

Computational Analysis of the Proton Transfer to the Secondary Quinone of Type II Photosynthetic Reaction Centers.

Dissertation zur Erlangung der Doktorwürde
der Fakultät für Biologie, Chemie und Geowissenschaften
der Universität Bayreuth

vorgelegt von

Eva-Maria Krammer

2008

Die vorliegende Arbeit wurde im Zeitraum von Juni 2005 bis November 2008 an der Universität Bayreuth unter der Leitung von Prof. Dr. Matthias Ullmann angefertigt.

Vollständiger Abdruck der von der Fakultät fuer Biologie, Chemie und Geowissenschaften der Universität Bayreuth genehmigten Dissertation zur Erlangung des akademischen Grades des Doktors der Naturwissenschaften (Dr. rer. nat.)

Datum der Einreichung der Arbeit: 05. 12. 2008

Datum des wissenschaftlichen Kolloquiums: 27. 04. 2009

Prüfungsausschuss: Prof. Dr. Matthias Ullmann (Erstgutachter)

Prof. Dr. Holger Dobbeck (Zweitgutachter)

Prof. Dr. Thomas Hellweg (Vorsitzender)

Prof. Dr. Stephan Clemens

Mein herzlicher Dank geht an:

Prof. Dr. Matthias Ullmann für die die exzellente fachliche Unterstützung, für die vielen interessanten und ergebnisreichen Diskussionen, sowie für die ausserordentlich guten Arbeitsbedingungen in seiner Gruppe.

Prof. Dr. Pierre Sebban (Université Paris-Sud, Frankreich) für die ergebnis- und aufschlussreiche Zusammenarbeit und die spannenden und lehrreichen Forschungsaufenthalte in seiner Gruppe. Merci beaucoup!

Dr. Günther Fritzsche und Dr. Jürgen Köpke (Max-Planck-Institut für Biophysik in Frankfurt am Main) für die gute Zusammenarbeit und für die Bereitstellung von benötigten Röntgenkristallstrukturen noch vor deren Veröffentlichung.

Timm Essigke für die hervorragende und ergebnisreiche Zusammenarbeit, für stetiges Interesse an immer neuen Ergebnistabellen, für das Korrekturlesen dieser Arbeit, und für stets schnelle und kompetente Hilfe nicht nur bei Computerproblemen.

Mirco Till für die interessante und produktive Zusammenarbeit sowie für die Umsetzung von Ideen in Programmen.

Der Arbeitsgruppe Strukturbiologie/Bioinformatik insbesondere Astrid, Torsten, Timm und Silke für die nette und freundschaftliche Atmosphäre in und ausserhalb der B14.

Christian Dahmann, weil er mir immer zur Seite steht.

Meinen Eltern, deren vielfältigen und ganz unterschiedlichen Beiträge zum Gelingen dieser Arbeit ich nicht in einen Satz fassen vermag.

Meinen Freundinnen Hanna Berkner, Anna Foik und Nina Link. Ohne Euch wären die letzten Jahre nicht das Gleiche gewesen.

Allen meiner Familie, meiner Verwandten, meiner Freunde und meiner Arbeitskollegen, die hier aus Platzgründen nicht zu Genüge erwähnt wurden und doch einen erheblichen Teil meines Lebens ausmachen. Danke!

Contents

Summary	7
Zusammenfassung	8
Type II Photosynthetic Reaction Centers	10
1 Photosynthesis and the Photosynthetic Reaction Center	10
1.1 Components of the Photosynthetic Apparatus	10
1.2 Evolution of Photosynthetic Reaction Centers	13
1.3 Structural Organization and Function of Type II Reaction Centers	13
2 Mechanism of the Bacterial Reaction Center	17
2.1 The Coenzyme Q Binding Sites	17
2.2 Proposed Catalytic Cycle	21
2.3 Proton Transfer to Q_B	24
3 Aim of the Thesis	29
Synopsis of the Manuscripts	30
Bibliography	35
List of Abbreviations	49
Manuscripts	51
List of Manuscripts	51
Manuscript A	53
Manuscript B	55
Manuscript C	57
Manuscript D	85
Manuscript E	87
Manuscript F	89
Manuscript G	91
Appendix	93
<i>Profile Hidden Markov Models</i>	95
1 Markov Chains and Hidden Markov Models	96
1.1 Markov Chains	96
1.2 Components of Hidden Markov Models	98

