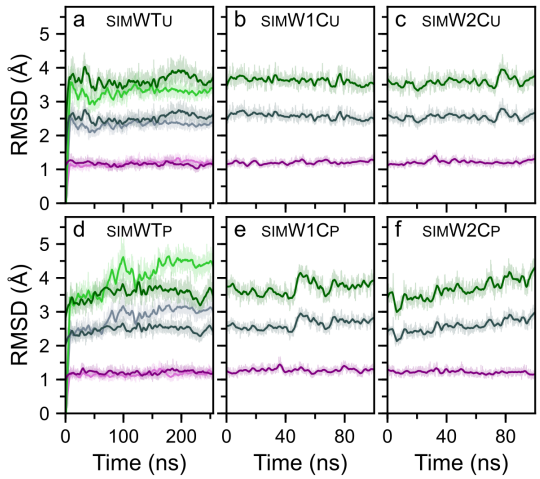


Figure B.1: C_α RMSD profiles for simulations of wild type C1C2. RMSD values are shown for: the whole protein (slate grey) and the α-helical (purple) and loop regions (green). Results from repeat simulations are shown in brighter colours.

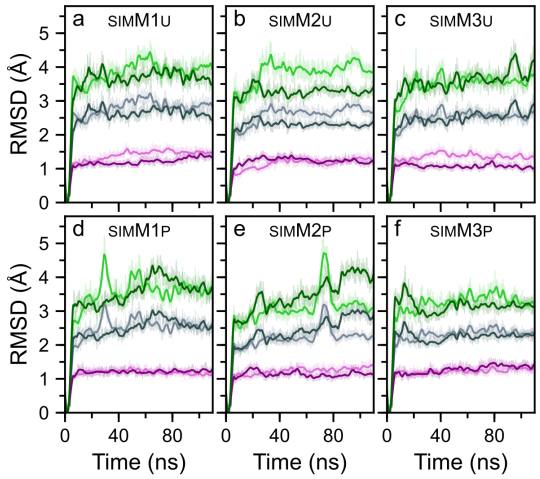


Figure B.2: C_α RMSD profiles for mutant simulations of C1C2. RMSD values are shown for: the whole protein (slate grey) and the α-helical (purple) and loop regions (green). Results from repeat simulations are shown in brighter colours.

B.2 | Hydration of Channelrhodopsin

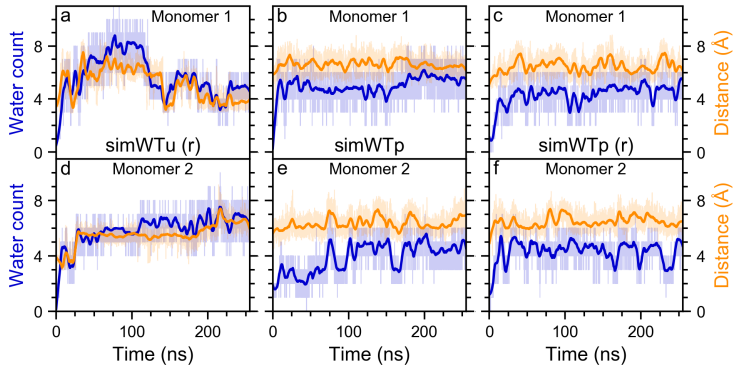


Figure B.3: Water in the Schiff base region in wild type simulations of C1C2 with all-*trans* retinal. We compared the number of water molecules hydrogen-bonding to the Schiff base or E162/D292 (blue) with the distance between the ammonium group nitrogen of K132 and C_γ of D292 (orange).

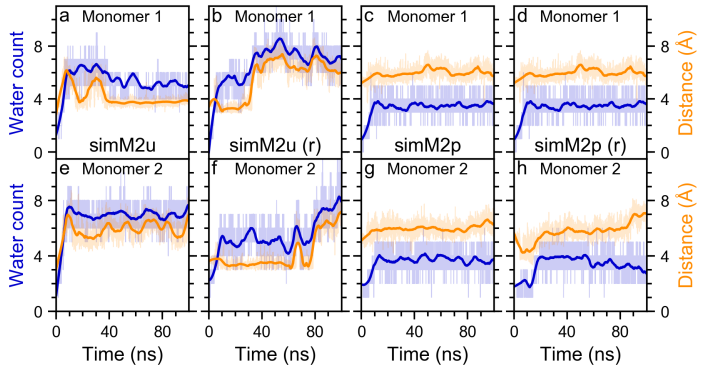


Figure B.4: Water in the Schiff base region in E129A mutant simulations of C1C2. We compared the number of water molecules hydrogen-bonding to the Schiff base or E162/D292 (blue) with the distance between the ammonium group nitrogen of K132 and C_γ of D292 (orange).

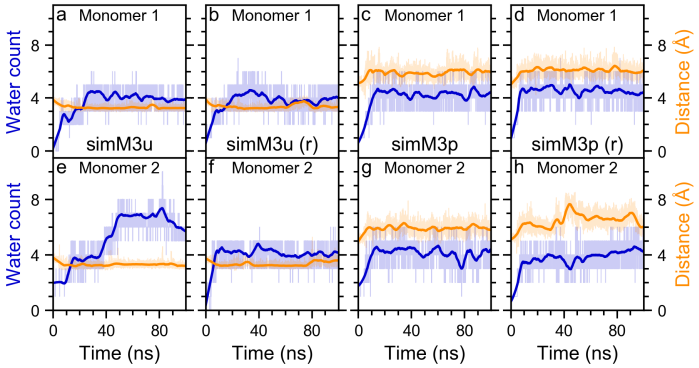


Figure B.5: Water in the Schiff base region in E162D mutant simulations of C1C2. We compared the number of water molecules hydrogen-bonding to the Schiff base or E162/D292 (blue) with the distance between the ammonium group nitrogen of K132 and C_γ of D292 (orange).

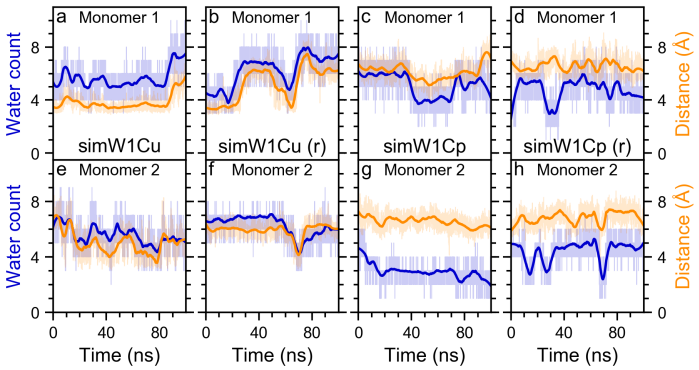


Figure B.6: Water in the Schiff base region in the wild type simulations simW1Cu and simW1Cp with 13-*cis* retinal in monomer 1. We compared the number of water molecules hydrogen-bonding to the Schiff base or E162/D292 (blue) with the distance between the ammonium group nitrogen of K132 and C_γ of D292 (orange).

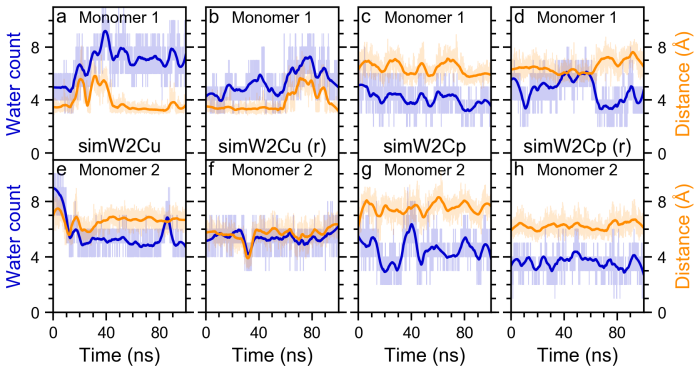


Figure B.7: Water in the Schiff base region in the wild type simulations simW2Cu and simW2Cp with 13-*cis* retinal in both monomers. We compared the number of water molecules hydrogen-bonding to the Schiff base or E162/D292 (blue) with the distance between the ammonium group nitrogen of K132 and C_γ of D292 (orange).

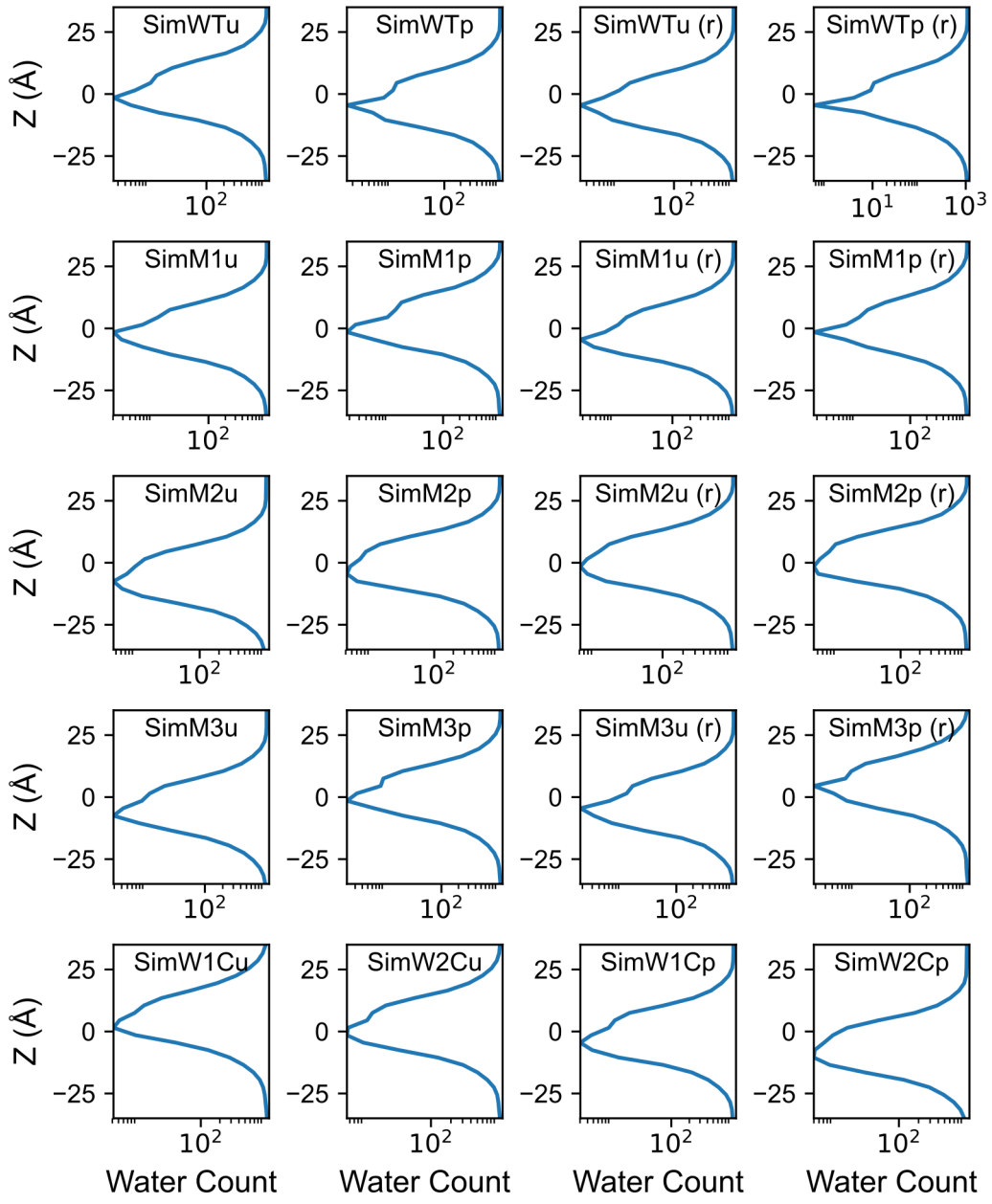


Figure B.8: Water molecules inside of the intrahelical region of C1C2 for all wild type and mutant simulations. The distribution of water molecules along the z-axis is shown for all simulations. Water numbers were calculated for the simulation box for the last 50 ns of the respective simulation. Because of the dynamics of the lipid bilayer, the hydration minimum was not always located at $Z = 0$ Å.

Table B.1: Number of water molecules forming hydrogen bonds with the Schiff base nitrogen or the side chains of E162 or D292 (data for repeat simulations).

Simulation	Number of Water Molecules Monomer 1	Monomer 2
simWTu	4.4 ± 0.9	6.6 ± 0.9
simWTp	4.7 ± 0.8	4.2 ± 0.9
simW1Cu	6.8 ± 1.2	5.7 ± 0.9
simW1Cp	4.9 ± 0.9	4.6 ± 1.0
simW2Cu	5.8 ± 1.2	5.4 ± 0.7
simW2Cp	4.4 ± 1.2	3.6 ± 0.7
simM1u	7.5 ± 1.4	6.1 ± 0.6
simM1p	5.2 ± 0.9	5.4 ± 1.0
simM2u	7.1 ± 0.8	6.7 ± 1.3
simM2p	3.9 ± 0.6	3.3 ± 0.8
simM3u	3.7 ± 0.6	4.4 ± 0.9
simM3p	4.4 ± 0.7	4.3 ± 0.6

B.3 | Hydrogen-Bonding in Channelrhodopsin

Table B.2: Hydrogen-bonding partners of the carboxyl group of E129 during the last 50 ns of first run MM MD.

Simulation	Monomer	Hydrogen-Bonding Percentage (%)						
		Water	N297	E162	D292	T98	S102	K132
simWTu	1	95	100	—	—	—	—	—
	2	89	100	—	—	—	—	—
simWTp	1	100	99	—	—	—	—	—
	2	61	100	—	—	7	—	—
simW1Cu	1	97	100	—	—	—	—	—
	2	92	100	—	—	—	—	—
simW1Cp	1	62	62	17	29	—	—	—
	2	69	0	44	52	—	—	6
simW2Cu	1	97	100	—	—	—	—	—
	2	83	100	—	—	—	—	—
simW2Cp	1	88	100	—	—	—	—	—
	2	39	96	—	—	10	—	—
simM1u	1	74	100	—	—	6	—	—
	2	79	100	—	—	6	—	—
simM1p	1	31	100	—	—	4	—	—
	2	12	100	—	—	4	—	—
simM3u	1	46	99	—	11	6	—	—
	2	98	92	—	—	—	9	—
simM3p	1	82	100	—	—	—	—	—
	2	51	90	—	—	6	—	—

To increase clarity, only percentages >3 % are shown.

Table B.3: Hydrogen-bonding partners of the carboxyl group of E129 during the last 50 ns of repeat MM MD.

Simulation	Monomer	Hydrogen-Bonding Percentage (%)						
		Water	N297	E162	D292	T98	S102	K132
simWTu	1	40	—	100	—	6	—	46
	2	70	—	92	—	—	—	84
simWTp	1	18	100	—	—	6	—	—
	2	21	100	—	—	7	—	—
simW1Cu	1	76	—	100	—	—	—	46
	2	46	—	97	—	—	—	91
simW1Cp	1	73	—	99	—	—	—	71
	2	56	—	64	—	—	—	94
simW2Cu	1	23	100	—	—	8	—	—
	2	76	100	—	—	4	—	—
simW2Cp	1	19	100	—	—	5	—	—
	2	6	100	—	—	8	—	—
simM1u	1	97	98	—	—	—	—	—
	2	5	100	—	—	—	—	—
simM1p	1	69	81	—	—	8	—	—
	2	30	100	—	—	4	—	—
simM3u	1	68	100	—	5	—	—	—
	2	14	100	—	4	6	—	—
simM3p	1	27	100	—	—	—	16	—
	2	100	—	100	—	4	—	9

To increase clarity, only percentages >3 % are shown.

Table B.4: Hydrogen-bonding partners of the carboxyl group of E129 during the last 750 ps of QM/MM MD.

Simulation	Monomer	Hydrogen-Bonding Percentage (%)						
		Water	N297	E162	D292	T98	S102	K132
simWTu	1	100	100	—	—	—	—	—
	2	98	100	—	—	—	—	—
simWTp	1	100	100	—	—	—	—	—
	2	99	100	—	—	5	—	—
simW1Cu	1	100	100	—	—	—	—	—
	2	99	100	—	—	4	—	—
simW1Cp	1	100	99	—	—	—	—	—
	2	99	100	—	—	—	—	—
simM1u	1	95	99	—	—	—	—	—
	2	95	100	—	—	—	—	—
simM1p	1	—	100	—	—	4	—	—
	2	23	100	—	—	—	—	—
simM3u	1	59	100	—	—	—	—	—
	2	100	43	—	—	—	—	—
simM3p	1	91	100	—	—	—	—	—
	2	83	100	—	—	8	—	—

To increase clarity, only percentages >3 % are shown.

Table B.5: Hydrogen-Bonding (HB), in %, and oxygen–oxygen (d_{OO}) and oxygen–nitrogen (d_{ON}) distances, in Å, between N297 and S102 during the last 50 ns of MM MD.

Simulation	Run*	Monomer 1			Monomer 2		
		HB	d_{OO}	d_{ON}	HB	d_{OO}	d_{ON}
simWTu	1	65	4.4 ± 0.7	3.5 ± 0.8	82	4.1 ± 0.3	3.2 ± 0.3
	2	86	3.8 ± 0.4	3.1 ± 0.2	5	5.2 ± 0.8	4.8 ± 0.7
simWTp	1	84	3.9 ± 0.3	3.2 ± 0.3	60	4.3 ± 0.4	3.3 ± 0.4
	2	19	4.2 ± 0.4	3.8 ± 0.6	29	4.5 ± 0.6	3.7 ± 0.5
simW1Cu	1	51	4.5 ± 0.5	3.6 ± 0.6	87	4.1 ± 0.3	3.1 ± 0.2
	2						
simW1Cp	1	75	4.1 ± 0.4	3.2 ± 0.4	63	3.6 ± 0.6	3.3 ± 0.4
	2						
simW2Cu	1	77	4.3 ± 0.4	3.2 ± 0.3	70	4.2 ± 0.4	3.3 ± 0.4
	2						
simW2Cp	1	69	4.2 ± 0.3	3.3 ± 0.3	12	4.1 ± 0.5	3.9 ± 0.6
	2						
simM1u	1	70	4.1 ± 0.4	3.3 ± 0.4	72	4.0 ± 0.4	3.2 ± 0.3
	2	89	4.0 ± 0.3	3.1 ± 0.3	30	4.0 ± 0.4	3.6 ± 0.5
simM1p	1	64	4.0 ± 0.3	3.3 ± 0.3	21	4.0 ± 0.4	3.7 ± 0.5
	2	47	4.6 ± 0.6	3.6 ± 0.7	41	4.2 ± 0.4	3.5 ± 0.4
simM2u	1	81	4.0 ± 0.3	3.2 ± 0.3	0	5.9 ± 0.6	5.0 ± 0.5
	2	38	4.7 ± 0.5	3.5 ± 0.4	35	4.5 ± 0.8	3.7 ± 0.5
simM2p	1	30	4.2 ± 0.5	3.7 ± 0.5	62	4.3 ± 0.5	3.4 ± 0.5
	2	21	4.9 ± 0.7	4.2 ± 0.8	42	4.4 ± 0.5	3.5 ± 0.5
simM3u	1	67	3.9 ± 0.3	3.3 ± 0.4	68	4.1 ± 0.4	3.3 ± 0.4
	2	0	6.7 ± 0.8	5.8 ± 0.6	50	3.9 ± 0.4	3.4 ± 0.4
simM3p	1	38	4.4 ± 0.5	3.7 ± 0.6	27	4.5 ± 0.5	3.7 ± 0.6
	2	1	6.7 ± 0.8	5.8 ± 0.7	37	4.1 ± 0.6	3.6 ± 0.5

*Initial simulations are denoted as run 1 and repeats as run 2.

Table B.6: Percentage of E162/D292 hydrogen-bonding to water (E162–w/D292–w) and of E162 interacting with D292 either via direct hydrogen-bonding or via a one-water bridge (E–D) as sampled during the last 50 ns of MM MD.

Simulation	Run*	Monomer 1 (%)			Monomer 2 (%)		
		E162–w	D292–w	E–D	E162–w	D292–w	E–D
simWTu	1	100	100	100	100	100	100
	2	100	100	97	100	100	99
simWTp	1	99	100	98	100	100	88
	2	96	100	91	93	100	97
simW1Cu	1	100	100	100	100	100	99
	2	100	100	98	100	100	98
simW1Cp	1	90	100	93	64	100	100
	2	93	100	93	96	100	79
simW2Cu	1	100	100	96	100	100	99
	2	100	100	96	100	100	98
simW2Cp	1	94	100	85	99	100	92
	2	90	100	98	89	100	99
simM1u	1	100	100	99	100	100	0
	2	100	100	98	100	100	98
simM1p	1	100	100	78	100	100	42
	2	99	100	58	99	100	47
simM2u	1	100	100	31	100	100	99
	2	100	100	100	100	100	98
simM2p	1	85	100	100	72	100	100
	2	76	100	100	74	100	99
simM3u	1	73	100	10	100	100	31
	2	95	100	78	95	100	70
simM3p	1	99	100	73	97	100	79
	2	99	100	70	39	100	34

*Initial simulations are denoted as run 1 and repeats as run 2.

Table B.7: Percentage of E162/D292 hydrogen-bonding to water (E162–w/D292–w) and of E162 interacting with D292 either via direct hydrogen-bonding or via a one-water bridge (E–D) as sampled during the last 750 ps of QM/MM MD.