1.3 Construction of a Hidden Markov Model.	100
2 A Profile Hidden Markov Model	100
2.1 Definition of the profile Hidden Markov Model.	101
2.2 Organization of a profile Hidden Markov Model	102
2.3 Construction of a profile Hidden Markov Model	104
2.4 Calculation of Emission Probabilities: Scoring Matrix and Gap Penalties . .	107
2.5 Multiple Sequence Alignment with a profile Hidden Markov Model.	109
2.6 Limitations and Advantages of the profile Hidden Markov Model Approach.	112
3. Bibliography	114
4 List of Abbreviations	116

Summary

For the life of a huge variety of different species molecular oxygen is needed. Photosynthesis is the main process on earth that produces molecular oxygen. A crucial step in photosynthesis is catalyzed by the Type II photosynthetic reaction center (RC): the conversion of chemical energy into an electrochemical gradient by reducing and protonating a Coenzyme Q bound in the Q_B binding site of the protein. The pigments of Type II RC, namely of the plant Photosystem II RC (PSII RC) and of its evolutionary ancestor, the bacterial RC (bRC), are arranged in two (pseudo-)symmetrical branches, the A- and the B-branch. In Type II RC proteins, the electrons are transferred to Q_B via the A-branch while the B-branch is electron-transfer inactive. In the thesis presented here the degree of conservation was analyzed for residues that tune the redox properties of the pigments and direct the electron transfer along the A-branch. The quality of such a conservation analysis depends critically on a correct multiple sequence alignment. Since the bRC and PSII RC share only very little sequence identity, profile Hidden Markov Models including structural information of the bRC and PSII RC were used to ensure a correct alignment. The conservation analysis showed that the tuning of the pigment redox properties and direction of electron transfer are conserved in bRC proteins but differ in PSII RC proteins. Correspondingly the character of the two Coenzyme Q (Q_A and Q_B) binding sites differs between PSII RC and bRC making it possible, that in PSII RC proteins Q_A can be protonated under stress conditions (such as high light) whereas such a protonation is not possible in the bRC.

Interestingly, two alternative binding positions (proximal and distal) have been observed for Q_B in the bRC of *Rhodobacter (Rb.) sphaeroides*. Experiments indicated, that Q_B changes its orientation by 180° during movement from the distal to the proximal position. Together with crystallographic experiments, my quantum chemical and continuum electrostatic calculations showed that Q_B is likely to have the same orientation in both binding positions. A coupling between the protonation of the ultimate proton donor groups (GluL212 and AspL213 of the L subunit) and the population of the two Q_B positions was identified explaining the observed pH- and illumination state dependence of the Q_B population. Moreover the protonation of these residues is needed to keep the first reaction intermediate, the potentially cell damaging semiquinone state $Q^{\cdot-}$, bound to the protein.

In contrast to the electron transfer via the A-branch, the mechanism of proton transfer to Q_B differs significantly between PSII RC and bRC. For the bRC of *Rb. sphaeroides* key residues of the proton transfer to Q_B were experimentally determined and proton entry points have been proposed. However, the exact organization of proton transfer to Q_B is still not known. Two alternative ideas are debated: either the protons are transferred via distinct proton transfer pathways or via a huge network without distinct pathways, a proton sponge. The analysis of a

multiple sequence alignment for the bRC subunits showed, that while the non-surface key residues of the proton transfer to Q_B are conserved, the proposed proton entry points are not conserved to the same extent. In addition, the hydrogen bonded network analysis revealed a huge network spanning from the cytoplasm to Q_B in the bRC of *Rb. sphaeroides* and *Blastochloris viridis*. Interestingly, these networks show a similar organization and both include all important non-surface key residues, but the networks differ in respect of the determined proton entry points. Both, the analysis of the conservation study and of the hydrogen bonded network, counter the idea of distinct proton transfer pathways and heavily support the idea of a proton sponge. By the combination of different approaches, such as conservation analysis based on multiple sequence alignments, continuum electrostatics, quantum mechanics and hydrogen bond network analysis, the work presented here succeeded in gaining further insights into the molecular details of the Q_B binding site and the proton and electron transfer reactions to Q_B .