Simulation	Monomer 1 (%)			Monomer 2 (%)		
	E162–w	D292–w	E–D	E162–w	D292–w	E–D
simWTu	100	100	100	100	100	97
simWTp	100	100	85	100	100	97
simW1Cu	100	100	94	100	100	98
simW1Cp	100	100	51	100	100	87
simM1u	100	100	88	100	100	0
simM1p	100	100	99	100	100	63
simM2u	100	100	80	100	100	99
simM2p	94	100	100	99	100	100
simM3u	94	100	87	100	100	100
simM3p	100	100	96	99	100	86

Table B.8: Distances, in Å, between carboxyl(ate) oxygens of E162 and D292 in wild type and mutant simulations of C1C2.

Simulation	First Run		Repeats		Average
	Mon 1	Mon 2	Mon 1	Mon 2	
simWTu	5.1 ± 0.2	5.4 ± 0.3	4.8 ± 0.2	5.5 ± 0.3	5.2
simWTp	4.9 ± 0.5	5.2 ± 0.3	5.2 ± 0.5	4.9 ± 0.6	5.1
simW1Cu	5.4 ± 0.2	4.9 ± 0.3	5.4 ± 0.4	5.2 ± 0.4	5.2
simW1Cp	4.8 ± 0.7	4.1 ± 0.2	4.6 ± 0.6	5.3 ± 0.5	4.7
simW2Cu	5.5 ± 0.2	5.8 ± 0.3	5.3 ± 0.4	5.4 ± 0.3	5.5
simW2Cp	4.8 ± 0.6	5.0 ± 0.5	4.7 ± 0.5	4.6 ± 0.5	4.8
simM1u	5.7 ± 0.3	5.7 ± 0.2	5.5 ± 0.4	5.3 ± 0.2	5.5
simM1p	5.6 ± 0.7	5.5 ± 0.5	5.6 ± 0.7	5.9 ± 0.6	5.6
simM2u	4.8 ± 0.2	5.6 ± 0.3	5.5 ± 0.3	5.4 ± 0.3	5.3
simM2p	4.0 ± 0.2	4.2 ± 0.2	4.2 ± 0.2	4.2 ± 0.3	4.2
simM3u	5.7 ± 0.2	5.8 ± 0.2	5.4 ± 0.2	5.5 ± 0.2	5.6
simM3p	5.7 ± 0.4	5.7 ± 0.5	5.7 ± 0.4	6.1 ± 0.4	5.8

Table B.9: Percentage of K132 hydrogen-bonding to E162 (K132–E) or D292 (K132–D) and of K132 simultaneously forming hydrogen bonds with E162 and D292 (K–ED) as sampled during the last 50 ns of MM MD.

Simulation	Run*	Monomer 1 (%)			Monomer 2 (%)		
		K132–E	K132–D	K–ED	K132–E	K132–D	K–ED
simWTu	1	33	32	32	80	—	—
	2	100	73	72	91	—	—
simWTp	1	5	—	—	—	—	—
	2	—	—	—	—	—	—
simW1Cu	1	99	76	76	92	27	27
	2	80	14	14	99	—	—
simW1Cp	1	5	—	—	—	—	—
	2	5	—	—	—	—	—
simW2Cu	1	100	98	98	100	—	—
	2	99	50	50	99	—	—
simW2Cp	1	—	—	—	—	—	—
	2	—	—	—	—	—	—
simM2u	1	99	100	99	87	—	—
	2	69	—	—	37	22	22
simM2p	1	—	—	—	—	—	—
	2	—	—	—	—	—	—
simM3u	1	97	100	98	100	100	100
	2	100	100	100	100	98	98
simM3p	1	—	—	—	—	—	—
	2	—	—	—	—	—	—

*Initial simulations are denoted as run 1 and repeats as run 2.

To increase clarity, only percentages >3 % are shown.

Table B.10: Percentage of K132 hydrogen-bonding to E162 (K132–E) or D292 (K132–D) and of K132 simultaneously forming hydrogen bonds with E162 and D292 (K–ED) as sampled during the last 750 ps of QM/MM MD.

Simulation	Monomer 1 (%)			Monomer 2 (%)		
	K132–E	K132–D	K–ED	K132–E	K132–D	K–ED
simWTu	100	91	91	100	—	—
simWTp	35	88	67	89	—	—
simW1Cu	100	99	99	99	—	—
simW1Cp	70	74	62	31	—	—
simM2u	100	99	99	97	—	—
simM2p	21	—	—	98	—	—
simM3u	93	100	97	96	96	95
simM3p	52	56	46	6	—	—

To increase clarity, only percentages >3 % are shown.

Table B.11: Hydrogen-bonding partners of the Schiff base nitrogen during the last 50 ns of repeat MM simulations.

Simulation	Monomer	Hydrogen-Bonding Percentage (%)				
		Water	E162	D292	S295	T166
simWTu	1	56	38	9	—	—
	2	85	—	36	—	—
simWTp	1	76	—	40	—	—
	2	60	—	54	—	—
simW1Cu	1	96	—	—	—	—
	2	57	—	61	—	—
simW1Cp	1	8	—	—	58	—
	2	72	—	52	—	—
simW2Cu	1	94	—	—	—	—
	2	18	20	75	—	—
simW2Cp	1	—	—	—	67	—
	2	6	—	98	—	—
simM1u	1	64	28	22	—	—
	2	95	—	10	—	—
simM1p	1	54	—	66	—	—
	2	61	—	60	—	—
simM2u	1	94	—	21	—	—
	2	23	54	30	—	—
simM2p	1	—	50	78	—	—
	2	—	23	93	—	—
simM3u	1	—	100	—	—	—
	2	15	85	—	—	—
simM3p	1	35	—	71	—	—
	2	20	—	82	—	—

To increase clarity, only percentages >3 % are shown.

Table B.12: Hydrogen-bonding partners of the QM Schiff base nitrogen during the last 750 ps of QM/MM MD.

Simulation	Monomer	Hydrogen-Bonding Percentage (%)				
		Water	E162	D292	S295	T166
simWTu	1	90	41	—	—	—
	2	84	—	83	—	—
simWTp	1	93	—	15	—	—
	2	97	—	4	—	—
simW1Cu*	1	—	—	—	24	—
simW1Cp*	1	—	—	—	59	—
simW2Cu	1	43	—	—	—	—
	2	—	—	—	—	95
simW2Cp	1	19	22	82	—	—
	2	—	—	—	65	—
simM1u	1	93	—	14	—	—
	2	—	—	95	—	—
simM1p	1	5	30	91	—	—
	2	97	—	12	—	—
simM2u	1	—	99	—	—	—
	2	95	—	21	—	—
simM2p	1	—	31	90	—	—
	2	—	25	90	—	—
simM3u	1	5	89	27	—	—
	2	97	—	6	—	—
simM3p	1	47	—	49	—	—
	2	5	12	67	—	—

*Results for the monomer containing 13-*cis* retinal only.

To increase clarity, only results for QM regions with percentages >3 % are shown.

B.4 | Water Density inside of Channelrhodopsin

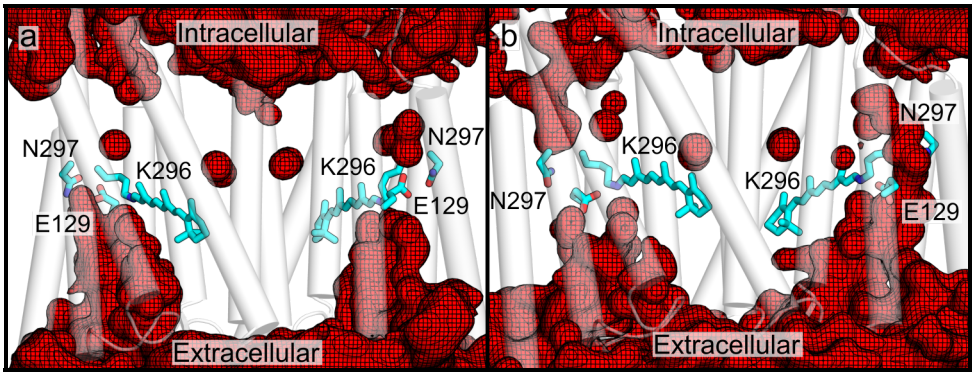


Figure B.9: Water density inside of channelrhodopsin in simWTu for the initial run (a) and the repeat simulation (b). (a) The hydrogen bond between E129 and N297 was stable throughout the simulation. (b) During the repeat simulation, he E129–N297 bond broke in both monomers. While in monomer 1 (left), no water entered the space between E129 and N297, in monomer 2, a water channel can be seen to form. Water densities were calculated for the last 25 ns of the respective trajectory.

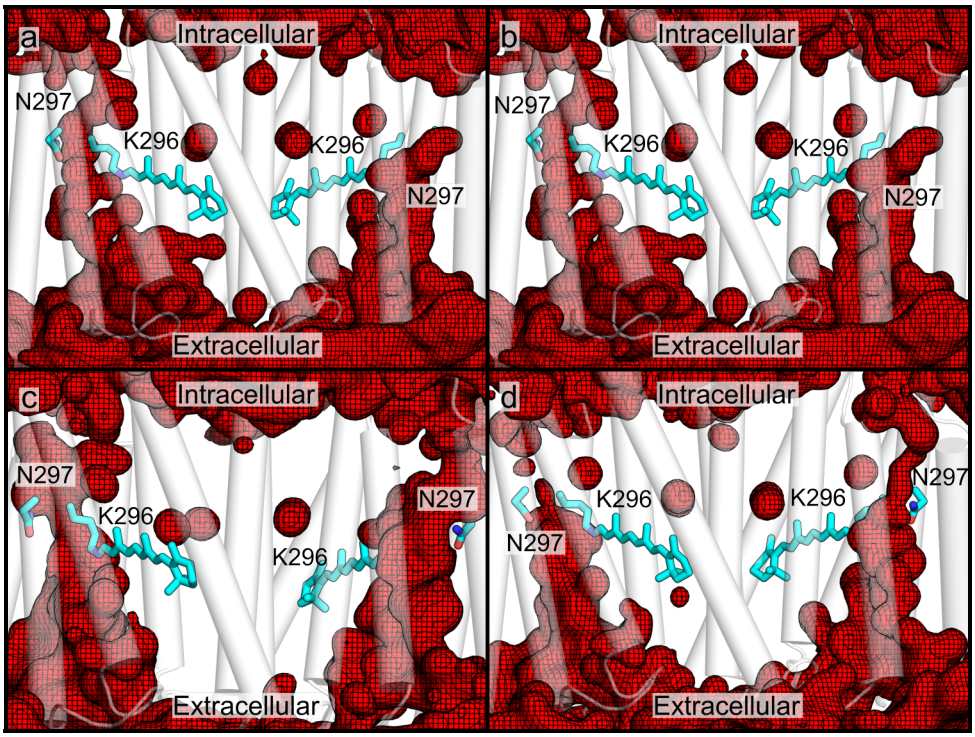


Figure B.10: Water density inside of channelrhodopsin in simW2u and simW2p for initial runs (a, b) and repeat simulations (c, d). In the mutant E129A, not E129–N297 hydrogen bond could be formed and a water pathway was seen to form in all simulations. (a) Initial run of simW2u (b) Initial run of simW2p (c) Repeat of simW2u (d) Repeat of simW2p

Papers C

C.1 | Revised CHARMM force field parameters for iron-containing cofactors of photosystem II

Adam, S., Knapp-Mohammady, M., Yi, J. & Bondar, A.-N. Revised CHARMM force field parameters for iron-containing cofactors of photosystem II. *Journal of Computational Chemistry* **39**, 7–20 (2018). `http://dx.doi.org/10.1002/jcc.24918`

For copyright reasons this publication is not included in the online version. The paper can be retrieved from `http://dx.doi.org/10.1002/jcc.24918`.

C.2 | Mechanism by Which Water and Protein Electrostatic Interactions Control Proton Transfer at the Active Site of Channelrhodopsin

Adam, S. & Bondar, A.-N. Mechanism by Which Water and Protein Electrostatic Interactions Control Proton Transfer at the Active Site of Channelrhodopsin. PLOS ONE. Under Revision.

A manuscript for this submitted publication can be found in this section.

Mechanism by Which Water and Protein Electrostatic Interactions Control Proton Transfer at the Active Site of Channelrhodopsin

Suliman Adam and Ana-Nicoleta Bondar*

Department of Physics, Freie Universität Berlin, Berlin, Germany

*Corresponding author

E-mail: nbondar@zedat.fu-berlin.de (ANB)

Abstract

Channelrhodopsins are light-sensitive ion channels whose reaction cycles involve conformation-coupled transfer of protons. Understanding how channelrhodopsins work is important for applications in optogenetics, where light activation of these proteins triggers changes in the transmembrane potential across excitable membranes. A fundamental open question is how the protein environment ensures that unproductive proton transfer from the retinal Schiff base to the nearby carboxylate counterion is avoided in the resting state of the channel. To address this question, we performed combined quantum mechanical/molecular mechanical proton transfer calculations with explicit treatment of the surrounding lipid membrane. The free energy profiles computed for proton transfer to the counterion, either via a direct jump or mediated by a water molecule, demonstrate that, when retinal is all-*trans*, water and protein electrostatic interactions largely favour the protonated retinal Schiff base state. We identified a conserved lysine group as an essential structural element for the proton transfer energetics in channelrhodopsins.

Introduction

Proton transfer reactions are of vital importance for biology(1, 2). Transfer of protons from donor to acceptor groups occurs, for example, during photosynthesis(3) and during the reaction cycles of enzymes and membrane transporters that have been implicated in human disease(4, 5). A fundamental issue for proteins whose reaction cycles involve proton transfer reactions is that the timing of the proton transfer event must be tightly controlled. A proton transfer reaction is typically coupled to changes in the protein conformational dynamics and can couple to other chemical reactions such as breaking of peptide bonds. Channelrhodopsins (ChRs) are light-driven cation channels, responsible for phototaxis in green algae(6). ChRs are used in modern neurobiological applications(7) and present an intriguing structural arrangement at their active site, where the primary proton donor—the retinal Schiff base—is found in a highly polar

environment that includes two nearby carboxylate groups. Yet, proton transfer from the retinal Schiff base to one of these nearby carboxylate group occurs only once interactions at the active site have been destabilized upon photoisomerization of the retinal chromophore from all-*trans* to 13-*cis*, thus allowing ChRs to avoid wasteful deprotonation of all-*trans* retinal. Knowledge of the molecular mechanism by which ChRs stabilize the protonated state of the retinal Schiff base would be valuable for our general understanding of structural and energetic determinants of proton transfer in polar protein environments, and could assist with fine-tuning properties of ChRs for their usage as optogenetic tools.

ChRs are members of the large family of microbial rhodopsins(8), which are transmembrane proteins characterized by a seven-transmembrane helix fold, as first observed for bacteriorhodopsin(9), with the retinal molecule covalently bound via a protonated Schiff base to a lysine amino acid residue on helix G. Among the best-studied ChRs are channelrhodopsin-1 (ChR1) and channelrhodopsin-2 (ChR2) from *Chlamydomonas reinhardtii*. Various issues pertaining to the reaction cycles of these two proteins have been characterized using electrophysiological(6, 10, 11) and spectroscopic methods(12-20). An electron microscopy structure of ChR2 has been solved at a low resolution of 6 Å (ref.(21)), and a chimaera made of helices A–E of ChR1 and helices F–G of ChR2 (C1C2) has been solved at a resolution of 2.3 Å using X-ray crystallography(22). Later, a blue-shifted C1C2 mutant was solved at 2.5 Å resolution(23), and most recently, the crystal structures of wild-type ChR2 from *C. reinhardtii* and its C128T mutant were solved at a resolution of 2.39 Å and 2.7 Å, respectively(24).

The crystal structures of C1C2(22) and of wild-type ChR2(24) placed the nitrogen atom of the protonated Schiff base within hydrogen-bonding distance of the carboxylate oxygen atoms of the nearby carboxylate groups from helices C and G (E162 and D292 in C1C2, see Fig 1b). In both crystal structures(22, 24), the retinal Schiff base lacks a direct hydrogen bond to water. This structural arrangement at the active site of ChRs is intriguing, because it raises the fundamental

question as to how ChRs avoid deprotonation of the retinal Schiff base in the all-*trans* resting state of the channel. In bacteriorhodopsin, for example, hydrogen-bonding to an active site water molecule (w402) is known to be essential for the stability of the protonated Schiff base state (25-29). Computations for proton transfer from the retinal Schiff base to the carboxylate counterion in bacteriorhodopsin (D85) indicate that the absence of w402 significantly lowers the energetic penalty for retinal deprotonation(28, 29).

Fig 1. Structure and Active Site Interactions of C1C2. (a) Cut-away view of the C1C2 dimer in a hydrated POPC lipid membrane environment. The protein coordinates are from the crystal structure of C1C2 (PDB ID: 3UG9(22)). The C1C2 monomers are shown in yellow and lime cartoons, with water oxygen atoms from the crystal structure shown as small light blue spheres. The lipid membrane is shown with white spheres for all heavy atoms, except for phosphorus and oxygen atoms, which are shown in orange and red, respectively. (b) Close-up of the retinal Schiff base region from the crystal structure. The retinal molecule and selected protein side chains are shown in stick representation with carbon, oxygen and nitrogen atoms highlighted in yellow, red and blue, respectively. The active site water molecule w19 is shown as a red sphere.

Lack of electron density for a water oxygen atom within hydrogen-bonding distance from the retinal Schiff base in the crystal structure of C1C2(22) and ChR2(24) might suggest that ChRs, unlike bacteriorhodopsin, may rely on structural elements other than hydrogen-bonding water to stabilize the proton on the Schiff base of all-*trans* retinal. Important indications that ChRs could use a different strategy for controlling the protonation state at their active sites are provided by experimental data on ChRs variants(30, 31). In *C. augustae* ChR1, the all-*trans* retinal resting state of has E169 (corresponds to E162 in C1C2) protonated and D299 (corresponds to D292 in C1C2) negatively charged(32). Deprotonation of the retinal Schiff base is thought to occur via a two-step mechanism, whereby a proton is first transferred from E169 to D299 and, in the second step, the Schiff base proton is transferred to D299(32). The observation that E169 is protonated in the resting state of *C. augustae* ChR1(32) appears compatible with initial suggestions based

on estimations of pK_a values, that E162 (Fig 1) is protonated in the crystal structure of all-*trans* C1C2(22). By contrast, FTIR spectroscopy data were interpreted to suggest that both E162 and D292 of C1C2 are negatively charged, and that E162 hydrogen-bonds directly with the retinal Schiff base(19). In the case of *C. reinhardtii* ChR2, D253 (corresponds to D292 in C1C2) was identified as the primary proton acceptor of the retinal Schiff base proton(17).

An important structural element that distinguishes some of the ChR variants is whether they contain a specific lysine group on helix B (corresponding to K132 in C1C2) or not(31) (Fig 1b). K132 is present in *C. reinhardtii* ChR1 and ChR2, but not in *C. augustae* ChR1, where it is replaced by a phenylalanine group(31, 33). Based on observations from site-directed mutagenesis, it has been proposed that K132 controls the effective charge of the counterions E162 and D292(31). Inside C1C2, the presence of complex hydrogen-bonding networks that involve K132 has been suggested by FTIR(19) and was observed in a recent molecular dynamics (MD) study(34).

The reaction mechanism for proton transfer is given by the sequence of structural changes and associated energetics along the pathway from the reactant state, where the proton is on the proton donor group and the proton acceptor group is negatively charged, to the neutral product state, where both the proton donor and acceptor groups are electrostatically neutral(35). The reaction mechanism will then be given by the transition state of the pathway(29, 35). In protein environments, where the proton donor and acceptor groups can engage in complex intramolecular interactions, more than one reaction pathway is possible(29, 36, 37). In the case of the all-*trans* retinal resting state of ChRs, stability of the protonated retinal Schiff base would require a proton transfer energy profile characterized by unfavourable reaction energetics, i.e. the free energy of the product state would need to be sufficiently larger than that of the reactant state. By dissecting the reaction energetics of such proton transfer pathways, we can

understand what the molecular interactions are that help control the protonation states at the active site of the protein.

To find out the structural and energetic determinants of the protonated retinal Schiff base state in ChRs we pursued systematic computations of proton transfers in C1C2. We used a combined quantum mechanical/molecular mechanical (QM/MM) approach(38, 39) that allows us to treat with quantum mechanics (QM) the breaking and forming of covalent bonds during proton transfer and to account for the protein environment with a simpler, classical mechanical (MM) description. Because protein flexibility(40) and water interactions(36) can impact significantly the energetics of proton transfer pathways, we performed computations of the potential of mean force (PMF) for proton transfer at 300 K, in the flexible environment of C1C2 embedded in a hydrated lipid membrane. The results of our computations indicate that K132 and active site water molecules largely stabilize the protonated Schiff state in all-*trans* retinal C1C2.

Methods

We performed MM and QM/MM all-atom MD simulations of ChR using the CHARMM36 protein(41, 42) and lipid(43) force fields, the ion parameters of Roux and coworkers(44) and the TIP3P water model(45). The retinal parameters were based on work performed by Hayashi(27, 46), Nina(47), Baudry(48) and Tajkhorshid(49-51). As QM method, we used the self-consistent charge density functional tight binding (SCC-DFTB) method(52), whose applicability to describing retinal geometry and proton transfer energetics has been documented extensively(29, 53-55).