Zusammenfassung

Für das Leben einer Vielzahl unterschiedlicher Arten wird molekularer Sauerstoff benötigt. Auf der Erde ist der wichtigste Prozess zur Herstellung von molekularem Sauerstoff die Photosynthese. Ein entscheidender Schritt der Photosynthese wird durch den Typ II des photosynthetischen Reaktionszentrums (RC) katalysiert: Die Umwandlung von chemischer Energie in einen elektrochemischen Gradienten durch die Reduzierung und Protonierung eines Coenzym Q Moleküls, dass in der Q_B Bindungstasche des Proteins gebunden ist. Die Pigmente des Typ II RC, nämlich des pflanzlichen Photosystem II RC (PSII RC) und dessen evolutionären Vorfahren, dem bakteriellen RC (bRC), sind in zwei (pseudo)-symmetrischen Zweigen angeordnet, dem A- und dem B-Zweig. In Typ II RC Proteinen werden die Elektronen entlang des A-Zweiges auf Q_B übertragen, während der B-Zweig keine Elektronen übertragen kann. In dieser Arbeit wurde der Konservierungsgrad von Resten untersucht, für die eine Beeinflussung der Redoxeigenschaften der Pigmente und der Lenkung des Elektronentransfers entlang des A-Zweiges bekannt ist. Die Qualität einer Konservierungsanalyse hängt massgeblich von einem korrekten multiplen Sequenzalignment ab. Da bRC und PSII RC nur eine sehr kleine Sequenzidentität haben, wurden profile Hidden Markov Modelle verwendet, welche die strukturellen Informationen der Proteine berücksichtigen, um ein korrektes Alignment zu erhalten. Die Konservierungsanalyse zeigte, dass die Abstimmung von Redoxeigenschaften der Pigmente und die Lenkung des Elektronentransfers im bRC konserviert sind, aber in PSII RC abweichen. Zwischen bRC und PSII RC Proteinen gibt es dementsprechend auch Unterschiede in den Eigenschaften der beiden Coenzym Q (Q_A und Q_B) Bindungstaschen, die es ermöglichen, dass im PSII RC Q_A unter Stressbedingungen (wie hoher Lichtintensität)

protonieren kann, während eine solche Protonierung im bRC nicht möglich ist.

Interessanterweise wurden im bRC von *Rhodobacter (Rb.) sphaeroides* zwei alternative Bindungspositionen für Q_B (proximal und distal) festgestellt. Experimente deuteten an, dass Q_B seine Orientierung um 180° ändert, während es sich von der distalen in die proximale Position bewegt. Zusammen mit kristallographischen Experimenten zeigten meine quantenchemischen und elektrostatischen Berechnungen, dass im bRC von *Rb. sphaeroides* Q_B wahrscheinlich die gleiche Orientierung in beiden Positionen einnimmt. Eine Kopplung des Protonierungszustands der terminalen Protonendonoren (GluL212 und AspL213 der L Untereinheit) und der Population der beiden Q_B Bindungspositionen erklärt die beobachtete pH- und Zustandsabhängigkeit der Q_B Population. Darüber hinaus müssen diese Reste protoniert sein, um das erste Reaktionszwischenprodukt, das zellschädigende Semichinon $Q^{\cdot-}$, gebunden zu halten.

Im Unterschied zum Elektronentransfer entlang des A-Zweiges unterscheidet sich der Mechanismus des Protonentransfers zu Q_B massgeblich zwischen PSII RC und bRC. Im bRC von *Rb. sphaeroides* wurden die wesentlichen Reste des Protonentransfers zu Q_B experimentell bestimmt und Protoneneintrittspunkte wurden vorgeschlagen. Die genaue Organisation des Protonentransfers zu Q_B ist allerdings nicht bekannt. Zwei sich ausschliessende Ideen werden diskutiert: Die Protonen werden entweder über unterschiedliche Pfade oder über ein grosses Netzwerk ohne klar definierte Pfade, einem Protonenschwamm, transportiert. Die Auswertung eines multiplen Sequenzalignments der bRC Untereinheiten zeigte, dass die wesentlichen, nicht auf der Proteinoberfläche liegenden Reste des Protonentransfers konserviert sind. Die vorgeschlagenen Protoneneintrittspunkte sind aber nicht im gleichen Ausmass konserviert. Zusätzlich zeigte die Auswertung des Wasserstoffbrückennetzwerks der bRC Proteine von *Rb. sphaeroides* und *Blastochloris viridis* jeweils ein grosses Netzwerk, dass vom Cytoplasma bis zur Q_B Bindungstasche reicht. Interessanterweise haben diese Netzwerke einen ähnlichen Aufbau und beinhalten alle wesentlichen nicht auf der Oberfläche liegenden Reste des Protonentransfers, unterscheiden sich aber in den ermittelten Protoneneintrittspunkten. Sowohl die Konservierungsstudie und als auch die Analyse des Netzwerkes widersprechen der Idee von unabhängigen Protonentransferpfaden und unterstützen die Idee des Protonenschwamms. Durch die Kombination unterschiedlicher Ansätze wie der Konservierungsanalyse basierend auf multiplen Sequenzalignments, der Kontinuumselktrostatik, der Quantenchemie und der Analyse der Wasserstoffbrückennetzwerke, gelang es in dieser Arbeit ein breiteres Wissen über die molekularen Details der Q_B Bindungstasche und des Elektronen- und Protonentransfer zu Q_B zu gewinnen.

Type II Photosynthetic Reaction Centers

1 Photosynthesis and the Photosynthetic Reaction Center

For the evolution of life on earth, the oxygen-producing photosynthesis is of central importance. Photosynthesis takes place in single-cell organisms such as purple bacteria as well as in chloroplasts of highly-organized multi-cell organisms like plants. The photosynthetic apparatus is positioned in specialized membranes such as invaginations of the cytoplasmic membrane of bacteria or the thylakoid membrane of chloroplasts in plants. In the photosynthetic process light is utilized as the external energy source and converted into the chemical energy of adenosinetriphosphate (ATP) inside the cell or cell component. ATP can then be used by other enzymes for the production of organic biomass. In several species such as in green algae or in plants, oxygen is produced during the photosynthetic process, whereas for example in purple bacteria no oxygen is produced. Thus, one discriminates between oxygenic and anoxygenic photosynthesis. Independent of the species and whether the process is anoxygenic or oxygenic, photosynthesis is always coupled to the existence of a certain cofactor molecule: (bacterio-)chlorophyll (Bcl/Chl)¹⁻³.

1.1 Components of the Photosynthetic Apparatus

As an example of an anoxygenic photosynthetic apparatus the one of purple bacteria is schematically depicted in Figure 1 and described in the following. The energy of sun light is absorbed by specialized light-harvesting antenna proteins (LH1 and LH2) and transferred as excitation energy to the photosynthetic reaction center (RC). In the RC, the excitation energy leads to charge separation at the level of the special pair (formed by two Bcl molecules and named P870 in Figure 1 due to its absorption maximum). Charge separation is followed by electron transfer to a terminal electron acceptor species. By the following enzymes (the membrane spanning cytochrome *bc*₁ complex and the soluble electron carrier cytochrome *c*₂), the special pair is rereduced, the terminal electron acceptor is recycled and protons are pumped out of the cell leading to a proton gradient through the cell membrane. The created proton gradient is then used to produce ATP out of adenosinediphosphate and inorganic phosphate⁴. The here described photosynthetic apparatus of purple bacteria is rather simple compared to the oxygenic photosynthetic apparatus of green algae or higher plants. However, the photosynthetic apparatus of purple bacteria is not organized in a linear chain in the cytoplasmic membrane as it is depicted in Figure 1 but in supercomplexes⁵⁻⁷. Joliot *et al.* (2005)⁶ proposed, that a dimer of the *bc*₁ complex is surrounded by four RC/LH1 complexes. The RC/LH1 complexes are themselves surrounded by a layer of LH2 proteins. The LH1 protein consists of two alternating subunits and forms a ring-like structure around a central RC protein. In some purple bacteria such as *Rb. sphaeroides* the ring is