MM simulations of all-*trans* C1C2

As starting coordinates, we used the crystal structure of the chimaera C1C2 (PDB ID: 3UG9(22)). Using the Phyre2 web portal(56), we modelled amino acid residues 110–117 that were missing from the crystal structure of C1C2. We assembled the protein dimer with PISA(57)

and linked the dimer with three disulphide bridges for C66, C73 and C75. We used standard protonation for all amino acid residues, except for E122, E129 and D195, which were considered protonated, and E162, for which both neutral and negatively charged states were generated in independent simulations. Hydrogen atoms were constructed with HBUILD in CHARMM(58). The protein dimer was placed inside a hydrated 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer using the CHARMM-GUI(59, 60) Membrane Builder(61-63). The complete system consisted of the protein dimer, 86 waters found in the crystal structure (43 per monomer), 300 POPC molecules and ~28 500 bulk water molecules in a simulation box of size 110×110×120 $Å^3$. For charge neutrality, we added two chloride ions to setups with protonated E162.

The MM simulations were performed using NAMD(64, 65) with a Langevin dynamics scheme and a Nosé–Hoover Langevin piston(66, 67) at a temperature of 300 K. We constrained covalent bonds involving hydrogen atoms using the SHAKE algorithm(68).

The first nanosecond of heating and equilibration was done following the CHARMM-GUI protocol(59, 60), while maintaining an integration time step of 1 fs and reassigning velocities every 0.5 ps. This was followed by two additional steps of equilibration of 2 ns each, where we placed on the protein and lipid heavy atoms a harmonic constraint of 5 kcal/mol/$Å^2$. Starting from the last 2 ns of the equilibration, we used to the reversible multiple time step integration scheme(69, 70) with steps of 1 fs for the bonded forces, 2 fs for short-range non-bonded and 4 fs for long-range non-bonded interactions. With the exception of the first 100 ps, where a canonical ensemble (*NVT*) was used, we used an isothermal–isobaric ensemble (*NPT*) with isotropic cell fluctuations at a pressure of 1 bar. After 5 ns of equilibration, production runs were started. The Langvin damping coefficient was 1 ps^{-1} during equilibration and 5 ps^{-1} during production runs. We generated 250 ns trajectories for each simulation of all-*trans* C1C2. We saved coordinates every 10 ps.

Modelling of the K132A mutant

To understand the impact K132 has on the proton transfer energetics of the retinal Schiff base, we performed additional calculations with the K132 to alanine (K132A) mutant of C1C2. K132A has faster kinetics than wild-type C1C2 and changes the ion selectivity of the channel from an unselective one to a potassium channel(22).

For the mutant computations, we used the wild type starting coordinates of C1C2 with protonated and unprotonated E162. We performed the equilibration as described above for wild-type C1C2, with the only exception being that the 5 kcal/mol/Å^2 harmonic constraint placed on the protein atoms was not relaxed. The equilibration was followed by 20 ns of simulation with the protein still constrained, after which we used CHARMM to mutate K132 to alanine in both monomers of C1C2. We then repeated the equilibration procedure using the same approach as described above for the wild type simulations, before prolonging the trajectories of the production runs to 110 ns each.

QM/MM computations of all-*trans* C1C2

We employed the CHARMM software package(42, 58, 71) with the SCCDFTB module(72) to perform the QM/MM calculations using third-order SCC-DFTB(52, 73, 74). To prepare the all-*trans* QM/MM systems, we took one snapshot each from the end of each of the MM trajectories. In case of the wild type simulations, we saved the coordinates at the end of the 250 ns production run. We used the same protocol for the QM/MM simulations of the K132A mutant, which were started from the end of their MM production run at 110 ns. During the first 50 ps of QM/MM dynamics, we placed on all heavy atoms a mass-scaled harmonic positional constraint given by the constraining potential:

$$U(x) = p\sum_i \left\{k_i m_i \left(x_i - x_{i,ref}\right)^2\right\},$$

where, for an atom i, k_i is the force constant of the constraint in kcal/mol/Å^2, m_i is the atomic weight, x_i is the current position in Å, $x_{i,ref}$ is the position of a reference set of coordinates in Å, and p is a pre-factor used to scale the constraint. We started with a pre-factor value of 1.0, and every 10 ps we reduced the value by 0.25 until 0 was reached. The equilibration was followed by a production run of 1 ns.

Because on the 1 ns timescale water molecules inside the Schiff base region could be replaced by water molecules initially located farther away from the Schiff base, we constrained MM water molecules at the entrance of the Schiff base region by 2.5 kcal/mol/Å^2. This keeps QM water molecules inside the Schiff base region without having to directly apply a constraining potential to them. A similar approach has been used before(75).

The QM/MM dynamics calculations were run using the Leapfrog Verlet algorithm and a canonical ensemble with a damping constant of 5 ps^{-1} at a temperature of 300 K. We used an integration time step of 1 fs and saved coordinates every 1 ps.

Choice of QM region

Because ChR is a dimer, the QM/MM treatment could be applied to either one of the active sites in the two C1C2 monomers. For simplicity, we treated the active site of only one of the protein monomers with QM, and we performed independent QM/MM MD and proton transfer computations in which the QM treatment was applied either to monomer 1 or to monomer 2 of C1C2.

The QM region consisted of the side chains of E129, K132, E162, D292 and Lys296 with the covalently bound retinal molecule (Fig 1b). Three water molecules that were located close to the Schiff base nitrogen in the starting crystal structure, were also included in the QM region (later in the text, these three water molecules are labelled as w1, w2 and w3). We used the divided frontier link atom scheme(76) and placed link atoms on CB for D292, on CD for K132 and on CG

for E129, E162 and Lys296. In K132A, we did not treat residue 132 with QM. The QM/MM computations were initiated from the end of each MM trajectory.

Proton transfer calculations

Proton transfer calculations were performed starting from the last coordinate snapshot of the QM/MM simulations using Alan Grossfield's implementation(77) of the Weighted Histogram Approach Method (WHAM)(78), an extension of the umbrella sampling method(79, 80). By using WHAM, the PMF of a reaction coordinate can be derived combining bins with different constraints placed on the reaction coordinate.

The reaction coordinate *D* of the proton transfer was defined as the difference between donor–donor hydrogen distance and acceptor–donor hydrogen distance:

$$D = d_{DH} - d_{AH},$$

where d_{DH} is the distance between donor and donor hydrogen, and d_{AH} is the distance between donor hydrogen and acceptor. We used the restrained distances (RESD) command(81) inside CHARMM to constrain *D* to generate bins for sampling the deprotonation potential of mean force. The constraint is given by:

$$E = \frac{1}{2}k(D - D_{ref}),$$

where *k* is the constraining force constant in kcal/mol/Å^2, and D_{ref} is the minimum of the constraint in Å. We used a force constant *k* of 150 kcal/mol/Å^2.

To generate starting points to begin sampling the bins for the WHAM analysis, we performed a short initial sampling as follows: Using the initial coordinates and starting from a value of $D_{ref,0}$ = −1.2 Å for an amino acid as acceptors, or $D_{ref,0}$ = −1.0 Å for a water molecule as acceptor, the RESD constraint was applied; the system was equilibrated for 2.5 ps and a restart file was saved; D_{ref} was decreased by 0.1 Å and another equilibration followed. This was done until

D_{ref} = −1.7 Å was reached. A second run was started from the initial coordinates; this time $D_{ref,0}$ was increased by 0.1 Å, followed by a 2.5 ps equilibration, until D_{ref} = 1.7 Å was reached. Once all 35 restart files had been generated, a 100 ps equilibration was started for each bin, using the respective D_{ref} value. The last 50 ps of each bin were employed for the WHAM calculations, with the values of the reaction coordinate D being saved at every time step.

To characterize structural and energetic determinants of the proton transfer energetics, we performed additional WHAM analyses with reduced systems as summarized below.

In the first test system, denoted as the PROT setup, we started from the end of the QM/MM simulation of C1C2 and deleted all lipid molecules and all water molecules farther than 5 Å away from the protein. The PMF computation on the PROT setup was then performed without additional geometry optimization. We performed 10 ps of equilibration, and then performed PMF computations using the same protocol as described above for the complete system.

In the second test system, denoted as DRY, starting from the end of the QM/MM simulation of C1C2 as well, we deleted, except for the water molecule closest to the Schiff base nitrogen, all water molecules within 4 Å of the Schiff base nitrogen or the carboxyl(ate) oxygen atoms of E162 or D292. In the DRIER setup, we deleted all water molecules within 4 Å of the Schiff base nitrogen or carboxyl(ate) oxygens without any exception. We found that the deletion of QM waters in the DRY test setup led to the need for a brief energy minimization before the initial 10 ps equilibration and the PMF computation could be performed.

Finally, the crystal setup consisted of the starting crystal structure of a C1C2 monomer (PDB ID: 3UG9(22)). In the crystal structure, there is only one water molecule (w19) in the Schiff base region; the distance between the oxygen atom of w19 and the Schiff base nitrogen of 4.43 Å, is too long for a direct hydrogen bond. Consequently, the PMF computations on the crystal test system considered only proton transfer pathways from the retinal Schiff base to E162 and D292.

To further test the importance of hydration for the energetics of the proton transfer pathways, we used the DOWSER plugin(82, 83) inside VMD(84) to generate possible missing water molecules in the Schiff base region, and then performed PMF computations. These test systems are labelled here as the DOWSER setups.

To preserve the shape of the protein during computations on the PROT, crystal and DOWSER test systems described above, we constrained coordinates of the heavy atoms of the loop regions and of all water molecules within 5 Å of the loop regions by using a harmonic constraint of 10 kcal/mol/Å^2 and 5 kcal/mol/Å^2 for loops and water molecules, respectively.

All molecular graphics were prepared using PyMOL(85).

Unless specified otherwise, all average values were computed from the last 50 ns of each MM simulation.

Results and Discussion

We performed MM and QM/MM studies on the C1C2 chimaera of *C. reinhardii* ChR (PDB ID: 3UG9(22)) to assess the impact the environment of the Schiff base has on the stability of the Schiff base proton in the dark state.

A summary of all MM simulations performed is given in Table 1. For each simulation and its repeat (indicated by a prime symbol), the C_α root-mean-square deviation (RMSD) had achieved plateau values for the last 50 ns that we used here for data analysis (S1 Fig). We note that during the last 50 ns of the wild type simulation with unprotonated E162 the RMSD for the entire protein was 2.6 ± 0.1 Å, and that the RMSD profile had reached its plateau within less than 10 ns (S1 Fig). The α-helical regions stayed even closer to the crystal structure with an RMSD of 1.1 ± 0.1 Å. This observation of a very low α-helical RMSD was shared across all simulations, with average RMSDs of 1.1–1.5 Å. Average RMSD values for the whole protein were between 2.3–3.1 Å, indicating overall good structural stability. For the loop regions, we obtained larger

RMSD values of 3.3–4.4 Å; such values are consistent with previous indications from MD simulations that the loop regions can be the most dynamic part of a protein, such that their motions cannot be sampled sufficiently on the timescale of ~100 ns(86).

Table 1. List of MM Simulations Performed for Wild-Type C1C2 and for the K132A Mutant.

Simulation	Protein	E162	Length
simWu	wild type	unprotonated	250 ns
simWu′	wild type	unprotonated	250 ns
simWp	wild type	protonated	250 ns
simWp′	wild type	protonated	250 ns
simMu	K132A	unprotonated	110 ns
simMu′	K132A	unprotonated	110 ns
simMp	K132A	protonated	110 ns
simMp′	K132A	protonated	110 ns

A prime symbol indicates repeat simulations. See S1 Fig for C_α RMSD profiles of the simulations.

In what follows, we first summarize observations from the MM simulations, with focus on internal water molecules and hydrogen bonding at the retinal Schiff base region, before proceeding to discuss the QM/MM computations.

Water molecules visit the intrahelical region of wild-type C1C2

The number and location of water molecules close to the proton transfer groups can significantly impact the proton transfer pathways and their associated energetics(29, 36). Our simulations on C1C2 indicate that, when the protein is found in a flexible, hydrated lipid membrane environment, numerous water molecules visit the interhelical region of C1C2, where they sample hydrogen bonds with protein groups. Indeed, the number of water molecules inside the interhelical region of C1C2 was significantly higher in our simulations as compared to the starting crystal structure (Fig 2). The increase in the number of water molecules was more pronounced in the extracellular half of the protein, where water molecules entered rapidly close to the Schiff base region (Figs 2 and 3). An important outcome of waters entering the extracellular half of C1C2 was that, in all simulations performed on wild-type C1C2, there were

on average ~5–6 water molecules hydrogen-bonding to either the Schiff base or E162/D292 (Figs 2 and 3, Table 2). In simulations on wild-type C1C2, regardless of the protonation state of E162, we observed the retinal Schiff base in both C1C2 monomers sampling direct hydrogen-bonding to water (Table 3).

Fig 2. Hydration of the Intrahelical Region in Wild-Type C1C2 with Unprotonated E162. (a) The number of water molecules along the membrane normal (z-axis) for water molecules within 6 Å of C1C2. The initial profile (orange) indicates the distribution of water molecules at the start of the simulations, that is, positions of waters inside C1C2 are as found in the starting crystal structure (PDB ID: 3UG9(22)). The simulation profile (blue) indicates the average number of water molecules computed for the last 50 ns of simWu (wild-type C1C2 with unprotonated E162). Note the increase in the number of water molecules inside C1C2, in particular in the extracellular half. (b) Close view of C1C2, overlaying starting crystal water oxygen atoms (orange spheres) with water oxygen atoms taken from a coordinate snapshot of simWu (blue).

Fig 3. Dynamics of the K132–D292 Interaction Compared to the Hydration of the Active Site. (a–b) Time series of the number of active site water molecules (blue profiles) and of the distance between NZ of K132 and CG of D292 (red profiles) in monomer 1 (a) and monomer 2 (b) of simWu. For each profile, thin lines indicate the calculated values, while thick lines indicate a smoothed fit. As water count we defined the number of water molecules forming hydrogen bonds with the Schiff base nitrogen or the carboxyl(ate) oxygens of E162/D292. The Pearson correlation coefficient for the K132–D292 distance and the water count was 0.63 and 0.79 for monomer 1 and monomer 2, respectively. Illustrations (1) and (2) are coordinate snapshots of monomer 1 with short (Illustration 1) and long (Illustration 2) distances between K132 and D292, as observed in the trajectory used for panel a. Note that there are more active site water molecules in Illustration 2 than in Illustration 1. Additional data analyses on the water count and the distance between E162 and D292 are given in S2 Fig.

Table 2. Number of water Molecules Forming Hydrogen Bonds with the Schiff Base Nitrogen or the Side Chains of E162 or D292.

Simulation	Number of Water Molecules	
	Monomer 1	Monomer 2
Wild-type C1C2		
simWu	6.0 ± 1.0	6.6 ± 0.6
simWu′	4.4 ± 0.9	6.6 ± 0.9
simWp	5.6 ± 0.7	4.5 ± 0.7
simWp′	4.7 ± 0.8	4.2 ± 0.9
K132A mutant		
simMu	5.5 ± 0.7	4.7 ± 0.5
simMu′	7.5 ± 1.4	6.1 ± 0.6
simMp	5.3 ± 1.0	5.7 ± 0.7
simMp′	5.2 ± 0.9	5.4 ± 1.0

A prime symbol indicates repeat simulations.

Table 3. Hydrogen-bonding Partners of the Schiff Base Nitrogen during the Last 50 ns of the MM Simulations.

Sim	Monomer 1 (%)			Monomer 2 (%)		
	Water	E162	D292	Water	E162	D292
Wild-type C1C2						
simWu	19	92	—	68	—	51
simWu′	56	38	9	85	—	36
simWp	99	—	—	57	—	68
simWp′	60	—	54	96	—	—
K132A mutant						
simMu	82	—	38	—	—	100
simMu′	64	28	22	95	—	10
simMp	37	—	85	86	—	24
simMp′	54	—	66	61	—	60

For clarity, only percentages >3 % are shown. A prime symbol indicates repeat simulations. Additional hydrogen-bonding data are summarized in S1–S4 Tables.

We note that, although the two C1C2 monomers had overall similar numbers of water molecules close to the retinal Schiff base region, for some simulations, the two monomers had slightly different levels of internal hydration (Table 2) and showed different tendencies for water hydrogen-bonding directly to the retinal Schiff base (Table 3). In SimWu′, for example, there

were 4.4 ± 0.9 and 6.6 ± 0.9 water molecules in monomer 1 and 2, respectively, and direct hydrogen bonding between the retinal Schiff base and water was sampled only transiently (19 %) in monomer 1, whereas in monomer 2 it was sampled frequently (68 %) (Table 3). These differences in the occupancies of dynamic hydrogen bonds in the two C1C2 monomers could be interpreted to suggest that full sampling of the dynamics of some of the intramolecular interactions would require simulations to be prolonged to timescales beyond the 250 ns reported here (Table 2). To account for water active site interactions as suggested by the analyses of both C1C2 monomers, the QM/MM computations for proton transfers reported below were performed separately for monomer 1 and monomer 2.

Hydrogen-Bonding Networks in the Schiff Base Region

Direct hydrogen bonding of water molecules with proton donor and/or acceptor groups can impact significantly the proton affinity and the geometry of these groups, and consequently, the proton transfer energetics (55, 87). We found that, for both wild-type C1C2 and K132A the Schiff base could hydrogen-bond only to E162, D292 or with nearby water molecules (Tables 3 and 4). This observation is compatible with the C1C2 crystal structure (PDB ID: 3UG9(22)), with suggestions from experiments that D292 is the primary proton acceptor(17, 22), that E162 could hydrogen-bond with the Schiff base(19), and with previous MD work on homology models of ChR(34, 75, 88, 89). Importantly, however, details of water hydrogen-bonding at the active site depend significantly on the protonation state of E162.

Table 4. Hydrogen-bonding Partners of the Schiff Base Nitrogen during the last 750 ps of the QM/MM Simulations of C1C2.

Sim	Monomer 1 (%)			Monomer 2 (%)		
	Water	E162	D292	Water	E162	D292
Wild-type C1C2						
simWu	90	41	—	84	—	83
simWp	93	—	15	97	—	4
K132A mutant						
simMu	93	—	14	—	—	95
simMp	5	30	91	97	—	12

For clarity, only percentages >3 % are shown.

The distance between CD of E162 and CG of D292 informs on whether the carboxylate groups hydrogen-bond directly with each other, or via one or more water molecules (Fig 4). Indeed, the average E162–D292 distance was 5.2 Å when E162 is negatively charged (Fig 4a), which corresponds to a water-mediated interaction between the two carboxylate groups. In the case of protonated E162, we observed that the distribution of the distances between E162 and D292 had a peak at 4.6 Å, corresponding to a direct hydrogen bond between E162 and D292, and a shoulder at larger values of the distance, which corresponds to water-mediated bridging of the two carboxylate groups (Fig 4b); for comparison, the distance between E162 and D292 in the starting crystal structure is 4.78 Å.(22)

Fig 4. Dependence of the Dynamics of the Interaction between E162 and D292 on the Protonation State of E162. We computed histograms of the distance between CD of E162 and CG of D292 in monomer 1 of simWu (panel a, wild-type C1C2 with unprotonated E162) and simWu′ (panel b, wild-type C1C2 with protonated E162). (a) In simWu, a one-water bridge mediates interactions between E162 and D292, and the distance between E162 and D292 is 5.1 ± 0.2 Å during the last 50 ns. (b) In simWp, D292 and the protonated E162 can either hydrogen-bond directly, resulting in the peak at ~4.5 Å, or interact via a one-water bridge, resulting in the shoulder from ~4.9 Å to ~5.4 Å. Additional data on the distance between E162 and D292 in simulations performed here are summarized in S3 Fig.

The protonation state of E162 additionally had an impact on the dynamics of its interactions with the K132 side chain (Fig 3). For negatively charged E162, K132 hydrogen-bonded with either E162 or D292, or with both residues at the same time (S4 Fig). These dynamics of the interactions between K132 and the carboxylate counterions were associated with changes in the number of water molecules close to the retinal Schiff base region (Fig 3). In simulations with E162 negatively charged, an increase in the distance between K132 and D292 led to an increase of the number of water molecules in the active site and vice versa.

To further assess the relationship between the K132–D292 distance and active site hydration, we computed the linear correlation between the distance that characterizes the K132–D292 interaction and the number of active site water molecules within hydrogen-bonding distance of either the Schiff base or E162/D292. We found that the K132–D292 distance and the number of active site waters were strongly linearly correlated in simulations with unprotonated E162, with a linear correlation of $\sim$70 %. The high correlation persisted even in the absence of direct hydrogen bonding between K132 and D292. Protonation of E162 eliminated this correlation and was associated with the K132–E162 salt bridge being broken. Taken together, these analyses suggest that K132 helps to control the access of waters to the active site of C1C2.

K132A alters internal water dynamics and protein hydrogen bonding

K132, which is present in many of the ChR sequences(33), appears to be involved in important electrostatic interactions at the active site of ChR. K132 influences the effective charge of E162 and D292(31) and affects the absorption maximum of the retinal (90). These observations are compatible with our computations above that indicated that the dynamics of the K132–D292 distance correlates with the dynamics of water at the active site (Fig 3).

To further probe the role of K132 in helping control the hydrogen-bond dynamics at the active site of C1C2, we used the crystal structure of C1C2(22) to model the K132A mutant, and we utilized MM simulations to evaluate the impact that the mutation would have on hydrogen

bonding at the active site of C1C2. We found that, when the K132 side chain was absent, R159 on helix C, which initially faced the extracellular side, could orient itself towards the active site region and engage in largely stable interactions with E162 and D292, i.e. we could associate the K132A mutation with altered dynamics of interhelical hydrogen bonds in the extracellular half of C1C2, including the counterions E162 and D292 (S4 Table).