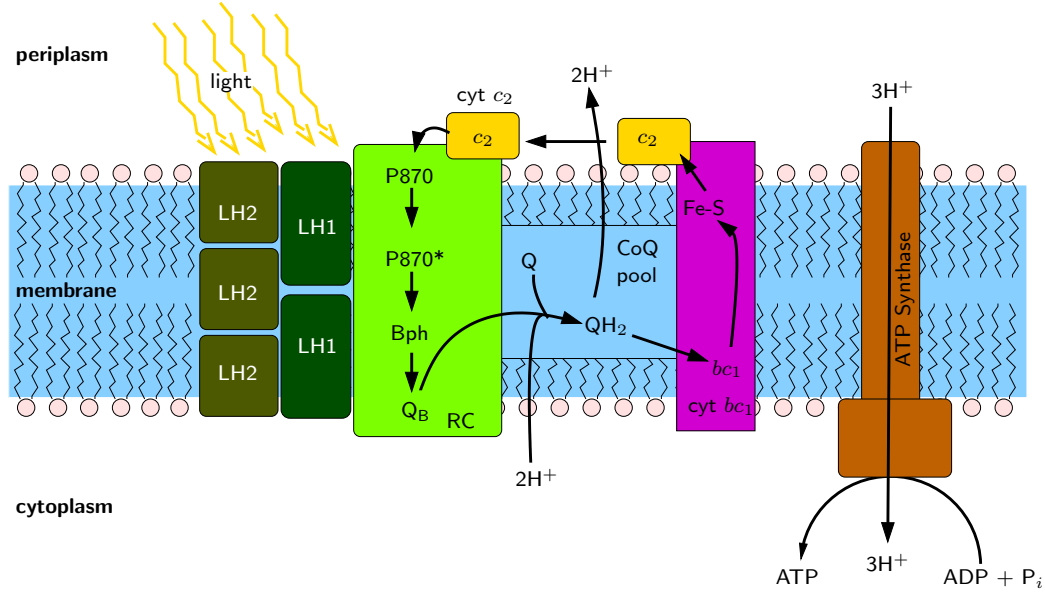


Figure 1: Schematic representation of the bacterial photosynthetic apparatus. Light energy is converted into a proton gradient by the action of LH1 and LH2 (colored in olive and dark-green) and the reaction center RC (colored light-green). Cytochrome bc_1 (cyt bc_1 ; colored purple) and cytochrome c_2 (cyt c_2 ; colored yellow) restore the function of the RC by oxidizing and deprotonating the terminal electron acceptor of the RC (Q_B) and by reducing the primary electron acceptor group of the RC (P870). The generated proton gradient is used to produce ATP out of ADP and P_i by an ATP-synthase (colored brown).

opened meaning that the subunits of the LH1 do not form a complete ring-like structure, but that several subunits are missing to close the ring. For other purple bacteria such as *Blastochloris* (*B.*) *viridis* a closed ring-like structure around the central RC is observed for the LH1 complex.^{8–10}. Apparently, the number of LH1 and LH2 complexes per RC and their organization differs from species to species^{11–14}. But the exact organization of the described supercomplexes is under investigation^{8,14}.

The oxygenic photosynthetic apparatus of green algae is depicted in Figure 2. As in purple bacteria (see Figure 1), light is collected by two light harvesting complexes (LHC). In contrast to the bacterial system, the LHCs are integrated as subunits in large protein complexes, the two photosystems, namely Photosystem I (PSI) and Photosystem II (PSII). These protein complexes consist of more than ten subunits and perform the central reactions^{15–17}. By the LHC subunits light is transferred as excitation energy to the special pair of the corresponding RC subunits. In the RC of the PSI and PSII absorption of the excitation energy leads to charge separation of the special pair (named P700 and P680 in Figure 2 for the PSI RC and PSII RC, respectively), followed by electron transfer to a terminal electron acceptor. PSI RC and PSII RC differ in the structure of their terminal electron acceptor: in PSI RC an iron-sulfur cluster is the electron