Proton transfer pathways in all-*trans* C1C2

The MM simulations summarized above indicate that the active site of C1C2 is characterized by the presence of a dynamic hydrogen-bonding network that includes the retinal Schiff base, protein side chains and water molecules. The retinal Schiff base can interact with a water molecule, with E162 or with D292, its preferred interaction partner depending to some extent on the protonation state of E162 (Table 3). For unprotonated E162, we observed persistent hydrogen bonding between E162 and the retinal Schiff base for monomer 1 of C1C2, which was associated with lower occupancies of water hydrogen bonding of the retinal Schiff base in both simWu and the repeat simWu′ (Table 3).

To characterize the energetics of proton transfer in all-*trans* C1C2, we pursued an exhaustive set of QM/MM computations, in which we studied the dynamics and calculated proton transfer pathways for conformations of wild-type C1C2 that were representative for the active site dynamics observed. To this effect, we performed PMF computations for proton transfer from the retinal Schiff base to E162 or to D292, either directly (Type 1, Fig 5a) or via one water molecule that acted as intermediate carrier for the proton (Type 2, Fig 5b).

Fig 5. Illustrations of Proton Transfer. Proton transfer from the all-*trans* retinal Schiff base can occur via a direct jump or via an intermediate water molecule. The reaction coordinate *D* used for the proton transfer calculations is the difference between donor–hydrogen distance d_{DH} and acceptor–hydrogen distance d_{AH}. (a) In a Type 1 proton transfer pathway, the proton directly moves from the retinal Schiff base to either E162 or D292. (b) In a Type 2 proton transfer

pathway, the Schiff base proton is transferred to the carboxylate acceptor group via an intermediate water molecule.

As starting point of our QM/MM proton transfer computations, we first used the last coordinate snapshot of the MM simulations simWu and simWp (Table 1) to perform 1 ns of QM/MM equilibration for each system. The occupancies of selected active site hydrogen bonds computed for these QM/MM simulations were mostly higher than in the corresponding MM simulations (Tables 3 and 4). We suggest that these differences in the occupancies of selected hydrogen bonds in MM vs QM/MM simulations are likely due to the fact that the QM/MM simulations, whose length is limited by the computational costs, provide an incomplete picture of the dynamics of the complex hydrogen-bonding network of C1C2.

The equilibrated QM/MM simulations provided the starting point for the PMF computations of proton transfer. For simplicity, we denote as reactant state (R) structures in which the proton is located on the retinal Schiff base and as product state (P) structures in which the retinal Schiff base is deprotonated and E162 and/or D292 are protonated; structures at the energy barrier of the pathways are labelled as intermediates (I, Fig 6). The proton transfer computations performed are illustrated in Figs 6–10 and are summarized in Table 5. Given the large number of paths computed here (27 in total, see Table 5), we assigned a unique path number to each PMF proton transfer calculation. In all of our setups, proton transfer calculations via water (Pathway Type 2, Fig 5) resulted in the proton being transferred to D292 (Table 5).

Fig 6. Proton Transfer in Wild-Type C1C2. We computed PMF profiles for direct (panel a, Paths 1–4) and water-mediated (panel b, Paths 5–8) proton transfer pathways with E162 protonated (green curves) and unprotonated (black curves). PMF profiles computed for monomer 1 and monomer 2 are plotted as solid and dashed lines, respectively. Illustrations R_1, I_1 and P_1 are reactant, intermediate and product states of the direct proton transfer pathway to D292, computed with E162 protonated in monomer 2 (panel a, green dashes). The red arrows in Illustrations R_1 and I_1 indicate the direction for proton transfer. Illustrations R_2, I_2 and P_2

correspond to water-mediated proton transfer to D292 computed for monomer 2 with E162 unprotonated (black dashes in panel b).

Fig 7. Assessing the Influence of the Lipid and Water Environment on the Proton Transfer Energetics. We performed all computations for water-mediated proton transfer (Table 5). (a–b) PMF profiles computed for wild-type C1C2 and K132A with E162 unprotonated (a) and with E162 protonated (b). The PMF profiles were computed for the setups illustrated in panels c–f, namely: full (solid black line), PROT (grey line), DRY (black dashes), K132A (solid magenta line) and K132A-DRY (magenta dashes). (c–f) Representative snapshots for PMF calculations for: wild-type C1C2 with or without a lipid membrane (full/PROT, c); wild-type C1C2 with a lipid membrane, but with some active site waters removed (DRY, d); the K132A mutant in a lipid membrane (K132A, e); and K132A with a lipid membrane, but some active site waters removed (K132A-DRY, f). The corresponding path numbers are given in brackets (Table 5). Note that active site hydration, E162 protonation and the presence of the K132 side chain contribute to the stability of the proton on the retinal Schiff base.

Fig 8. Assessing the Impact of the E162 Protonation on the Proton Transfer Energetics. We computed the PMF for water-mediated proton transfer to D292 in monomer 1 of wild-type C1C2 in a hydrated lipid membrane environment (Table 5). We compared the PMF profiles of negatively charged E162 (full, solid black line, Path 5) to the PMF profile computed by first removing some of the active site water molecules (DRY, black dashes, Path 15) and the profile computed for the DRY setup, but with a proton added to E162 (DRY*, solid orange, Path 16). Note that, compared to computations with E162 unprotonated, adding a proton to E162 (DRY*) made the product state with neutral Schiff base and protonated D292 less favourable relative to the reactant state.

Fig 9. Role of Water and Protein Electrostatic Interactions in the Stability of the Protonated Retinal Schiff Base of Resting State C1C2. (a) PMF profiles computed for direct proton transfer to D292 in wild-type C1C2 with unprotonated E162. We compare the PMF profiles computed for C1C2 in a hydrated lipid membrane environment (full, black line, Path 1 in Table 5), for C1C2 with lipids and bulk water removed (PROT, grey line, Path 9) and for C1C2 with all active site waters

removed (DRIER, dashed line, Path 13). (b) Testing the impact of the K132A mutation on proton transfer. We show PMF profiles computed for direct proton transfer to D292, when E162 is protonated. We compare PMF profiles computed for wild-type C1C2 in the full (solid black line, Path 4), the PROT (solid grey, Path 10) and the DRIER setups (black dashes, Path 14) to PMF profiles computed for the K132A mutant in the full (solid magenta, Path 18) and the DRIER setup (magenta dashes, Path 21). We note that, for direct proton transfer in K132A with unprotonated E162, pathway calculations did not converge and were not used in the analysis.

Fig 10. Proton Transfer Calculations in the Crystal Structure Monomer of C1C2 in Vacuo. (a) PMF profiles for direct transfer of the Schiff base proton to D292. In the first step, we deleted w19 (Fig 1b) und considered the remaining crystal structure water molecules for the PMF calculations with E162 unprotonated (black dashes, Path 24 in Table 5) and protonated (green dashes, Path 25). In the second step, we used DOWSER to add three water molecules to the active site of the crystal structure. We then computed PMF profiles for E162 unprotonated (solid black, Path 26) and protonated (solid green, Path 27). (b) Proton transfer from E162 to D292. We performed PMF calculations for direct proton transfer from E162 to D292 in the C1C2 crystal structure without w19 (dashes) and in the crystal structure with DOWSER-added waters (solid line). In both of these PMF computations, the Schiff base was considered protonated. Note that adding waters makes the proton prefer staying on E162 instead of D292.

Table 5. List of Proton Transfer Pathways with their Respective Reaction Energies and Barrier Heights.

Path	Proton Transfer	Final Acceptor	Monomer	E162 Protonated	Reaction Energy	Barrier Height
Wild type with membrane, active site waters present (full setup)						
1	direct	E162	1	−	17.6	18.7
2	direct	D292	2	−	20.6	20.6
3	direct	D292	1	+	19.6	22.6
4	direct	D292	2	+	16.8	17.6
5	via water	D292	1	−	16.0	18.7
6	via water	D292	2	−	16.7	21.1
7	via water	D292	1	+	17.3	21.1
8	via water	D292	2	+	16.8	21.4
Wild type without membrane, active site waters present (PROT setup)						
9	direct	E162	1	−	17.7	19.5
10	direct	D292	2	+	15.5	16.3
11	via water	D292	1	−	16.1	17.6
12	via water	D292	2	+	17.4	21.1
Wild type with membrane, all active site waters removed (DRIER setup)						
13	direct	E162	1	−	12.5	13.5
14	direct	D292	2	+	19.8	19.9
Wild type with membrane, single active site water kept (DRY setup)						
15	via water	D292	1	−	11.4	16.5
16	via water	D292	2	+	15.7	19.7
17	via water	D292	1	+	19.8	20.8
K132A with membrane, active site waters present (K132A setup)						
18	direct	D292	2	+	16.6	18.6
19	via water	D292	1	−	10.0	15.8
20	via water	D292	2	+	15.2	20.9
K132A with membrane, all active site waters removed (K132A-DRIER setup)						
21	direct	D292	2	+	8.8	11.7
K132A with membrane, single active site water kept (K132A-DRY setup)						
22	via water	D292	1	−	6.6	16.1
23	via water	D292	2	+	8.8	13.9
Wild type monomer in gas-phase, no active site waters (crystal setup)						
24	direct	D292	—	−	6.1	7.2
25	direct	D292	—	+	17.9	18.1
Wild type monomer in gas-phase, DOWSER-added waters (DOWSER setup)						
26	direct	D292	—	−	19.0	19.0
27	direct	D292	—	+	17.9	19.4

Reaction energies and barrier heights are given in kcal/mol and have been computed for C1C2 containing all-*trans* retinal.

We found that proton transfer in wild-type C1C2, either direct or concerted via one water molecule, was associated with significant energy barriers of ~18–23 kcal/mol and with

unfavourable reaction energies of ~16–21 kcal/mol (Table 5, Fig 6); that is, the energetics of the PMF proton transfer pathways computed here was consistent with the notion that the protonated state of the retinal Schiff base is stable in all-*trans* C1C2.

The PMF computations for the direct proton transfers tend to largely lack clear valleys for the free energy minimum of the product state (Table 5, Fig 6a). This could be interpreted to suggest that, in these direct paths, the active site of C1C2 might be driven into energetically unfavourable geometries, and that longer sampling might be necessary to allow groups from the active site to adjust to the altered protonation state. By contrast, all water-mediated proton transfer pathways (Table 5, Fig 6b), display similar values of their reaction energies, and clear minima for the product state. This suggests that our PMF computations for proton transfer via a mediating water molecule were overall well converged.

A surprising observation from the PMF computations we performed was that, in simulation setups that include the hydrated lipid membrane environment of C1C2 (Fig 1a), the protonation state of E162 appeared to have had only a minor impact on the barrier for proton transfer. The water-mediated proton transfer pathways Path 6 (E162 unprotonated) and Path 7 (E162 protonated) were associated with largely the same reaction energetics (Table 5). Likewise, the direct proton transfer pathways Paths 2 and 3 had largely similar energy barriers of 20.6–22.6 kcal/mol (Table 5).

Impact of the lipid and water environment on proton transfer energetics

To derive further insight into the structural elements essential for the stability of the protonated state of the all-*trans* retinal Schiff base, we pursued PMF computations in which we probed the role of the lipid and water environment. For these tests we used monomer 1 of the wild-type C1C2 setup with unprotonated E162 (simWu in Table 1) and monomer 2 of the setup with protonated E162 (simWp in Table 1). For either protonation state of E162, we then computed PMF profiles for: (i) the PROT setup, in which we deleted the lipid membrane and considered

only the protein and water molecules within 5 Å of the protein (Fig 7a); (ii) the DRY setup, where we removed waters within 4 Å of the retinal Schiff base region, such that only a single water molecule close to the Schiff base remained (DRY, Fig 7b); and (iii) the DRY setup applied to K132A (K132A-DRY, Fig 7d). We compare these PMF profiles with the computations considering the full hydrated lipid membrane system for wild-type C1C2 (full, Fig 7a) and the K132A mutant (K132A, Fig 7c)

Comparison of the energetics of pathways initiated from reactant states that were the same except for the presence or absence of the lipid environment suggests that the lipid membrane has only a minor influence on the energetics of proton transfer from the Schiff base to the nearby carboxylate groups: Contrasting Path 1 with 9, Path 4 with 10, Path 5 with 11 and Path 8 with 12, we see that the effect that removing the lipid membrane environment has on the corresponding path energetics is within 1 kcal/mol (Table 5, Fig 7f). Such a small impact of lipids on the proton transfer energetics could be due to the fact that the proton transfer site is located close to the centre of the hydrophobic core of the lipid membrane (Fig 1a).

The PMF computations for water-mediated proton transfer in the absence of active site water other than the water closest to the Schiff base (Fig 7b) indicated that removal of waters from the active site leads to significantly lower reaction energies and energy barriers, when both E162 and D292 are negatively charged. In the case of Path 5 compared to Path 15, the energy barrier and reaction energy were lowered by 2.2 kcal/mol and 4.6 kcal/mol, respectively (Table 5, Fig 7e). The effect of removing waters from the active site is less pronounced when E162 is protonated, where the energy barrier decreased by 0.6 kcal/mol and the reaction energy increased by 3 kcal/mol (see Path 8 vs 17 in Table 5 and Fig 7f).

The computations above, on which water-mediated proton transfer was computed in the presence of a single active site water (Fig 7b), indicated that the energetics of Schiff base proton transfer is influenced significantly by waters hydrogen-bonding to the counterions. To further

dissect the role of waters in Schiff base proton transfer, we performed PMF computations for direct proton transfer in which we first removed all waters within 4 Å of the Schiff base, E162 and D292, i.e. we also removed the water molecule that can hydrogen-bond to the retinal Schiff base. Comparison of Paths 1 and 13 indicated that, relative to the full setup (Path 1), the reaction energy and energy barrier for direct proton transfer to the negatively charged E162 were ~5 kcal/mol lower, when active site waters were removed (Table 5, Fig 9a). For protonated E162, removing all active site water had a smaller impact on the proton transfer energetics: Comparing Path 4 and Path 14, we see that the reaction energy and the energy barrier increased by 3.0 kcal/mol and 2.3 kcal/mol, respectively.

To further test the impact of E162 protonation, we started from the DRY setup with unprotonated E162 for water-mediated proton transfer (Path 15) and changed the protonation state of E162 by placing on its carboxylate group a proton, which had been added to the system (Path 16). We then recalculated the PMF profile using this new setup (DRY*). Protonating E162 completely negated the effect of the dehydration and increased the barrier height and reaction energy to 19.7 kcal/mol and 15.7 kcal/mol, respectively (Table 5, Fig 8). The results of this protonation test were added evidence for the strong stabilizing effect protonation of E162 has on the Schiff base proton.

K132 as an important determinant of proton transfer energetics

The PMF results summarized above indicated that the protein and water environment largely stabilize the protonated state of the all-*trans* retinal Schiff base, such that proton transfer is energetically prohibitive. The analysis of the MM simulations (Fig 3 and S1–3 Tables) indicated that K132 has complex hydrogen-bond dynamics, where it always hydrogen-bonds with water and where it further samples hydrogen bonds with E136, E162 and D292. We reasoned that these interactions could affect the energetics of proton transfer to E162 or D292, and consequently, studied the dynamics of the K132A mutant with MM simulations.

The overall structure of the K132A mutant, as indicated by the C_α RMSD profile, is largely stable (S1 Fig), i.e. at least on the timescale of our simulations, the mutation does not appear associated with large structural rearrangements of the protein. Important details of the active site geometry were, however, altered by the mutation. When E162 was negatively charged it favoured interactions with R159, and R159 could additionally hydrogen-bond to D292, E136 or H288 (S1, S2 and S4 Tables). In wild-type C1C2, R159 mainly interacted with water (S4 Table). In K132A simulations with E162 protonated, E162 hydrogen-bonded to water and N297, whereas the occupancies of hydrogen bonds to D292 and R159 were small.

As discussed above for wild-type C1C2, we used the equilibrated MM simulations of K132A to perform QM/MM MD simulations followed by PMF computations for proton transfer. We found that the mutation lowered significantly the energetic penalty for water-mediated proton transfer when E162 was negatively charged. Path 19, computed for the K132A mutant, had an energy barrier and a reaction energy of 2.9 kcal/mol and 6.0 kcal/mol, respectively, smaller than the corresponding Path 5 computed for wild-type C1C2 (Fig 7e). Removing the water molecules from inside the Schiff base region of the K132A setup (K132A-DRY, Fig 7d) led to an additional decrease in reaction energy by 3.4 kcal/mol to 6.6 kcal/mol (see Path 22 in Table 5, Fig 7e, compare to Path 5 for wild-type C1C2).

When E162 is protonated, the effect of the K132A mutation on the energetics of water-mediated proton transfer to D292 was mild—within ~1 kcal/mol (compare Paths 8 and 20 in Table 5, see also Fig 7f). This result is compatible with the PMF computation for direct proton transfer in K132A with protonated E162, for which we note that the energy barrier increased by 1.0 kcal/mol and the reaction energy remained the same as computed for wild-type C1C2 (compare Paths 4 and 18 in Table 5, see Fig 9b).

Removing waters from the active site of K132A can alter drastically the energetics of proton transfer. The reaction energy and energy barrier for water-mediated proton transfer computed

for the K132A mutant with the DRY setup and protonated E162 were 8.8 kcal/mol and 13.9 kcal/mol, respectively (Path 23 in Table 5, Fig 7e–f); these energy values were ~6–7 kcal/mol lower than for the corresponding Path 20. Similarly, the direct proton transfer to D292 computed with protonated E162 and all active site waters removed (Path 21, K132A-DRIER) had its reaction energy and energy barrier ~7–8 kcal/mol lower than the corresponding Path 18 for K132A with full internal hydration.

Based on the computations discussed above, we conclude that K132 and active site waters help stabilize the protonated state of the all-*trans* retinal Schiff base. The stabilizing effect of the positively charged K132 side chain depends on the protonation state of E162. For unprotonated E162, the absence of K132 and a dehydrated Schiff base region were associated with the energetic penalty for proton transfer decreasing by ~10 kcal/mol. For protonated E162, removal of active site waters or the K132A mutation had milder effects on the proton transfer energetics, but, in the absence of both active site waters and of the K132 side chain, the reaction energy decreased by 8.0 kcal/mol. Consequently, to effectively stabilize the Schiff base proton, a protonated E162 seems to need at least either the K132 side chain or active site water molecules to bridge the interaction between E162 and D292.

Proton transfer in the crystal structure

The computations presented here support an important role of water molecules as determinants of the proton transfer energetics. As an independent test of the effect of water on the proton transfer energetics, we performed an additional set of PMF computations for an isolated monomer of C1C2 starting from the crystal structure (PDB ID: 3UG9(22)). These tests were performed for direct proton transfer using D292 as acceptor, and we considered protonated and unprotonated E162. For either protonation state of E162, we performed computations in which we used two different setups for the internal water molecules: In the first one (crystal), we removed the single crystal water that was located close to the Schiff base region (w19 in Fig 1b);

in the second one (DOWSER), we considered all crystal structure waters and used DOWSER(82, 83) to add three water molecules close to the active site. Results of these test PMF computations are summarized in Table 5 (Paths 24–27).

In the crystal system with unprotonated E162, the barrier height and reaction energy were 7.2 kcal/mol and 6.1 kcal/mol, respectively (Path 24 in Table 5, Fig 10a). For protonated E162, the energy barrier was 18.1 kcal/mol and is followed by a shallow minimum with a reaction energy of 17.9 kcal/mol (Path 25 in Table 5, Fig 10a). The DOWSER setups gave similar results for both unprotonated and protonated E162 and had energy barriers of ∼19 kcal/mol and reaction energies of 19.0 kcal/mol and 17.9 kcal/mol for unprotonated and protonated E162, respectively (Path 26 and 27 in Table 5, Fig 10a).

As an additional test for the crystal structure with protonated E162, we performed a PMF computation for proton transfer from E162 to D292. For the crystal system, we found that proton transfer was favourable, with an energy barrier of 0.1 kcal/mol and a reaction energy of −2.2 kcal/mol; that is, a protonated D292 was more favourable than a protonated E162 (10b). Once water molecules were added with DOWSER, however, proton transfer from E162 to D292 had a barrier height of 3.5 kcal/mol and a reaction energy of 1.7 kcal/mol (10b), that is, the crystal structure with added water molecules favours a protonated E162 over D292. These test computations further support our proposal that water molecules can largely shape the energetics of proton transfer in the active site of C1C2.

Overall, the results of the crystal computations were compatible with the results on the effect of water molecules on the proton transfer energetics computed for the wild type and K132A in hydrated lipid membrane environments: Active site water molecules have a pronounced impact on the energetics of proton transfers computed for negatively charged E162 and only a mild effect on paths computed for neutral E162.

Conclusions

The reaction cycles of retinal proteins involve proton transfer reactions that couple to the dynamics of the protein environment and internal water molecules. A fundamental aspect common to these proteins is the need to control the protonation state of the Schiff base of their retinal chromophore (35). An overall productive reaction cycle requires proper timing of the proton transfer reactions and protein conformational changes that might be required, for example, to ensure proper accessibility of water to the interior of the protein. Computations with isolated model compounds indicated that the proton affinities of retinal Schiff base and acetate models strongly favour the deprotonated retinal state(54, 55). The structure of the retinal binding pocket of ChRs, with its direct hydrogen bonds between the protonated retinal Schiff base and the carboxylate counterion(s)(22, 24) (Fig 1b), raises the important question of how the protein environment ensures stability of the protonated retinal Schiff base state. We addressed this question by performing systematic computations of proton transfer pathways using a QM/MM description of C1C2 in a hydrated lipid membrane environment. We performed independent computations with unprotonated and protonated E162, because the protonation state of the active site glutamate in ChRs has been controversial(17, 19, 22, 30).