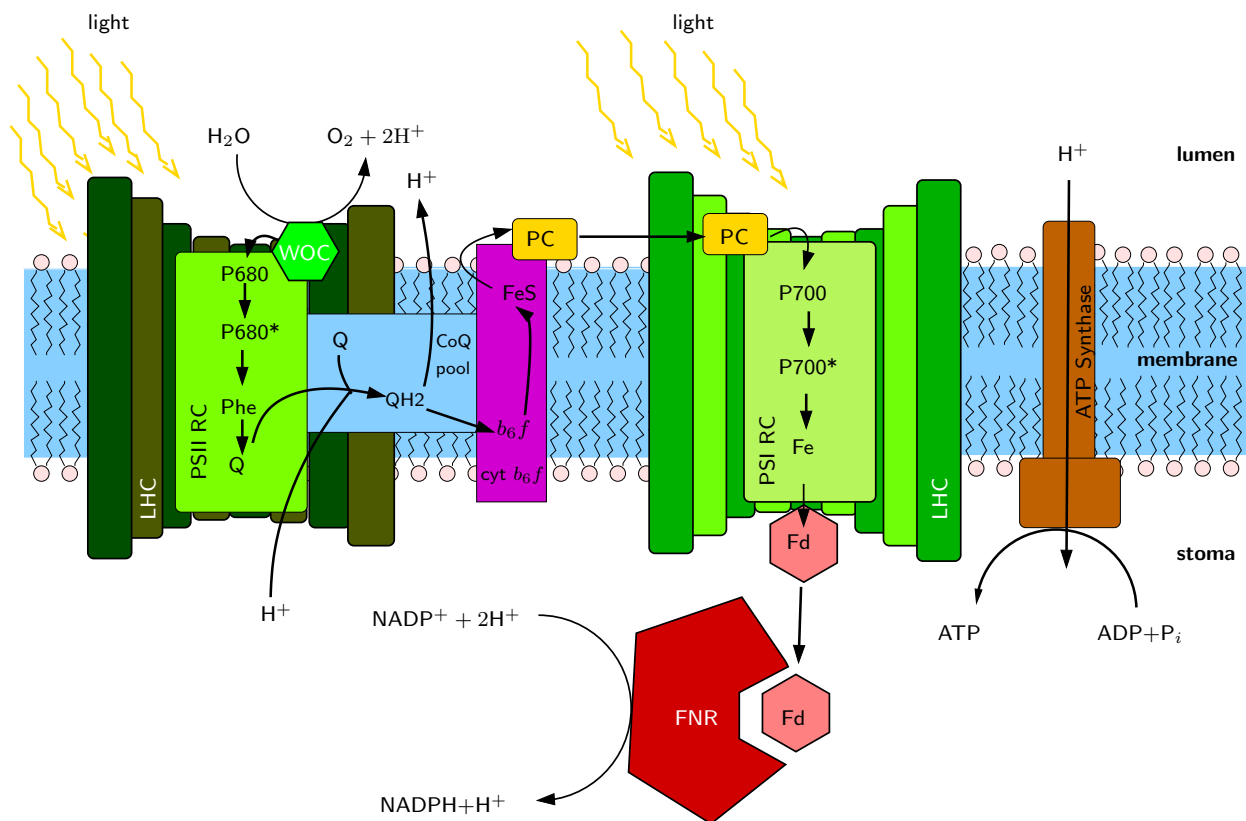


Figure 2: The photosynthetic apparatus of plants. Light is converted into a proton gradient by the action of two photosystems, PSI (colored light-olive) and PSII (colored light green). The functionality of both complexes is coupled by the cytochrome b_6f complex (cyt b_6f ; colored purple) and the soluble electron carrier Plastocyanin (PC; colored yellow). In addition to a created proton gradient, oxygen and $NADPH^+ + H^+$ are produced. Oxygen is formed in the WOC (green) of the PSII. $NADPH^+ + H^+$ is formed by the action of the electron carrier protein Ferredoxin (Fd; colored rose) and the FNR complex (colored red). The ATP synthase uses the created proton gradient to produce ATP.

acceptor (Fe in Fig 2) whereas it is a Coenzyme Q (CoQ) molecule for PSII RC (Q in Fig 2). In the PSII RC, the special pair is rereduced under the production of oxygen by the water oxidizing complex (WOC) bound to the PSII RC^{18,19}. The WOC is also named oxygen evolving complex or water splitting complex in literature. The terminal electron acceptor of PSII RC is recycled by the cytochrome b_6f complex, which is evolutionary linked to the cytochrome bc_1 complex in purple bacteria (see Figure 1)²⁰. During this recycling process, protons are pumped out of the cell and electrons are transferred stepwise by the soluble protein plastocyanin from the b_6f complex to PSI. In the PSI RC, the electrons are used to rereduce the special pair. As in PSII RC, the light-induced charge separation of the special pair is followed by electron transfer to the terminal electron acceptor in PSI RC. The electrons are abstracted stepwise from the terminal electron acceptor by the soluble protein ferredoxin. Two electron transfer steps by ferredoxin from PSI RC

to the ferredoxin-NADP oxidoreductase (FNR) are needed to recycle $\text{NADPH} + \text{H}^+$ from NADP^+ and two protons. This process also contributes to proton gradient set-up photosynthetically. The proton gradient is then used to produce ATP (like in purple bacteria; see Figure 1). Both ATP and $\text{NADPH} + \text{H}^+$ are used for the conversion of carbondioxide into biomass in the dark reaction, the Calvin cycle¹⁵. The described linear photosynthetic electron flow in the oxygenic photosynthetic apparatus can only be achieved, if the same amount of light is absorbed by the LHCs of PSI and PSII. Often the same light absorption at PSI and PSII is not possible. In the oxygenic photosynthetic apparatus several mechanisms exist to set up a proton gradient even without such light conditions like cyclic photosynthetic electron transport or special soluble light harvesting antenna, which can transport excitation energy from one complex to the other¹⁵.

1.2 Evolution of Photosynthetic Reaction Centers

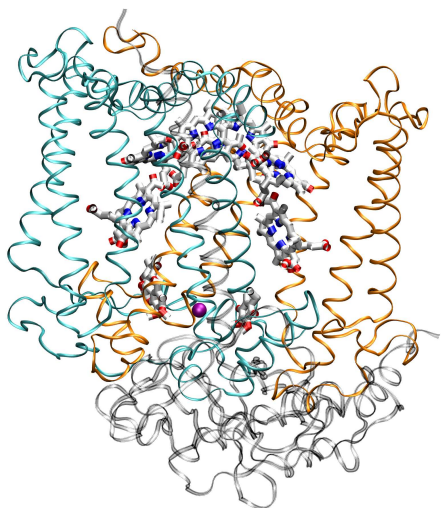
All RC proteins are evolutionary related and share a common design^{21–26}. However, since they differ in the terminal electron acceptor, the RC proteins can be grouped into two classes: Type I and Type II RC proteins. Type I RC proteins use an iron-sulfur cluster, whereas Type II RC proteins use a CoQ molecule as terminal electron acceptor. Both groups contain RC proteins of bacteria as well as of eukaryotes². The previously described anoxygenic photosynthetic apparatus of purple bacteria (see Figure 1) contains a Type II bacterial RC (bRC). The oxygenic photosynthetic apparatus of green algae (see Figure 2) contains a Type I RC protein in PSI and a Type II RC protein in PSII (PSII RC). It is generally accepted, that the Type I and Type II RC proteins share a common evolutionary ancestor^{23–26}. But until now, the exact evolutionary tree of RCs is still under debate²⁴. The selective loss model assumes, that the Type I bRC is the ancestral RC for all subsequent forms. During evolution the loss of the iron-sulfur center, gene splitting, gene duplication and divergence resulted in the other now existing RC proteins^{23,24}. In contrary, it is assumed in the fusion model, that the Type I and Type II RC proteins evolved independently in different bacterial species and were brought together in one species by gene fusion^{24–26}.

1.3 Structural Organization and Function of Type II Reaction Centers

In the following the components and structural features of Type II RC proteins will be described in more detail, since only Type II RC proteins were used in this study. However, apart from the terminal electron acceptor, similar cofactor organization and structural composition is also found in Type I RC proteins^{21,27}.

The bRC consists of up to four subunits, the L, M, H and C subunit²⁸. It is known, that in species like the purple bacterium *Rb. sphaeroides* with a bRC containing an H subunit, several residues of the H subunit are of functional importance for proton transfer to Q_B . In species with

a) Structure



b) Cofactor organization

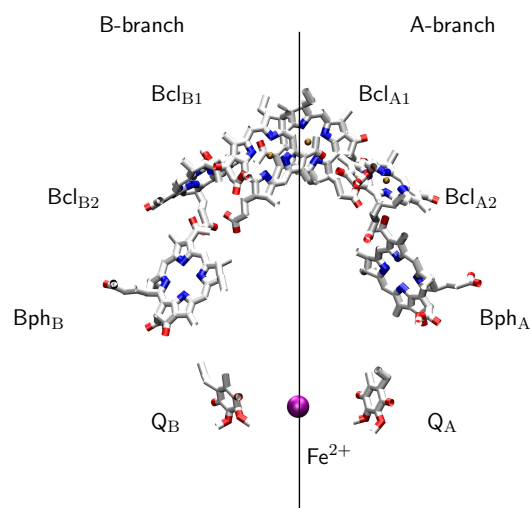
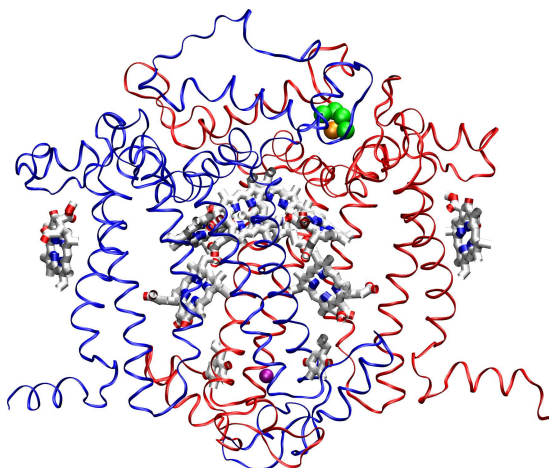