Proton transfer from the all-*trans* retinal Schiff base to a nearby carboxylate group may occur via a direct jump of the proton or via a water molecule (29). The energetics of the proton transfer pathways is influenced significantly by the relative orientation of the proton donor and acceptor groups and their water interactions(27, 28, 36, 55, 87, 91), as well as by protein flexibility(40) and by the protein electrostatic environment (54, 92, 93).

We accounted for the dynamics of the protein and water interactions by performing an exhaustive set of 27 PMF computations on C1C2 at room temperature, in which we considered both the direct and water-mediated proton transfers (Table 5). The ensemble of these computations indicates that the protonated state of the retinal Schiff base is energetically

favourable in all-*trans* C1C2 and that transfer of the proton from the retinal Schiff base to E162 or D292 is associated with a significant energy penalty of ~16 kcal/mol in the lowest-energy water-mediated pathway (Table 5, Fig 6).

The high energetic cost of proton transfer from the retinal Schiff base to D292 in wild-type C1C2 ensures that the protein avoids unproductive deprotonation of the Schiff base prior to photoisomerization. Since the energetics of proton transfer computed without a protein environment strongly favours the transfer of the proton from the retinal Schiff base to a nearby carboxylate side chain(54, 94), the important question that arises is that of the molecular interactions that stabilize the protonated retinal Schiff base state in all-*trans* C1C2. We addressed this question by calculating proton transfer pathways for the wild type and for the K132A mutant using different amounts of active site water molecules, different protonation states of E162 and performing our computations with and without a lipid membrane.

We found that, for negatively charged E162, hydration of the Schiff base region contributes largely to the stability of the protonated Schiff base state. Removing water molecules from the vicinity of the retinal active site lowered the reaction energy for water-mediated proton transfer by ~5 kcal/mol (compare Paths 5 and 15 in Table 5). In the crystal structure of C1C2(22), which resolves only one water molecule close to the counterion carboxylates (Fig 1b), we find that adding water increases the energetics for Schiff base deprotonation by ~13 kcal/mol (Paths 24 and 26, Table 5).

The observation that active site water interactions are essential for the stability of the protonated retinal Schiff base in C1C2 is compatible with previous work on retinal proteins (25, 27, 55); while the observation that the active site of C1C2 contains more water molecules than indicated by the crystal structure(22) is compatible with previous experiments (19) and computations(95, 96). Water molecules that we found to be important for the energetics of proton transfer had entered the active site region during the simulations (Fig 2). This fact

highlights the importance of sampling the dynamics of membrane proteins in fluid, hydrated lipid environments.

The impact of waters on the energetics of proton transfer from the retinal Schiff base to D292 appears to depend on the protonation state of E162. In the crystal structure with protonated E162, proton transfer to D292 is 12 kcal/mol less favourable than for negatively charged E162, and the energetics remain largely the same, when water is added (Paths 25–27 in Table 5). Likewise, in membrane-embedded C1C2 with protonated E162, removing waters from the vicinity of the active site shifts the reaction energy by only ~3 kcal/mol (Paths 8 and 17 in Table 5). Changing the protonation state of E162 alters the intermolecular interactions at the active site, which in return might explain the impact of the E162 protonation state on the extent to which waters can influence the energetics of proton transfer (Figs 7, 9 and 10).

Our simulations showed that K132 has a strong effect on the proton transfer energetics in C1C2. For negatively charged E162 in K132A, the reaction energy for water-mediated deprotonation of the retinal Schiff base to D292 is 10.0 kcal/mol, which is ~6 kcal/mol less than in wild-type C1C2 (Paths 19 and 5, respectively, in Table 5). The significant role of K132 on the stability of the retinal Schiff base appears associated with its interactions with E162. In contrast to the significantly changed proton transfer energetics in K132A with negatively charged E162 (Fig 7), in the presence of protonated E162, the effect of the mutation is within 2 kcal/mol (Paths 8 and 20 in Table 5). Taken together with the significant conservation of K132 in sequences of ChRs(33) and experimental data indicating a role of K132 in controlling the charge of the counterion groups (31), the PMF computations presented here suggest that K132 is an essential determinant of the energetics of proton transfer in ChRs.

The approach used here, whereby we sampled the conformational dynamics of C1C2 and performed PMF computations on membrane-embedded C1C2 at room temperature, allows us to account explicitly for motions of the protein and its environment. In the future, we envision that

such a setup could be used to perform proton transfer calculations at the proton uptake and release sites of ChR, and thus study proton transfer at the interface between proteins and lipid membranes.

Acknowledgements

We thank Jens Dreger and Milan Hodoscek for their excellent technical support. We acknowledge the use of NAMD: NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign.

References

1. Gutman M. Intra-protein proton transfer: presentation of the most massive flux in biosphere at quantum chemistry resolution. Structure. 2004;12:1123-5.
2. Wraight CA. Chance and design - proton transfer in water, channels and bioenergetic proteins. Biochim Biophys Acta. 2006;1757:886-912.
3. Barber J. Photosystem II: a multisubunit membrane protein that oxidises water. Current Opinion Struct Biol. 2002;12:523-30.
4. Pierdominici-Sottile G, Roitberg AE. Proton transfer facilitated by ligand binding. An energetic analysis of the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. Biochemistry. 2011;50:836-42.
5. Eicher T, Seeger MA, Anselmi C, Zhou W, Brandstätter L, Verrey F, et al. Coupling of remote alternating-access transport mechanisms for protons and substrates in the multidrug efflux pump AcrB. eLife. 2014;3:e03145.
6. Sineshchekov OA, Jung KH, Spudich JL. Two rhodopsins mediate phototaxis to low- and high-intensity light in Chlamydomonas reinhardtii. Proc Natl Acad Sci U S A. 2002;99(13):8689-94.
7. Schneider F, Grimm C, Hegemann P. Biophysics of Channelrhodopsin. Annu Rev Biophys. 2015;44:167-86.
8. Ernst OP, Lodowski DT, Elstner M, Hegemann P, Brown LS, Kandori H. Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. Chem Rev. 2014;114(1):126-63.
9. Henderson R, Unwin PNT. Three-dimensional model of purple membrane obtained by electron microscopy. Nature. 1975;257:28-32.
10. Nagel G, Ollig D, Fuhrmann M, Kateriya S, Misti AM, Bamberg E, et al. Channelrhodopsin-1: A light-gated proton channel in green algae. Science. 2002;296:2395-8.
11. Nagel G, Szelles T, Huhn W, Kateriya S, Adeishvili N, Berthold P, et al. Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. Proc Natl Acad Sci. 2003;100:13940-5.
12. Ritter E, Stehfest K, Berndt A, Hegemann P, Bartl FJ. Monitoring light-induced structural changes of channlerhodopsin-2 by UV-visible and Fourier transform infrared spectroscopy. J Biol Chem. 2008;283:35033-41.
13. Radu I, Bamann C, Nack M, Nagel G, Bamberg E, Heberle J. Conformational changes of channelrhodopsin-2. J Am Chem Soc. 2009;131:7313-9.

14. Bamann C, Gueta R, Kleinlogel S, Nagel G, Bamberg E. Structural guidance of the photocycle of channelrhodopsin-2 by an interhelical hydrogen bond. Biochemistry. 2010;49:267-78.
15. Nack M, Radu I, Schultz B-J, Resler T, Schlessinger R, Bondar A-N, et al. Kinetics of proton release and uptake by channelrhodopsin-2. FEBS Lett. 2012;586:1344-8.
16. Eisenhauer K, Kuhne J, Ritter E, Berndt A, Wolf S, Freier E, et al. In channelrhodopsin-2 Glu-90 is crucial for ion selectivity and is deprotonated during the photocycle. J Biol Chem. 2012;287:6904-11.
17. Lorenz-Fonfria VA, Resler T, Krause N, Nack M, Gossing M, von Mollard GF, et al. Transient protonation changes in channelrhodpsin-2 and their relevance to channel gating. Proc Natl Acad Sci USA. 2013;110:1273-81.
18. Lorenz-Fonfria VA, Muders V, Schlesinger R, Heberle J. Changes in the hydrogen-bonding strength of water molecules and cysteine residues in the conductive state of channelrhodopsin-1. J Chem Phys. 2014;141:22D507.
19. Ito S, Kato HE, Taniguchi R, Iwata T, Nureki O, Kandori H. Water-containing hydrogen-bonding network in the active center of channelrhodopsin. J Am Chem Soc. 2014;136:3475-82.
20. Kuhne J, Eisenhauer K, Ritter E, Hegemann P, Gerwert K, Bartl F. Early formation of the ion-conducting pore in channelrhodopsin-2. Angew Chem Int Ed. 2015;54:4953-7.
21. Müller M, Bamann C, Bamberg E, Kühlbrandt W. Projection structure of channelrhodopsin-2 at 6Å resolution by electron crystallography. J Mol Biol. 2011;414:86-95.
22. Kato HE, Zhang F, Yizhar O, Ramakhrishnan C, Nishizawa T, Hirata K, et al. Crystal structure of the channelrhodopsin light-gated cation channel. Nature. 2012;482:369-74.
23. Kato HE, Kamiya M, Sugo S, Ito J, Taniguchi R, Orito A, et al. Atomistic design of microbial opsin-based blue-shifted optogenetics tools. Nat Commun. 2015;6:7177.
24. Volkov O, Kovalev K, Polovinkin V, Borschchevskiy V, Bamann C, Astashkin R, et al. Structural insights into ion conduction by channelrhodopsin 2. Science. 2017;358:eaaan8862.
25. Luecke H, Schobert B, Richter H-T, Cartailler J-P, Lanyi JK. Structure of bacteriorhdopsin at 1.55 Å resolution. J Mol Biol. 1999;291:899-911.
26. Belrhali H, Nollert P, Royant A, Menzel C, Rosenbusch JP, Landau EM, et al. Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9Å resolution. Structure. 1999;7:909-17.
27. Hayashi S, Ohmine I. Proton transfer in bacteriorhodopsin: structure, excitation, IR spectra, and potential energy surface analyses by an ab initio QM/MM method. J Phys Chem B. 2000;104:10678-91.
28. Murata K, Fujii Y, Enomoto N, Hata M, Hoshino T, Tsuda M. A study on the mechanism of the proton transport in bacteriorhodopsin: the importance of the water molecule. Biophys J. 2000;79:982-91.
29. Bondar A-N, Elstner M, Suhai S, Smith JC, Fischer S. Mechanism of primary proton transfer in bacteriorhodopsin. Structure. 2004;12:1281-8.
30. Sineshchekov OA, Govorunova EG, Wang J, Li H, Spudich JL. Intramolecular proton transfer in channelrhodopsins. Biophys J. 2013;104:807-17.
31. Li H, Govorunova EG, Sineschchekov OA, Spudich EN. Role of a helix B lysine residue in the photoactive site in channelrhodopsins. Biophys J. 2014;106:1607-17.
32. Ogren JI, Mamaev S, Li H, Spudich JL, Rotschild KJ. Proton transfers in a channelrhodopsin-1 studied by Fourier Transform Infrared (FTIR) difference spectroscopy and site-directed mutagenesis. Journal of Biological Chemistry. 2015;290:12719-30.
33. del Val C, Royuela-Flor J, Milenkovic S, Bondar A-N. Channelrhodopsins - a bioinformatics perspective. Biochim Biophys Acta Bioenergetics. 2014;1837:643-55.
34. VanGordon MR, Gyawali G, Rick SW, Rempe SB. Atomistic Study of Intramolecular Interactions in the Closed-State Channelrhodopsin Chimera, C1C2. Biophys J. 2017;112(5):943-52.
35. Bondar A-N, Smith JC. Protonation-state coupled conformational dynamics in reaction mechanisms of channel and pump rhodopsins. Photochem Photobiol. 2017;93:1336-44.

36. Bondar A-N, Baudry J, Suhai S, Fischer S, Smith JC. Key role of active-site water molecules in bacteriorhodopsin proton-transfer reactions. J Phys Chem B. 2008;112:14729-41.
37. Kiani FA, Fischer S. Catalytic strategy used by the myosin motor to hydrolyze ATP. Proc Natl Acad Sci USA. 2014;111:2947-56.
38. Singh UC, Kollman PA. An approach to computing electrostatic charges for molecules. J Comput Chem. 1984;5:129-45.
39. Field MJ, Bash PA, Karplus M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics. J Comput Chem. 1990;11:700-33.
40. Bondar A-N, Smith JC, Fischer S. Structural and energetic determinants of primary proton transfer in bacteriorhodopsin. Photochem Photobiol Sci. 2006;5:547-52.
41. MacKerell Jr. AD, Feig M, Brooks CLI. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem. 2004;25:1400-15.
42. MacKerell Jr. AD, Bashford D, Bellot M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998;102:3586-616.
43. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. J Phys Chem B. 2010;114:7830-43.
44. Beglov D, Roux B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. J Chem Phys. 1994;100:9050-63.
45. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79:926-35.
46. Hayashi S, Tajkhorshid E, Pebay-Peyroula E, Royant A, Landau EM, Navarro J, et al. Structural Determinants of Spectral Tuning in Retinal ProteinsBacteriorhodopsin vs Sensory Rhodopsin II. The Journal of Physical Chemistry B. 2001;105(41):10124-31.
47. Nina M, Roux B, Smith JC. Functional interactions in bacteriorhodopsin: a theoretical analysis of retinal hydrogen bonding with water. Biophys J. 1995;68:25-39.
48. Baudry J, Crouzy S, Roux B, Smith JC. Quantum chemical and free energy simulation analysis of retinal conformational energetics. J Chem Inf Comput Sci. 1997;37:1018-24.
49. Tajkhorshid E, Paizs B, Suhai S. Conformational effects on the proton affinity of the Schiff base in bacteriorhodopsin: A density functional study. J Phys Chem B. 1997;101:8021-8.
50. Tajkhorshid E, Suhai S. Influence of the methyl groups on the structure, charge distribution, and proton affinity of the retinal Schiff base. J Phys Chem B. 1999;103:5581-90.
51. Tajkhorshid E, Baudry J, Schulten K, Suhai S. Molecular dynamics study of the nature and origin of retinal's twisted structure in bacteriorhodopsin. Biophys J. 2000;78:683-93.
52. Elstner M, Porezag D, Jungnickel G, Elner J, Haugk M, Frauenheim T, et al. Self-consistent-charge-density-functional tight-binding method for simulations of complex material properties. Phys Rev B. 1998;58:7260-8.
53. Zhou H, Tajkhorshid E, Frauenheim T, Suhai S, Elstner M. Performance of the AM1, PM3, and SCC-DFTB methods in the study of conjugated Schiff base models. Chem Phys. 2002;277:91-103.
54. Bondar A-N, Fischer S, Smith JC, Elstner M, Suhai S. Key role of electrostatic interactions in bacteriorhodopsin proton transfer. J Am Chem Soc. 2004;126:14668-77.
55. Bondar A-N, Suhai S, Fischer S, Smith JC, Elstner M. Suppression of the back proton-transfer from Asp85 to the retinal Schiff base in bacteriorhodopsin: a theoretical analysis of structural elements. J Struct Biol. 2007;157:454-69.
56. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10(6):845-58.

57. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007;372(3):774-97.
58. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem. 1983;4:187-217.
59. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. Journal of Computational Chemistry. 2008;29:1859-65.
60. Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput. 2016;12(1):405-13.
61. Jo S, Kim T, Im W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. PLoS One. 2007;2(9):e880.
62. Jo S, Lim JB, Klauda JB, Im W. CHARMM-GUI Membrane Builder. 2008.
63. Wu EL, Cheng X, Jo S, Rui H, Song KC, Dávila-Contreras EM, et al. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J Comput Chem. 2014;35:1997-2004.
64. Kalé L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, et al. NAMD2: greater scalability for parallel molecular dynamics. J Comput Phys. 1999;151:283-312.
65. Phillips JC, Braun B, Wang W, Gumbart J, Takjkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem. 2005;26:1781-802.
66. Feller SE, Zhang Y, Pastor RW, Brooks B. Constant pressure molecular dynamics simulation: The Langevin piston method. J Chem Phys 1995;103:4613-21.
67. Martyna GJ, Tobias DJ, Klein ML. Constant-pressure molecular-dynamics algorithms. J Chem Phys. 1994;101:4177-89.
68. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints. Molecular dynamics of n-alkanes. J Comput Phys. 1977;23:327-41.
69. Grubmüller H, Heller H, Windemuth A, Schulten K. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. Mol Simul. 1991;6:121-42.
70. Tuckermann M, Berne BJ, Martyna GJ. Reversible multiple time scale molecular dynamics. J Chem Phys. 1992;97:1990-2001.
71. Brooks BR, Brooks CLI, MacKerell Jr. AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. J Comput Chem. 2009;30:1545-614.
72. Cui Q, Elstner M, Kaxiras E, Frauenheim T, Karplus M. A QM/MM implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) method. J Phys Chem B. 2001;105:569-85.
73. Gaus M, Cui Q, Elstner M. DFTB3: extension of the self-consistent charge-density functional tight-binding method (SCC-DFTB). J Chem Theory Comput. 2011;7:931-48.
74. Gaus M, Goez A, Elstner M. Parametrization and benchmark of DFTB3 for organic molecules. J Chem Theory Comput. 2013;9:338-54.
75. Watanabe H, Welke K, Schneider F, Tsunoda S, Zhang F, Deisseroth K, et al. Structural model of channelrhodopsin. J Biol Chem. 2012;287:7456-66.
76. König PH, Hoffmann M, Frauenheim T, Cui Q. A critical evaluation of different QM/MM frontier treatments with SCC-DFTB as the QM method. J Phys Chem B. 2005;109:9082-95.
77. Grossfield A. WHAM: the weighted histogram analysis method. 2.0 ed2013.
78. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. Journal of Computational Chemistry. 1992;13(8):1011-21.
79. Torrie GM, Valleau JP. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. Chemical Physics Letters. 1974;28(4):578-81.

80. Mezei M, Mehrotra PK, Beveridge DL. Monte Carlo determination of the free energy and internal energy of hydration for the Ala dipeptide at 25.degree.C. Journal of the American Chemical Society. 1985;107(8):2239-45.
81. Brooks BR. RESDistance Module in CHARMM. ; 1995.
82. Zhang L, Hermans J. Hydrophilicity of cavities in proteins. Proteins: Structure, Function, and Genetics. 1996;24(4):433-8.
83. Gumbart J, Trabuco LG, Schreiner E, Villa E, Schulten K. Regulation of the protein-conducting channel by a bound ribosome. Structure. 2009;17(11):1453-64.
84. Humphrey W, Dalke W, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14:33-8.
85. Schrödinger L. The PyMol Molecular Graphics System, Version 1.8. 2015.
86. Grossfield A, Feller SE, Pitman MC. Convergence of molecular dynamics simulations of membrane proteins. Proteins. 2007;67(1):31-40.
87. Gat Y, Sheves M. A mechanism for controlling the pKa of the retinal protonated Schiff base in retinal proteins. A study with model compounds. J Am Chem Soc. 1993;115:3772-3.
88. Watanabe H, Welke K, Sindhikara DJ, Hegemann P, Elstner M. Towards an understanding of channelrhodopsin-2 function: simulations lead to novel insights of the channel mechanism. J Mol Biol. 2013;425:1795-814.
89. Guo Y, Beyle FE, Bold BM, Watanabe HC, Koslowski A, Thiel W, et al. Active site structure and absorption spectrum of channelrhodopsin-2 wild-type and C128T mutant. Chemical Science. 2016;7(6):3879-91.
90. Dokukina I, Weingart O. Spectral properties and isomerisation path of retinal in C1C2 channelrhodopsin. Phys Chem Chem Phys. 2015;17:25142-50.
91. Scheiner S, Duan X. Effect of intermolecular orientation upon proton transfer within a polarizable medium. Biophys J. 1991;60:874-83.
92. Warshel A. Charge stabilization mechanism in the visual and purple membrane pigments. Proc Natl Acad Sci USA. 1978;75:2558-62.
93. Warshel A. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. J Biol Chem. 1998;273:27035-8.
94. Buda F, Keijer T, Ganapathy S, de Grip WJ. A Quantum-mechanical Study of the Binding Pocket of Proteorhodopsin: Absorption and Vibrational Spectra Modulated by Analogue Chromophores. Photochem Photobiol. 2017;93(6):1399-406.
95. Wietek J, Wiegert S, Adeishvili N, Schneider F, Watanabe H, Tsunoda SP, et al. Conversion of channelrhodopsin into a light-gated chloride channel. Science. 2014;344:409-12.
96. Hontani Y, Marazzi M, Stehfest K, Mathes T, van Stokkum IHM, Elstner M, et al. Reaction dynamics of the chimeric channelrhodopsin C1C2. Sci Rep. 2017;7(1):7217.

Supporting information

S1 Fig. RMSD of the Backbone C_α Atoms of C1C2. (a) RMSD profile for wild-type C1C2 with unprotonated E162 (simWu). The RMSD computed from the last 50 ns of simWu is 2.6 ± 0.1 Å for the full protein (dark green), as compared to 1.1 ± 0.1 Å and 3.7 ± 0.2 Å for the α-helical regions (dark purple) and loops (dark slate grey), respectively. (b) RMSD profile for the K132A mutant simulation with unprotonated E162 (simMu). (c) RMSD profile for wild-type C1C2 with protonated E162 (simWp). (d) RMSD profile for

K132 with protonated E162 (simMp). Results from the corresponding repeat simulations are shown in brighter colours.

S2 Fig. Water inside the Schiff Base Region in Simulations of Wild-Type C1C2 with Unprotonated (simWu) and Protonated E162 (simWp). The number of water molecules hydrogen-bonding to the Schiff base or the side chains of E162 or D292 is shown in blue and the distance between NZ of K132 and CG of D292 is shown in red. (a, d) Data from the repeat simulation of wild-type C1C2 with unprotonated E162 for monomer 1 (a) and monomer 2 (d). (b, e) Data from the simulation with protonated E162 (simWp). (c, f) Data from the repeat simulation of wild-type C1C2 with protonated E162 (simWp′).

S3 Fig. Violin Plots of the Distance between CD of E162 and CG of D292. Violin plots have been calculated for the last 50 ns of each simulation, with simulations labels according to Table 1. The inlayed box-and-whiskers plots indicate the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentile.

S4 Fig. Interactions between K132 and the Counterions E162/D292. (a) K132 hydrogen-bonding to both E162 and D292. (b) K132 hydrogen-bonding to a single counterion (E162) only.

S1 Table. Hydrogen Bonding Partners of the E162 carboxyl(ate). Percentages have been computed for the last 50 ns of the MM simulations. For clarity, only percentages >3 % are shown. A prime symbol indicates repeat simulations.

S2 Table. Hydrogen Bonding Partners of the D292 Carboxyl(ate). Percentages have been computed for the last 50 ns of the MM simulations. For clarity, only percentages >3 % are shown. A prime symbol indicates repeat simulations.

S3 Table. Hydrogen Bonding Partners of the K132 Ammonium Group. Percentages have been computed for the last 50 ns of the MM simulations. For clarity, only percentages >3 % are shown. A prime symbol indicates repeat simulations.

S4 Table. Hydrogen Bonding Partners of the R159 Side Chain Amine and Imine Groups. Percentages have been computed for the last 50 ns of the MM simulations. For clarity, only percentages >3 % are shown. A prime symbol indicates repeat simulations.

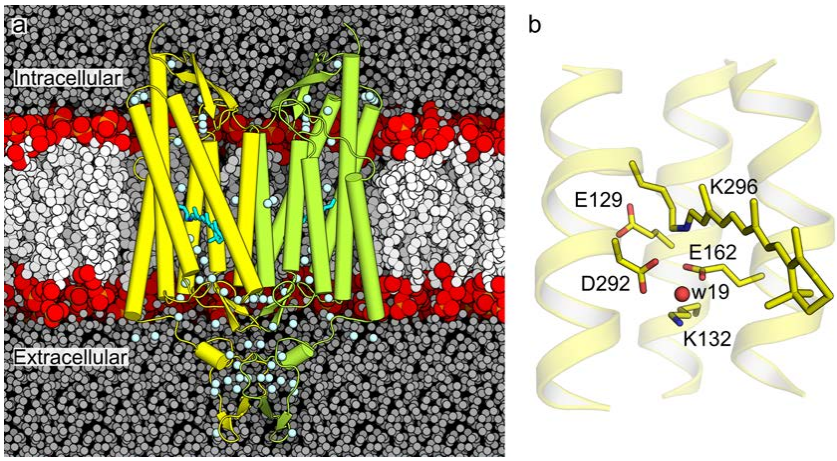

a
Intracellular
Extracellular
b
E129
K296
E162
D292
w19
K132

Figure 1

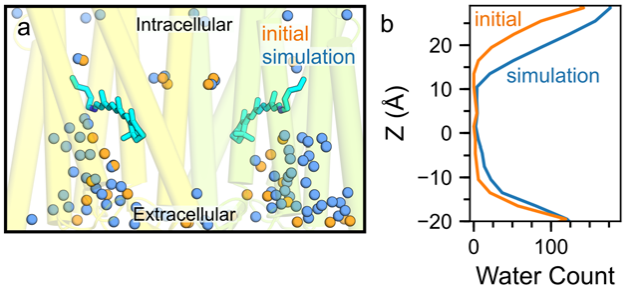

a
Intracellular
initial
simulation
Extracellular
b
initial
simulation
Z (Å)
20
10
0
−10
−20
0
100
Water Count

Figure 2.

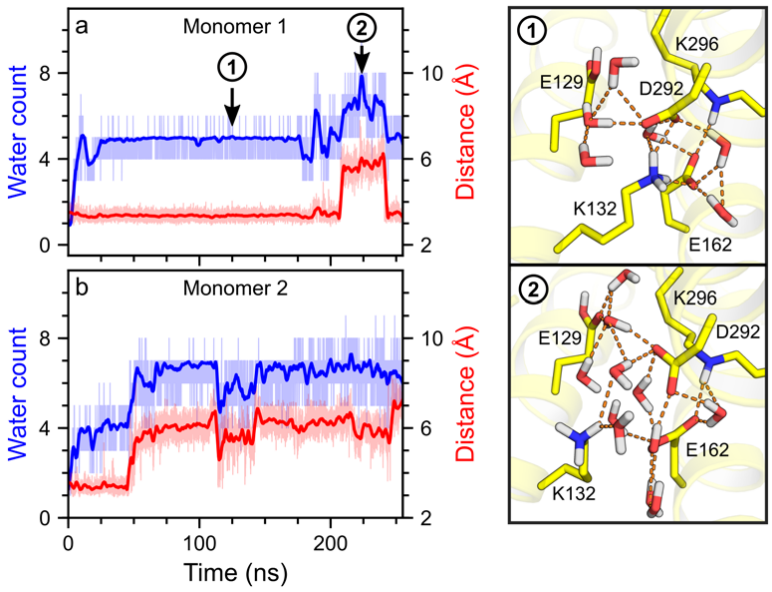

a
Monomer 1
b
Monomer 2
Water count
Distance (Å)
Time (ns)
0
100
200
0
4
8
2
6
10
1
2
K296
E129
D292
K132
E162

Figure 3.

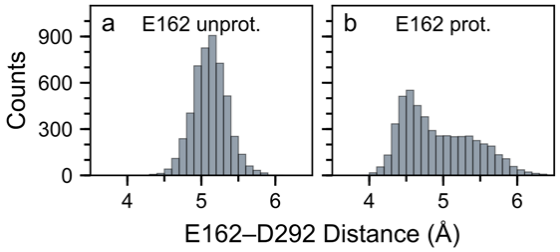

a
E162 unprot.
b
E162 prot.
Counts
900
600
300
0
4
5
6
4
5
6
E162–D292 Distance (Å)

Figure 4.

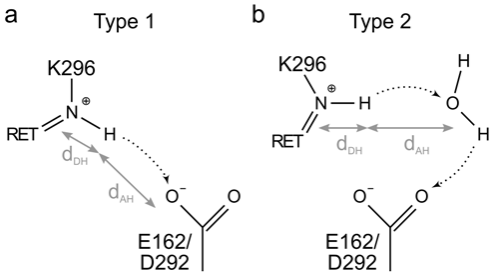

a
Type 1
K296
RET
N
H
d_{DH}
d_{AH}
O
O
E162/
D292
b
Type 2
K296
RET
N
H
O
H
H
d_{DH}
d_{AH}
O
O
E162/
D292

Figure 5.

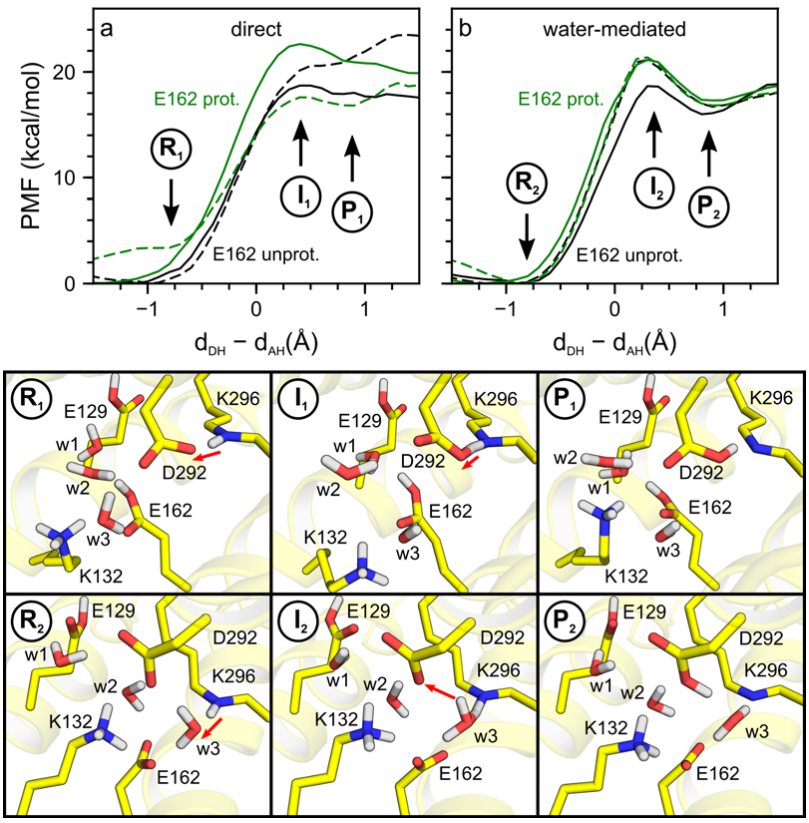

a
direct
b
water-mediated
E162 prot.
E162 unprot.
PMF (kcal/mol)
20
10
0
−1
0
1
$d_{DH} - d_{AH}$(Å)
R_1
I_1
P_1
R_2
I_2
P_2
E129
K296
D292
E162
K132
w1
w2
w3

Figure 6.

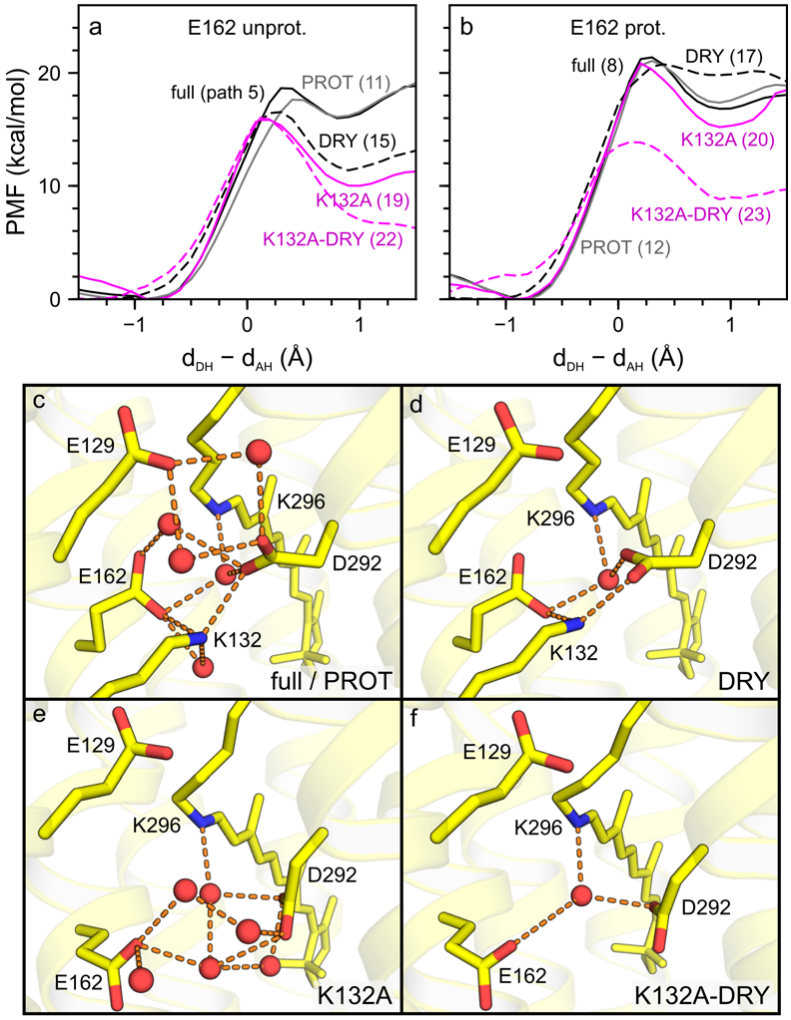

a
E162 unprot.
full (path 5)
PROT (11)
DRY (15)
K132A (19)
K132A-DRY (22)
PMF (kcal/mol)
20
10
0
−1
0
1
$d_{DH} - d_{AH}$ (Å)
b
E162 prot.
full (8)
DRY (17)
K132A (20)
K132A-DRY (23)
PROT (12)
−1
0
1
$d_{DH} - d_{AH}$ (Å)
c
E129
K296
D292
E162
K132
full / PROT
d
E129
K296
D292
E162
K132
DRY
e
E129
K296
D292
E162
K132A
f
E129
K296
D292
E162
K132A-DRY

Figure 7.

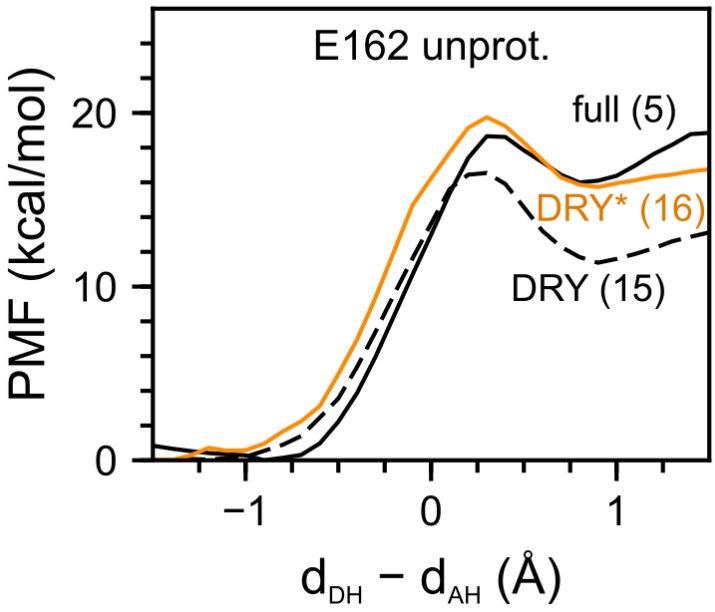

E162 unprot.
full (5)
DRY* (16)
DRY (15)
PMF (kcal/mol)
20
10
0
−1
0
1
d_DH − d_AH (Å)

Figure 8.

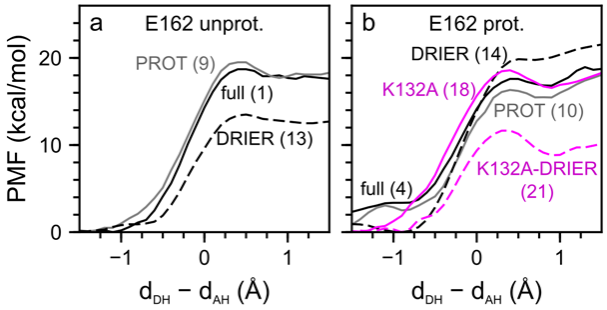

a
E162 unprot.
PROT (9)
full (1)
DRIER (13)
PMF (kcal/mol)
20
10
0
−1
0
1
d_DH − d_AH (Å)
b
E162 prot.
DRIER (14)
K132A (18)
PROT (10)
K132A-DRIER
(21)
full (4)
−1
0
1
d_DH − d_AH (Å)

Figure 9.

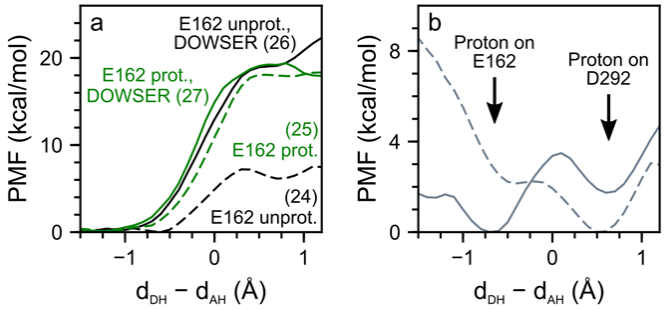

a
E162 unprot.,
DOWSER (26)
E162 prot.,
DOWSER (27)
(25)
E162 prot.
(24)
E162 unprot.
PMF (kcal/mol)
20
10
0
−1
0
1
$d_{DH} - d_{AH}$ (Å)
b
Proton on
E162
Proton on
D292
PMF (kcal/mol)
8
4
0
−1
0
1
$d_{DH} - d_{AH}$ (Å)

Figure 10.

Bibliography

1. Morrison, L. J., Gizis, F. & Shorter, B. Prevalent use of dietary supplements among people who exercise at a commercial gym. *International Journal of Sport Nutrition and Exercise Metabolism* **14**, 481–492 (2004). Cited on page 3.

2. Gao, M. *et al.* Multidrug resistance protein. *Journal of Biological Chemistry* **273**, 10733–10740 (1998). Cited on page 4.

3. Chang, G. Multidrug resistance ABC transporters. *FEBS Letters* **555**, 102–105 (2003). Cited on page 4.

4. Sharom, F. J. ABC multidrug transporters: structure, function and role in chemoresistance. *Pharmacogenomics* **9**, 105–127 (2008). Cited on page 4.

5. Zaidi, A. H. *et al.* The ABC-type multidrug resistance transporter LmrCD is responsible for an extrusion-based mechanism of bile acid resistance in lactococcus lactis. *Journal of Bacteriology* **190**, 7357–7366 (2008). Cited on page 4.

6. Tosteson, D. C. Regulation of cell volume by active cation transport in high and low potassium sheep red cells. *The Journal of General Physiology* **44**, 169–194 (1960). Cited on page 4.

7. Roux, B. *et al.* Ion selectivity in channels and transporters. *The Journal of General Physiology* **137**, 415–426 (2011). Cited on page 4.

8. Paoli, M., Liddington, R., Tame, J., Wilkinson, A. & Dodson, G. Crystal structure of t state haemoglobin with oxygen bound at all four haems. *Journal of Molecular Biology* **256**, 775–792 (1996). Cited on page 4.

9. Umena, Y., Kawakami, K., Shen, J.-R. & Kamiya, N. Crystal structure of oxygen-evolving photosystem ii at a resolution of 1.9 å. *Nature* **473**, 55–60 (2011). Cited on pp. 4, 5, 6, 21, 52, 55, 68, and 71.

10. Suga, M. *et al.* Native structure of photosystem II at 1.95 åresolution viewed by femtosecond x-ray pulses. *Nature* **517**, 99–103 (2014). Cited on page 4.

11. Kandel, G. L. Human color vision, by robert m. boynton, holt, rinehart and winston, new york, 1979, 448 pp. price: $29.95. *Color Research & Application* **5**, 185–186 (1980). Cited on page 4.

12. Merbs, S. L. & Nathans, J. Absorption spectra of human cone pigments. *Nature* **356**, 433–435 (1992). Cited on page 4.

13. Kochendoerfer, G. G., Lin, S. W., Sakmar, T. P. & Mathies, R. A. How color visual pigments are tuned. *Trends in Biochemical Sciences* **24**, 300–305 (1999). Cited on page 4.

14. Remy, A. & Gerwert, K. Coupling of light-induced electron transfer to proton uptake in photosynthesis. *Nature Structural Biology* **10**, 637–644 (2003). Cited on pp. 4 and 72.

15. Chernev, P., Zaharieva, I., Dau, H. & Haumann, M. Carboxylate shifts steer interquinone electron transfer in photosynthesis. *Journal of Biological Chemistry* **286**, 5368–5374 (2010). Cited on pp. 4 and 72.

16. Nic, M., Jirát, J., Košata, B., Jenkins, A. & McNaught, A. (eds.) *IUPAC Compendium of Chemical Terminology* (IUPAC, 2009). Cited on page 4.

17. Walker, F. A. Models of the bis-histidine-ligated electron-transferring cytochromes. comparative geometric and electronic structure of low-spin ferro- and ferrihemes. *Chemical Reviews* **104**, 589–616 (2004). Cited on pp. 4 and 71.

18. Berry, E. A. & Walker, F. A. Bis-histidine-coordinated hemes in four-helix bundles: how the geometry of the bundle controls the axial imidazole plane orientations in transmembrane cytochromes of mitochondrial complexes II and III and related proteins. *JBIC Journal of Biological Inorganic Chemistry* **13**, 481–498 (2008). Cited on pp. 4 and 71.

19. McEvoy, J. P. & Brudvig, G. W. Redox reactions of the non-heme iron in photosystem II: An EPR spectroscopic study†. *Biochemistry* **47**, 13394–13403 (2008). Cited on pp. 4, 22, and 72.

20. Kato, Y. & Noguchi, T. Long-range interaction between the mn4cao5cluster and the non-heme iron center in photosystem II as revealed by FTIR spectroelectrochemistry. *Biochemistry* **53**, 4914–4923 (2014). Cited on page 4.

21. Chu, H.-A. & Chiu, Y.-F. The roles of cytochrome b559 in assembly and photoprotection of photosystem II revealed by site-directed mutagenesis studies. *Frontiers in Plant Science* **6** (2016). Cited on page 4.

22. Ishikita, H. & Knapp, E.-W. Oxidation of the non-heme iron complex in photosystem II†. *Biochemistry* **44**, 14772–14783 (2005). Cited on pp. 4 and 22.

23. Sugiura, M., Nakamura, M., Koyama, K. & Boussac, A. Assembly of oxygen-evolving photosystem II efficiently occurs with the apo-cyt b 559 but the holo-cyt b 559 accelerates the recovery of a functional enzyme upon photoinhibition. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1847**, 276–285 (2015). Cited on page 5.

24. Hung, C.-H., Huang, J.-Y., Chiu, Y.-F. & Chu, H.-A. Site-directed mutagenesis on the heme axial-ligands of cytochrome b559 in photosystem II by using cyanobacteria synechocystis PCC 6803. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1767**, 686–693 (2007). Cited on page 5.

25. Müh, F. & Zouni, A. The nonheme iron in photosystem II. *Photosynthesis Research* **116**, 295–314 (2013). Cited on page 6.

26. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4**, 187–217 (1983). Cited on pp. 6, 17, 23, 28, 30, and 37.

27. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* NA–NA (2009). Cited on pp. 6, 15, 21, 23, 24, 25, 26, 27, 28, 43, 45, 47, 48, 51, 72, and 133.

28. Costanzi, S., Siegel, J., Tikhonova, I. & Jacobson, K. Rhodopsin and the others: A historical perspective on structural studies of g protein-coupled receptors. *Current Pharmaceutical Design* **15**, 3994–4002 (2009). Cited on page 6.

29. Sudo, Y. *et al.* A microbial rhodopsin with a unique retinal composition shows both sensory rhodopsin II and bacteriorhodopsin-like properties. *Journal of Biological Chemistry* **286**, 5967–5976 (2010). Cited on page 6.

30. Ernst, O. P. *et al.* Microbial and animal rhodopsins: Structures, functions, and molecular mechanisms. *Chemical Reviews* **114**, 126–163 (2014). Cited on page 6.

31. Kato, H. E. *et al.* Atomistic design of microbial opsin-based blue-shifted optogenetics tools. *Nature Communications* **6**, 7177 (2015). Cited on pp. 6, 7, 9, and 80.

32. Bondar, A.-N. & Smith, J. C. Protonation-state-coupled conformational dynamics in reaction mechanisms of channel and pump rhodopsins. *Photochemistry and Photobiology* **93**, 1336–1344 (2017). Cited on page 6.

33. Oesterhelt, D. & Stoeckenius, W. Rhodopsin-like protein from the purple membrane of halobacterium halobium. *Nature New Biology* **233**, 149–152 (1971). Cited on page 6.

34. Walter, J. M., Greenfield, D. & Liphardt, J. Potential of light-harvesting proton pumps for bioenergy applications. *Current Opinion in Biotechnology* **21**, 265–270 (2010). Cited on page 6.

35. Henderson, R. *et al.* Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *Journal of Molecular Biology* **213**, 899–929 (1990). Cited on page 6.

36. Berman, H. M. The protein data bank. *Nucleic Acids Research* **28**, 235–242 (2000). Cited on page 6.

37. Hayashi, S. *et al.* Structural determinants of spectral tuning in retinal proteins - bacteriorhodopsin vs sensory rhodopsin ii. *The Journal of Physical Chemistry B* **105**, 10124–10131 (2001). Cited on pp. 7, 29, and 103.

38. Houjou, H., Inoue, Y. & Sakurai, M. Physical origin of the opsin shift of bacteriorhodopsin. comprehensive analysis based on medium effect theory of absorption spectra. *Journal of the American Chemical Society* **120**, 4459–4470 (1998). Cited on page 7.

39. Nielsen, M. B. Model systems for understanding absorption tuning by opsin proteins. *Chemical Society Reviews* **38**, 913 (2009). Cited on page 7.

40. Melaccio, F., Ferré, N. & Olivucci, M. Quantum chemical modeling of rhodopsin mutants displaying switchable colors. *Physical Chemistry Chemical Physics* **14**, 12485 (2012). Cited on page 7.

41. Kamiya, M., Kato, H. E., Ishitani, R., Nureki, O. & Hayashi, S. Structural and spectral characterizations of c1c2 channelrhodopsin and its mutants by molecular simulations. *Chemical Physics Letters* **556**, 266–271 (2013). Cited on pp. 7, 30, 86, 87, and 112.

42. Sudo, Y. *et al.* A blue-shifted light-driven proton pump for neural silencing. *Journal of Biological Chemistry* **288**, 20624–20632 (2013). Cited on page 7.

43. Nina, M., Roux, B. & Smith, J. Functional interactions in bacteriorhodopsin: a theoretical analysis of retinal hydrogen bonding with water. *Biophysical Journal* **68**, 25–39 (1995). Cited on pp. 7 and 29.

44. Baudry, J., Crouzy, S., Roux, B. & Smith, J. C. Quantum chemical and free energy simulation analysis of retinal conformational energetics. *Journal of Chemical Information and Computer Sciences* **37**, 1018–1024 (1997). Cited on pp. 7 and 29.

45. Tajkhorshid, E., Paizs, B. & Suhai, S. Conformational effects on the proton affinity of the schiff base in bacteriorhodopsin: a density functional study. *The Journal of Physical Chemistry B* **101**, 8021–8028 (1997). Cited on pp. 7 and 29.

46. Tajkhorshid, E. & Suhai, S. Influence of the methyl groups on the structure, charge distribution, and proton affinity of the retinal schiff base. *The Journal of Physical Chemistry B* **103**, 5581–5590 (1999). Cited on pp. 7 and 29.

47. Tajkhorshid, E., Baudry, J., Schulten, K. & Suhai, S. Molecular dynamics study of the nature and origin of retinal's twisted structure in bacteriorhodopsin. *Biophysical Journal* **78**, 683–693 (2000). Cited on pp. 7 and 29.

48. Hayashi, S. & Ohmine, I. Proton transfer in bacteriorhodopsin: Structure, excitation, ir spectra, and potential energy surface analyses by an ab initio QM/MM method. *The Journal of Physical Chemistry B* **104**, 10678–10691 (2000). Cited on pp. 7, 29, and 103.

49. Gat, Y. & Sheves, M. A mechanism for controlling the pKa of the retinal protonated schiff base in retinal proteins. a study with model compounds. *Journal of the American Chemical Society* **115**, 3772–3773 (1993). Cited on pp. 7 and 103.

50. Bondar, A.-N., Elstner, M., Suhai, S., Smith, J. C. & Fischer, S. Mechanism of primary proton transfer in bacteriorhodopsin. *Structure* **12**, 1281–1288 (2004). Cited on pp. 7 and 29.

51. Bondar, A.-N., Fischer, S., Smith, J. C., Elstner, M. & Suhai, S. Key role of electrostatic interactions in bacteriorhodopsin proton transfer. *Journal of the American Chemical Society* **126**, 14668–14677 (2004). Cited on pp. 7 and 29.

52. Bondar, A.-N., Suhai, S., Fischer, S., Smith, J. C. & Elstner, M. Suppression of the back proton-transfer from asp85 to the retinal schiff base in bacteriorhodopsin: a theoretical analysis of structural elements. *Journal of Structural Biology* **157**, 454–469 (2007). Cited on pp. 7, 29, 103, and 113.

53. Hegemann, P. Algal sensory photoreceptors. *Annual Review of Plant Biology* **59**, 167–189 (2008). Cited on page 7.

54. Schneider, F., Grimm, C. & Hegemann, P. Biophysics of channelrhodopsin. *Annual Review of Biophysics* **44**, 167–186 (2015). Cited on page 7.

55. Müller, M., Bamann, C., Bamberg, E. & Kühlbrandt, W. Projection structure of channelrhodopsin-2 at 6 AA resolution by electron crystallography. *Journal of Molecular Biology* **414**, 86–95 (2011). Cited on page 7.

56. Berthold, P. *et al.* Channelrhodopsin-1 initiates phototaxis and photophobic responses in *Chlamydomonas* by immediate light-induced depolarization. *The Plant Cell* **20**, 1665–1677 (2008). Cited on pp. 7 and 8.

57. Sineshchekov, O. A., Jung, K.-H. & Spudich, J. L. Two rhodopsins mediate phototaxis to low- and high-intensity light in chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences* **99**, 8689–8694 (2002). Cited on page 8.

58. Nagel, G. Channelrhodopsin-1: A light-gated proton channel in green algae. *Science* **296**, 2395–2398 (2002). Cited on page 8.

59. Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proceedings of the National Academy of Sciences* **100**, 13940–13945 (2003). Cited on pp. 8 and 87.

60. Lórenz-Fonfría, V. A. & Heberle, J. Channelrhodopsin unchained: Structure and mechanism of a light-gated cation channel. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1837**, 626–642 (2014). Cited on pp. 8, 9, and 95.

61. Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K. Optogenetics in neural systems. *Neuron* **71**, 9–34 (2011). Cited on page 8.

62. Macé, E. *et al.* Targeting channelrhodopsin-2 to ON-bipolar cells with vitreally administered AAV restores ON and OFF visual responses in blind mice. *Molecular Therapy* **23**, 7–16 (2015). Cited on page 8.

63. Bruegmann, T. *et al.* Optogenetic control of contractile function in skeletal muscle. *Nature Communications* **6**, 7153 (2015). Cited on page 8.

64. Kato, H. E. *et al.* Crystal structure of the channelrhodopsin light-gated cation channel. *Nature* **482**, 369–374 (2012). Cited on pp. 8, 29, 30, 31, 32, 36, 77, 78, 79, 80, 81, 83, 87, 88, 92, 93, 94, 95, 105, 106, 111, and 112.

65. Inaguma, A. *et al.* Chimeras of channelrhodopsin-1 and -2 from *Chlamydomonas reinhardtii* exhibit distinctive light-induced structural changes from channelrhodopsin-2. *Journal of Biological Chemistry* **290**, 11623–11634 (2015). Cited on pp. 8, 83, and 88.

66. Wang, H. *et al.* Molecular determinants differentiating photocurrent properties of two channelrhodopsins fromChlamydomonas. *Journal of Biological Chemistry* **284**, 5685–5696 (2008). Cited on page 8.

67. Volkov, O. *et al.* Structural insights into ion conduction by channelrhodopsin 2. *Science* **358**, eaan8862 (2017). Cited on pp. 9, 81, 109, and 118.

68. Schrödinger, E. An undulatory theory of the mechanics of atoms and molecules. *Physical Review* **28**, 1049–1070 (1926). Cited on page 12.

69. Friesner, R. A. Ab initio quantum chemistry: methodology and applications. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 6648–6653 (2005). Cited on page 12.

70. Parr, R. G., Craig, D. P. & Ross, I. G. Molecular orbital calculations of the lower excited electronic levels of benzene, configuration interaction included. *The Journal of Chemical Physics* **18**, 1561–1563 (1950). Cited on page 12.

71. Head-Gordon, M., Pople, J. A. & Frisch, M. J. MP2 energy evaluation by direct methods. *Chemical Physics Letters* **153**, 503–506 (1988). Cited on pp. 12 and 15.

72. Cramer, C. J. *Essentials of Computational Chemistry* (John Wiley and Sons, Ltd., 2004). Cited on pp. 12, 13, and 14.

73. Roothaan, C. C. J. New developments in molecular orbital theory. *Reviews of Modern Physics* **23**, 69–89 (1951). Cited on page 13.

74. Jensen, F. *Introduction to Computational Chemistry* (John Wiley and Sons, Ltd., 2007), 2nd edn. Cited on page 14.

75. Leininger, M. L., Allen, W. D., Schaefer, H. F. & Sherrill, C. D. Is moller-plesset perturbation theory a convergent *ab initio* method. *The Journal of Chemical Physics* **112**, 9213–9222 (2000). Cited on page 14.

76. He, Y. & Cremer, D. Molecular geometries at sixth order møller-plesset perturbation theory. at what order does MP theory give exact geometries? *The Journal of Physical Chemistry A* **104**, 7679–7688 (2000). Cited on page 14.

77. Cremer, D. & He, Z. Sixth-order møller-plesset perturbation TheoryOn the convergence of the MPnSeries. *The Journal of Physical Chemistry* **100**, 6173–6188 (1996). Cited on page 14.

78. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Physical Review* **136**, B864–B871 (1964). Cited on page 15.

79. Ziegler, T. Approximate density functional theory as a practical tool in molecular energetics and dynamics. *Chemical Reviews* **91**, 651–667 (1991). Cited on page 15.

80. Jensen, F. *Introduction to Computational Chemistry* (John Wiley and Sons, Ltd., 2017), 3rd edn. Cited on pp. 15 and 16.

81. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **38**, 3098–3100 (1988). Cited on page 15.

82. Lee, C., Yang, W. & Parr, R. G. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **37**, 785–789 (1988). Cited on page 15.

83. Miehlich, B., Savin, A., Stoll, H. & Preuss, H. Results obtained with the correlation energy density functionals of becke and lee, yang and parr. *Chemical Physics Letters* **157**, 200–206 (1989). Cited on page 15.

84. Becke, A. D. Density-functional thermochemistry. III. the role of exact exchange. *The Journal of Chemical Physics* **98**, 5648–5652 (1993). Cited on page 16.

85. Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry* **98**, 11623–11627 (1994). Cited on page 16.

86. Vosko, S. H., Wilk, L. & Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics* **58**, 1200–1211 (1980). Cited on page 16.

87. Thiel, W. Semiempirical quantum-chemical methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**, 145–157 (2013). Cited on page 16.

88. Elstner, M. *et al.* Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Physical Review B* **58**, 7260–7268 (1998). Cited on pp. 16 and 29.

89. Elstner, M., Frauenheim, T., Kaxiras, E., Seifert, G. & Suhai, S. A self-consistent charge density-functional based tight-binding scheme for large biomolecules. *physica status solidi (b)* **217**, 357–376 (2000). Cited on pp. 16 and 29.

90. Yang, Yu, H., York, D., Cui, Q. & Elstner, M. Extension of the self-consistent-charge density-functional tight-binding method: third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction. *The Journal of Physical Chemistry A* **111**, 10861–10873 (2007). Cited on pp. 16 and 29.

91. Gaus, M., Cui, Q. & Elstner, M. DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *Journal of Chemical Theory and Computation* **7**, 931–948 (2011). Cited on pp. 16 and 29.

92. Elstner, M., Hobza, P., Frauenheim, T., Suhai, S. & Kaxiras, E. Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *The Journal of Chemical Physics* **114**, 5149–5155 (2001). Cited on page 16.

93. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-d) for the 94 elements h-pu. *The Journal of Chemical Physics* **132**, 154104 (2010). Cited on page 16.

94. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **32**, 1456–1465 (2011). Cited on page 16.

95. MacKerell, A. D. *et al.* CHARMM: The energy function and its parameterization. *Encyclopedia of Computational Chemistry* (2002). Cited on pp. 17 and 37.

96. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry* **34**, 2135–2145 (2013). Cited on page 17.

97. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. *Journal of chemical theory and computation* **8**, 3257–3273 (2012). Cited on pp. 17, 21, 24, 27, 28, 29, 43, 46, 51, 123, 131, 132, 134, and 135.

98. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **30**, 1545–1614 (2009). Cited on page 17.

99. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B* **102**, 3586–3616 (1998). Cited on pp. 17, 19, 21, 28, 29, 44, 72, 131, 133, and 134.

100. MacKerell, A. D., Feig, M. & Brooks, C. L. Improved treatment of the protein backbone in empirical force fields. *Journal of the American Chemical Society* **126**, 698–699 (2004). Cited on page 17.

101. Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of computational chemistry* **25**, 1400–1415 (2004). Cited on pp. 17, 28, and 29.

102. Mackerell, A. D. Empirical force fields for biological macromolecules: overview and issues. *Journal of computational chemistry* **25**, 1584–1604 (2004). Cited on page 18.

103. Halgren, T. A. Merck molecular force field. i. basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* **17**, 490–519 (1996). Cited on page 18.

104. Maple, J. R. *et al.* Derivation of class II force fields. i. methodology and quantum force field for the alkyl functional group and alkane molecules. *Journal of Computational Chemistry* **15**, 162–182 (1994). Cited on page 18.

105. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water.

The Journal of Chemical Physics **79**, 926–935 (1983). Cited on pp. 18, 19, 28, and 29.

106. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *The Journal of Physical Chemistry* **91**, 6269–6271 (1987). Cited on page 18.

107. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. Interaction models for water in relation to protein hydration. In *The Jerusalem Symposia on Quantum Chemistry and Biochemistry*, 331–342 (Springer Netherlands, 1981). Cited on page 18.

108. Mark, P. & Nilsson, L. Structure and dynamics of the TIP3p, SPC, and SPC/e water models at 298 k. *The Journal of Physical Chemistry A* **105**, 9954–9960 (2001). Cited on page 18.

109. Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics* **112**, 8910–8922 (2000). Cited on page 19.

110. Nutt, D. R. & Smith, J. C. Molecular dynamics simulations of proteins: can the explicit water model be varied? *Journal of Chemical Theory and Computation* **3**, 1550–1560 (2007). Cited on page 19.

111. Hayward, J. A. & Smith, J. C. Temperature dependence of protein dynamics: computer simulation analysis of neutron scattering properties. *Biophysical journal* **82**, 1216–1225 (2002). Cited on page 19.

112. Cui, Q., Elstner, M., Kaxiras, E., Frauenheim, T. & Karplus, M. A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *The Journal of Physical Chemistry B* **105**, 569–585 (2001). Cited on pp. 19, 20, and 32.

113. König, P. H., Hoffmann, M., Frauenheim, T. & Cui, Q. A critical evaluation of different QM/MM frontier treatments with SCC-DFTB as the QM method. *The Journal of Physical Chemistry B* **109**, 9082–9095 (2005). Cited on pp. 19, 20, 35, and 118.

114. Das, D. *et al.* Optimization of quantum mechanical molecular mechanical partitioning schemes: gaussian delocalization of molecular mechanical

charges and the double link atom method. *The Journal of Chemical Physics* **117**, 10534–10547 (2002). Cited on page 20.

115. Waszkowycz, B., Hillier, I. H., Gensmantel, N. & Payling, D. W. A combined quantum mechanical/molecular mechanical model of the potential energy surface of ester hydrolysis by the enzyme phospholipase a_2. *Journal of the Chemical Society, Perkin Transactions 2* 225 (1991). Cited on page 20.

116. Riccardi, D. *et al.* Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *The Journal of Physical Chemistry B* **110**, 6458–6469 (2006). Cited on page 20.

117. Riccardi, D., Schaefer, P. & Cui, Q. pKaCalculations in solution and proteins with QM/MM free energy perturbation simulations: a quantitative test of QM/MM protocols. *The Journal of Physical Chemistry B* **109**, 17715–17733 (2005). Cited on page 20.

118. Kuczera, K., Kuriyan, J. & Karplus, M. Temperature dependence of the structure and dynamics of myoglobin. a simulation approach. *Journal of molecular biology* **213**, 351–373 (1990). Cited on pp. 21, 24, 43, 44, 46, 49, 51, 71, 123, 131, 132, 133, 134, and 135.

119. Loncharich, R. J. & Brooks, B. R. Temperature dependence of dynamics of hydrated myoglobin. comparison of force field calculations with neutron scattering data. *Journal of molecular biology* **215**, 439–455 (1990). Cited on pp. 21, 24, 43, 44, 46, 71, 123, and 135.

120. Zoppellaro, G. *et al.* Review: Studies of ferric heme proteins with highly anisotropic/highly axial low spin (s= 1/2) electron paramagnetic resonance signals with bis-histidine and histidine-methionine axial iron coordination. *Biopolymers* **91**, 1064–1082 (2009). Cited on page 22.

121. Hariharan, P. C. & Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica Chimica Acta* **28**, 213–222 (1973). Cited on page 24.

122. Hehre, W. J., Ditchfield, R. & Pople, J. A. Self—consistent molecular orbital methods. XII. further extensions of gaussian—type basis sets for use in molecular orbital studies of organic molecules. *The Journal of Chemical Physics* **56**, 2257–2261 (1972). Cited on page 24.

123. Rassolov, V. A., Pople, J. A., Ratner, M. A. & Windus, T. L. 6-31g∗ basis set for atoms k through zn. *The Journal of Chemical Physics* **109**, 1223–1229 (1998). Cited on page 24.

124. Kim, K. & Friesner, R. A. Hydrogen bonding between amino acid backbone and side chain analogues: a high-level ab initio study. *Journal of the American Chemical Society* **119**, 12952–12961 (1997). Cited on page 25.

125. Huang, N. & MacKerell, A. D. An ab initio quantum mechanical study of hydrogen-bonded complexes of biological interest. *The Journal of Physical Chemistry A* **106**, 7820–7827 (2002). Cited on page 25.

126. MacKerell, A. D. & Karplus, M. Importance of attractive van der waals contribution in empirical energy function models for the heat of vaporization of polar liquids. *The Journal of Physical Chemistry* **95**, 10559–10560 (1991). Cited on page 25.

127. Mayne, C. G., Saam, J., Schulten, K., Tajkhorshid, E. & Gumbart, J. C. Rapid parameterization of small molecules using the force field toolkit. *Journal of Computational Chemistry* **34**, 2757–2770 (2013). Cited on page 25.

128. Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry* **5**, 129–145 (1984). Cited on page 26.

129. Jorgensen, W. L. Optimized intermolecular potential functions for liquid alcohols. *The Journal of Physical Chemistry* **90**, 1276–1284 (1986). Cited on page 26.

130. Beglov, D. & Roux, B. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *The Journal of Chemical Physics* **100**, 9050–9063 (1994). Cited on pp. 28 and 29.

131. Pesce, A. *et al.* Human brain neuroglobin structure reveals a distinct mode of controlling oxygen affinity. *Structure* **11**, 1087–1095 (2003). Cited on pp. 28, 62, 63, 65, 66, 68, 71, and 139.

132. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859–1865 (2008). Cited on pp. 29, 30, and 31.

133. Lee, J. *et al.* CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Journal of Chemical Theory and Computation* **12**, 405–413 (2016). Cited on pp. 29, 30, and 31.

134. Jo, S., T. Kim, V. I., & Im, W. CHARMM-GUI solvator (2008). Cited on page 29.

135. Loncharich, R. J. & Brooks, B. R. The effects of truncating long-range forces on protein dynamics. *Proteins* **6**, 32–45 (1989). Cited on page 29.

136. Kalé, L. *et al.* NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics* **151**, 283–312 (1999). Cited on pp. 29, 30, and 37.

137. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781–1802 (2005). Cited on pp. 29, 30, and 37.

138. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The langevin piston method. *The Journal of Chemical Physics* **103**, 4613–4621 (1995). Cited on pp. 29 and 30.

139. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **101**, 4177–4189 (1994). Cited on pp. 29 and 30.

140. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327–341 (1977). Cited on pp. 29 and 31.

141. Grubmüller, H., Heller, H., Windemuth, A. & Schulten, K. Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation* **6**, 121–142 (1991). Cited on pp. 29 and 31.

142. Tuckerman, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics* **97**, 1990–2001 (1992). Cited on pp. 29 and 31.

143. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The Journal of Physical Chemistry B* **114**, 7830–7843 (2010). Cited on page 29.

144. Zhou, H., Tajkhorshid, E., Frauenheim, T., Suhai, S. & Elstner, M. Performance of the AM1, PM3, and SCC-DFTB methods in the study of conjugated schiff base molecules. *Chemical Physics* **277**, 91–103 (2002). Cited on page 29.

145. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858 (2015). Cited on page 30.

146. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372**, 774–797 (2007). Cited on page 30.

147. Gunaydin, L. A. *et al.* Ultrafast optogenetic control. *Nature Neuroscience* **13**, 387–392 (2010). Cited on pp. 30 and 32.

148. Watanabe, H. C. *et al.* Structural model of channelrhodopsin. *Journal of Biological Chemistry* **287**, 7456–7466 (2012). Cited on pp. 30, 33, 83, 94, 111, and 112.

149. Scholz, F., Bamberg, E., Bamann, C. & Wachtveitl, J. Tuning the primary reaction of channelrhodopsin-2 by imidazole, pH, and site-specific mutations. *Biophys. J.* **102**, 2649–2657 (2012). Cited on pp. 30, 32, and 93.

150. Lorenz-Fonfria, V. A. *et al.* Transient protonation changes in channelrhodopsin-2 and their relevance to channel gating. *Proceedings of the National Academy of Sciences* **110**, E1273–E1281 (2013). Cited on pp. 30, 31, 83, 87, 88, 94, and 112.

151. Kuhne, J. *et al.* Early formation of the ion-conducting pore in channelrhodopsin-2. *Angewandte Chemie International Edition* **54**, 4953–4957 (2014). Cited on pp. 30, 83, 86, 87, and 112.

152. Sneskov, K. *et al.* Computational screening of one- and two-photon spectrally tuned channelrhodopsin mutants. *Phys. Chem. Chem. Phys.* **15**, 7567 (2013). Cited on page 30.

153. S. Jo, T. K., Iyer, V. & Im, W. Charmm-gui membrane builder (2008). Cited on page 30.

154. Jo, S., Kim, T. & Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLOS ONE* **2**, e880 (2007). Cited on page 30.

155. Wu, E. L. *et al.* CHARMM-GUIMembrane buildertoward realistic biological membrane simulations. *Journal of Computational Chemistry* **35**, 1997–2004 (2014). Cited on page 30.

156. Eisenhauer, K. *et al.* In channelrhodopsin-2 glu-90 is crucial for ion selectivity and is deprotonated during the photocycle. *Journal of Biological Chemistry* **287**, 6904–6911 (2012). Cited on pp. 31, 86, 87, and 112.

157. Takemoto, M. *et al.* Molecular dynamics of channelrhodopsin at the early stages of channel opening. *PLOS ONE* **10**, e0131094 (2015). Cited on page 31.

158. Wietek, J. *et al.* Conversion of channelrhodopsin into a light-gated chloride channel. *Science* **344**, 409–412 (2014). Cited on pp. 31, 83, 88, and 111.

159. Luecke, H., Schobert, B., Richter, H.-T., Cartailler, J.-P. & Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 Å resolution. *Journal of Molecular Biology* **291**, 899 – 911 (1999). Cited on pp. 32, 81, and 103.

160. Grossfield, A. Wham: the weighted histogram analysis method, version 2.0.9 (2013). URL `http://membrane.urmc.rochester.edu/content/wham`. Cited on page 35.

161. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry* **13**, 1011–1021 (1992). Cited on pp. 35 and 95.

162. Ferrenberg, A. M. & Swendsen, R. H. Optimized monte carlo data analysis. *Physical Review Letters* **63**, 1195–1198 (1989). Cited on page 35.

163. Torrie, G. M. & Valleau, J. P. Monte carlo free energy estimates using non-boltzmann sampling: Application to the sub-critical lennard-jones fluid. *Chemical Physics Letters* **28**, 578–581 (1974). Cited on page 35.

164. Mezei, M., Mehrotra, P. K. & Beveridge, D. L. Monte carlo determination of the free energy and internal energy of hydration for the ala dipeptide at 25.degree.c. *Journal of the American Chemical Society* **107**, 2239–2245 (1985). Cited on page 35.

165. Roux, B. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* **91**, 275–282 (1995). Cited on page 35.

166. Bondar, A.-N., Baudry, J., Suhai, S., Fischer, S. & Smith, J. C. Key role of active-site water molecules in bacteriorhodopsin proton-transfer reactions. *The Journal of Physical Chemistry B* **112**, 14729–14741 (2008). Cited on page 35.

167. Kirkwood, J. G. Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics* **3**, 300–313 (1935). Cited on page 35.

168. Brooks, B. R. RESDistance module in CHARMM (1995). Cited on page 35.

169. Zhang, L. & Hermans, J. Hydrophilicity of cavities in proteins. *Proteins: Structure, Function, and Genetics* **24**, 433–438 (1996). Cited on page 36.

170. Gumbart, J., Trabuco, L. G., Schreiner, E., Villa, E. & Schulten, K. Regulation of the protein-conducting channel by a bound ribosome. *Structure* **17**, 1453–1464 (2009). Cited on page 36.

171. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996). Cited on pp. 36 and 37.

172. Frisch, M. J. *et al.* Gaussian 09, revision d.01 (2013). Cited on page 37.

173. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **30**, 1545–1614 (2009). Cited on page 37.

174. Brooks, B. R. *et al.* CHARMM version 41b2 (2015). Cited on page 37.

175. Phillips, J. C. *et al.* NAMD version 2.9 (2014). Cited on page 37.

176. Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8 (2015). Cited on page 37.

177. Humphrey, W., Dalke, A. & Schulten, K. VMD version 1.9.3 (2016). Cited on page 37.

178. Stone, J. *An Efficient Library for Parallel Ray Tracing and Animation.* Master's thesis, Computer Science Department, University of Missouri-Rolla (1998). Cited on page 37.

179. Adam, S., Knapp-Mohammady, M., Yi, J. & Bondar, A.-N. Revised CHARMM force field parameters for iron-containing cofactors of photosystem II. *Journal of Computational Chemistry* **39**, 7–20 (2018). Cited on pp. 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 57, 59, 60, 62, 63, 64, 65, 66, 67, 68, 73, 123, 124, and 125.

180. Petrich, J. W. *et al.* Ligand binding and protein relaxation in heme proteins: a room temperature analysis of nitric oxide geminate recombination. *Biochemistry* **30**, 3975–3987 (1991). Cited on pp. 46, 52, 56, 71, and 131.

181. Vanommeslaeghe, K., Yang, M. & MacKerell, A. D. Robustness in the fitting of molecular mechanics parameters. *Journal of Computational Chemistry* **36**, 1083–1101 (2015). Cited on page 58.

182. Holroyd, L. F. & van Mourik, T. Insufficient description of dispersion in B3LYP and large basis set superposition errors in MP2 calculations can hide peptide conformers. *Chemical Physics Letters* **442**, 42–46 (2007). Cited on page 59.

183. Yesselman, J. D., Price, D. J., Knight, J. L. & Brooks, C. L. MATCH: an atom-typing toolset for molecular mechanics force fields. *Journal of Computational Chemistry* **33**, 189–202 (2011). Cited on page 59.

184. Vanommeslaeghe, K. & MacKerell, A. D. Automation of the CHARMM general force field (CGenFF) i: bond perception and atom typing. *Journal of Chemical Information and Modeling* **52**, 3144–3154 (2012). Cited on pp. 59 and 118.

185. Vanommeslaeghe, K., Raman, E. P. & MacKerell, A. D. Automation of the CHARMM general force field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *Journal of Chemical Information and Modeling* **52**, 3155–3168 (2012). Cited on pp. 59 and 118.

186. McLean, A. D. & Chandler, G. S. Contracted gaussian basis sets for molecular calculations. i. second row atoms, z=11–18. *The Journal of Chemical Physics* **72**, 5639–5648 (1980). Cited on page 60.

187. Krishnan, R., Binkley, J. S., Seeger, R. & Pople, J. A. Self-consistent molecular orbital methods. XX. a basis set for correlated wave functions. *The Journal of Chemical Physics* **72**, 650–654 (1980). Cited on page 60.

188. Clark, T., Chandrasekhar, J., Spitznagel, G. W. & Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. iii. the 3-21+g basis set for first-row elements, li-f. *Journal of Computational Chemistry* **4**, 294–301 (1983). Cited on page 60.

189. Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **90**, 1007–1023 (1989). Cited on page 60.

190. Autenrieth, F., Tajkhorshid, E., Baudry, J. & Luthey-Schulten, Z. Classical force field parameters for the heme prosthetic group of cytochrome *c*. *Journal of Computational Chemistry* **25**, 1613–1622 (2004). Cited on pp. 60 and 64.

191. Gill, P. M. & Radom, L. Deceptive convergence in møller-plesset perturbation energies. *Chemical Physics Letters* **132**, 16–22 (1986). Cited on page 62.

192. Vallone, B., Nienhaus, K., Brunori, M. & Nienhaus, G. U. The structure of murine neuroglobin: novel pathways for ligand migration and binding. *Proteins: Structure, Function, and Bioinformatics* **56**, 85–92 (2004). Cited on pp. 66 and 71.

193. Arcovito, A. *et al.* An x-ray diffraction and x-ray absorption spectroscopy joint study of neuroglobin. *Archives of Biochemistry and Biophysics* **475**, 7–13 (2008). Cited on pp. 66 and 71.

194. Lechauve, C. *et al.* Neuroglobins, pivotal proteins associated with emerging neural systems and precursors of metazoan globin diversity. *Journal of Biological Chemistry* **288**, 6957–6967 (2013). Cited on pp. 66 and 71.

195. Guimarães, B. G., Hamdane, D., Lechauve, C., Marden, M. C. & Golinelli-Pimpaneau, B. The crystal structure of wild-type human brain neuroglobin reveals flexibility of the disulfide bond that regulates oxygen affinity. *Acta Crystallographica Section D Biological Crystallography* **70**, 1005–1014 (2014). Cited on pp. 66 and 71.

196. Schreiter, E. R., Rodríguez, M. M., Weichsel, A., Montfort, W. R. & Bonaventura, J. *S*-nitrosylation-induced conformational change in blackfin tuna myoglobin. *Journal of Biological Chemistry* **282**, 19773–19780 (2007). Cited on pp. 68 and 71.

197. shar Huang, L. *et al.* 3-nitropropionic acid is a suicide inhibitor of mitochondrial respiration that, upon oxidation by complex II, forms a covalent adduct with a catalytic base arginine in the active site of the enzyme. *Journal of Biological Chemistry* **281**, 5965–5972 (2005). Cited on page 71.

198. shar Huang, L., Cobessi, D., Tung, E. Y. & Berry, E. A. Binding of the respiratory chain inhibitor antimycin to the mitochondrial bc1 complex: A new crystal structure reveals an altered intramolecular hydrogen-bonding pattern. *Journal of Molecular Biology* **351**, 573–597 (2005). Cited on page 71.

199. Leu, B. M. *et al.* Resilience of the iron environment in heme proteins. *Biophysical Journal* **95**, 5874–5889 (2008). Cited on page 72.

200. Berthomieu, C. & Hienerwadel, R. Iron coordination in photosystem II: interaction between bicarbonate and the q_b pocket studied by fourier transform infrared spectroscopy. *Biochemistry* **40**, 4044–4052 (2001). Cited on page 72.

201. Koepke, J. *et al.* pH modulates the quinone position in the photosynthetic reaction center from *Rhodobacter sphaeroides* in the neutral and charge separated states. *Journal of Molecular Biology* **371**, 396–409 (2007). Cited on page 73.

202. Yeh, A. P., Hu, Y., Jenney, F. E., Adams, M. W. W. & Rees, D. C. Structures of the superoxide reductase from *Pyrococcus furiosus* in the oxidized and reduced states. *Biochemistry* **39**, 2499–2508 (2000). Cited on page 73.

203. Hewitson, K. S. *et al.* Evidence that two enzyme-derived histidine ligands are sufficient for iron binding and catalysis by factor inhibiting HIF (FIH). *Journal of Biological Chemistry* **283**, 25971–25978 (2008). Cited on page 73.

204. Day, C. L., Anderson, B. F., Tweedie, J. W. & Baker, E. N. Structure of the recombinant N-Terminal lobe of human lactoferrin at 2·0 Å resolution. *Journal of Molecular Biology* **232**, 1084–1100 (1993). Cited on page 73.

205. Noinaj, N. *et al.* Structural basis for iron piracy by pathogenic *Neisseria*. *Nature* **483**, 53–58 (2012). Cited on page 73.

206. VanGordon, M. R., Gyawali, G., Rick, S. W. & Rempe, S. B. Atomistic study of intramolecular interactions in the closed-state channelrhodopsin chimera, c1c2. *Biophysical Journal* **112**, 943–952 (2017). Cited on pp. 80, 86, 87, 89, 93, 94, and 112.

207. Richards, R. & Dempski, R. E. Adjacent channelrhodopsin-2 residues within transmembranes 2 and 7 regulate cation selectivity and distribution of the two open states. *Journal of Biological Chemistry* **292**, 7314–7326 (2017). Cited on pp. 80, 87, 89, and 112.

208. Edholm, O., Berger, O. & Jähnig, F. Structure and fluctuations of bacteriorhodopsin in the purple membrane: A molecular dynamics study. *Journal of Molecular Biology* **250**, 94–111 (1995). Cited on page 80.

209. Bernèche, S. & Roux, B. Molecular dynamics of the KcsA k^+ channel in a bilayer membrane. *Biophysical Journal* **78**, 2900–2917 (2000). Cited on page 80.

210. Anashkin, V. A. *et al.* Engineering a carotenoid-binding site in dokdonia sp. PRO95 na^+-translocating rhodopsin by a single amino acid substitution. *Photosynthesis Research* (2017). Cited on page 80.

211. Grossfield, A., Feller, S. E. & Pitman, M. C. Convergence of molecular dynamics simulations of membrane proteins. *Proteins: Structure, Function, and Bioinformatics* **67**, 31–40 (2007). Cited on page 80.

212. Matsui, Y. *et al.* Specific damage induced by x-ray radiation and structural changes in the primary photoreaction of bacteriorhodopsin. *Journal of Molecular Biology* **324**, 469–481 (2002). Cited on page 81.

213. Borshchevskiy, V. *et al.* Low-dose x-ray radiation induces structural alterations in proteins. *Acta Crystallographica Section D Biological Crystallography* **70**, 2675–2685 (2014). Cited on page 81.

214. Nikolaev, M. *et al.* Integral membrane proteins can be crystallized directly from nanodiscs. *Crystal Growth & Design* **17**, 945–948 (2017). Cited on page 81.

215. Hildebrandt, P. & Stockburger, M. Role of water in bacteriorhodopsin's chromophore: resonance raman study. *Biochemistry* **23**, 5539–5548 (1984). Cited on page 81.

216. Althaus, T., Eisfeld, W., Lohrmann, R. & Stockburger, M. Application of raman spectroscopy to retinal proteins. *Israel Journal of Chemistry* **35**, 227–251 (1995). Cited on page 81.

217. Kandori, H. Role of internal water molecules in bacteriorhodopsin. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1460**, 177–191 (2000). Cited on page 81.

218. Ito, S. *et al.* Water-containing hydrogen-bonding network in the active center of channelrhodopsin. *Journal of the American Chemical Society* **136**, 3475–3482 (2014). Cited on pp. 81, 89, 94, and 112.

219. Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from

published macromolecular structures. *FEBS Journal* **275**, 1–21 (2007). Cited on page 82.

220. Hontani, Y. *et al.* Reaction dynamics of the chimeric channelrhodopsin c1c2. *Scientific Reports* **7** (2017). Cited on pp. 83 and 111.

221. Watanabe, H. C., Welke, K., Sindhikara, D. J., Hegemann, P. & Elstner, M. Towards an understanding of channelrhodopsin function: Simulations lead to novel insights of the channel mechanism. *Journal of Molecular Biology* **425**, 1795–1814 (2013). Cited on pp. 83, 94, 111, and 112.

222. Guo, Y. *et al.* Active site structure and absorption spectrum of channelrhodopsin-2 wild-type and c128t mutant. *Chem. Sci.* **7**, 3879–3891 (2016). Cited on pp. 83, 94, 111, and 112.

223. Ritter, E., Stehfest, K., Berndt, A., Hegemann, P. & Bartl, F. J. Monitoring light-induced structural changes of channelrhodopsin-2 by UV-visible and fourier transform infrared spectroscopy. *Journal of Biological Chemistry* **283**, 35033–35041 (2008). Cited on page 87.

224. Sugiyama, Y. *et al.* Photocurrent attenuation by a single polar-to-nonpolar point mutation of channelrhodopsin-2. *Photochemical & Photobiological Sciences* **8**, 328 (2009). Cited on page 87.

225. Ruffert, K. *et al.* Glutamate residue 90 in the predicted transmembrane domain 2 is crucial for cation flux through channelrhodopsin 2. *Biochemical and Biophysical Research Communications* **410**, 737–743 (2011). Cited on page 87.

226. Gradmann, D., Berndt, A., Schneider, F. & Hegemann, P. Rectification of the channelrhodopsin early conductance. *Biophysical Journal* **101**, 1057–1068 (2011). Cited on pp. 87 and 112.

227. Nack, M., Radu, I., Bamann, C., Bamberg, E. & Heberle, J. The retinal structure of channelrhodopsin-2 assessed by resonance raman spectroscopy. *FEBS Letters* **583**, 3676–3680 (2009). Cited on pp. 94 and 112.

228. Harbison, G. S. *et al.* Dark-adapted bacteriorhodopsin contains 13-cis, 15-syn and all-trans, 15-anti retinal schiff bases. *Proceedings of the National Academy of Sciences* **81**, 1706–1709 (1984). Cited on page 95.

229. Zhu, F. & Hummer, G. Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *Journal of Computational Chemistry* **33**, 453–465 (2011). Cited on page 95.

230. Han, M. & Smith, S. O. NMR constraints on the location of the retinal chromophore in rhodopsin and bathorhodopsin. *Biochemistry* **34**, 1425–1432 (1995). Cited on page 103.

231. Furutani, Y., Shibata, M. & Kandori, H. Strongly hydrogen-bonded water molecules in the schiff base region of rhodopsins. *Photochemical & Photobiological Sciences* **4**, 661 (2005). Cited on page 103.

232. Buss, V., Sugihara, M., Entel, P. & Hafner, J. Thr94 and wat2b effect protonation of the retinal chromophore in rhodopsin. *Angewandte Chemie International Edition* **42**, 3245–3247 (2003). Cited on pp. 103 and 113.

233. Sugihara, M., Buss, V., Entel, P. & Hafner, J. The nature of the complex counterion of the chromophore in rhodopsin. *The Journal of Physical Chemistry B* **108**, 3673–3680 (2004). Cited on pp. 103 and 113.

234. Janz, J. M. & Farrens, D. L. Role of the retinal hydrogen bond network in rhodopsin schiff base stability and hydrolysis. *Journal of Biological Chemistry* **279**, 55886–55894 (2004). Cited on page 103.

235. Krause, N., Engelhard, C., Heberle, J., Schlesinger, R. & Bittl, R. Structural differences between the closed and open states of channelrhodopsin-2 as observed by EPR spectroscopy. *FEBS Letters* **587**, 3309–3313 (2013). Cited on page 109.

236. Sattig, T., Rickert, C., Bamberg, E., Steinhoff, H.-J. & Bamann, C. Light-induced movement of the transmembrane helix b in channelrhodopsin-2. *Angewandte Chemie International Edition* **52**, 9705–9708 (2013). Cited on page 109.

237. Müller, M., Bamann, C., Bamberg, E. & Kühlbrandt, W. Light-induced helix movements in channelrhodopsin-2. *Journal of Molecular Biology* **427**, 341–349 (2015). Cited on page 109.

238. Shaw, D. E. *et al.* Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM* **51**, 91 (2008). Cited on page 117.

239. Wolf, M. G. & Groenhof, G. Explicit proton transfer in classical molecular dynamics simulations. *Journal of Computational Chemistry* **35**, 657–671 (2014). Cited on page 118.

240. Lamoureux, G. & Roux, B. Modeling induced polarization with classical drude oscillators: Theory and molecular dynamics simulation algorithm. *The Journal of Chemical Physics* **119**, 3025–3039 (2003). Cited on page 118.

241. Lopes, P. E. M. *et al.* Polarizable force field for peptides and proteins based on the classical drude oscillator. *Journal of Chemical Theory and Computation* **9**, 5430–5449 (2013). Cited on page 118.

Selbstständigkeitserklärung

Hiermit bestätige ich, dass die vorliegende Doktorarbeit von mir allein und nur unter Nutzung der angegebenen Hilfsmittel verfasst wurde. Ferner bestätige ich, dass diese Arbeit nicht in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt wurde.

Berlin, den 7. Dezember 2017

Suliman Adam