Figure 3: Structure and cofactor organization of the bacterial reaction center. (a) The structure of the bRC from *Rb. sphaeroides* is shown. The subunits L (orange), M (cyan), and H (grey) and the cofactors are depicted. (b) The cofactor molecule organization with the pseudo-symmetry axis of the bRC from *Rb. sphaeroides* is shown. The figure is made using vmd³³ and based on the crystal structure 2C8J³⁴.

bRC containing a C subunit like *B. viridis*, the electrons from cytochrome c_2 are accepted by one of the heme cofactor molecules bound by the C subunit and are transferred through the Bcl cofactors to the special pair.^{29,30} The L and M subunits form the core of the protein. In Figure 3 the bRC of *Rb. sphaeroides* and its cofactor organization is depicted. The PSII RC is formed by the PSII subunits D1 and D2, and forms together with more than twenty subunits the PSII complex³¹. The other subunits are involved in regulation, light harvesting, proton transfer to CoQ and self-protection against too much light^{19,31,32}. The structure of the PSII RC and the organization of the cofactors bound by these subunits is depicted in Figure 4. The subunits L and M of the bRC and the subunits D1 and D2 of the PSII RC are superimposable and show the same fold of five membrane-spanning helices³⁵. In both proteins, the two subunits are symmetrically arranged (see Figure 3a and Figure 4a). However, the L, M, D1, and D2 subunits show only about 5 % sequence identity³⁶, which is a value often found in comparison of random protein sequences³⁷. Thus, based on the sequence identity the observed structural similarity and the evolutionary relationship would not have been suggested.

By the L and M, and the D1 and D2 subunits, respectively, the redox cofactors are bound, which are needed for the light-induced reduction reaction. The cofactors are four Bcl/Chl molecules, two (bacterio-) pheophytin (Bph/Phe) molecules, two CoQ molecules and a non-heme

a) Structure



b) Cofactor organization

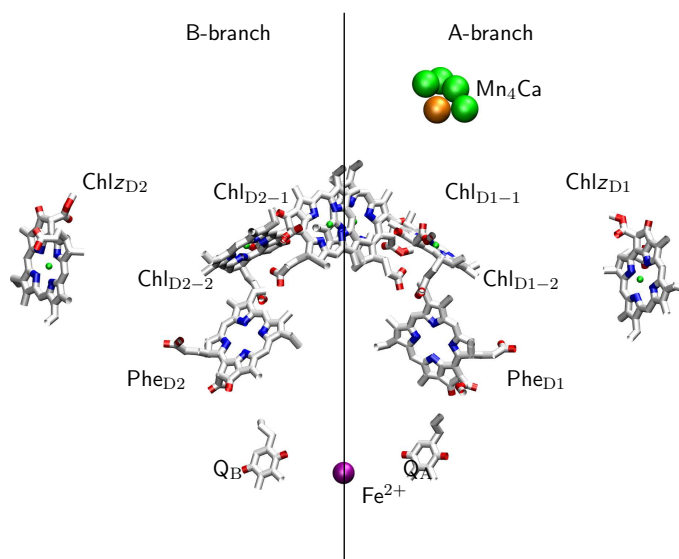


Figure 4: Structure and cofactor organization of the Photosystem II reaction center. (a) The PSII RC structure of *T. elongatus* consisting of the D1 (red) and D2 (blue) subunits and cofactors is shown. (b) Cofactor organization of the PSII RC of *T. elongatus*. Also the WOC (Mn_4Ca ; Mn and Ca ions are shown in green and orange, respectively) and the peripheral accessory chlorophylls (Chl_{D1} , Chl_{D2}) are depicted. The figure is made with vmd³³ based on the crystal structure 2AXT³¹.

iron atom³⁵. The cofactors are arranged in two membrane-spanning branches A and B as shown in Figure 3b for the bRC and in Figure 4b for the PSII RC. Each branch consists of two Bcl/Chl molecules, a Bph/Phe molecule and a CoQ molecule. In the bRC protein, the transfer of excitation energy leads to charge separation of the special pair, which is formed by the two first Bcl molecules of the A and B branch (by Bcl_{A1} and Bcl_{B1} ; see Figure 3b). The A and B branch show the same symmetric cofactor composition, but the electron transfer through the protein is asymmetric. Only the A-branch is electron transfer active in *wild-type* RC of *Rb. sphaeroides*. Initiated by the transfer of excitation (or light) energy, the electron is transferred from the special pair via the named cofactors of the A-branch to the first CoQ bound in the Q_A site and from there finally to the second CoQ molecule bound in the Q_B site of the RC protein. However, also the B-branch is able to transfer electrons as well^{38–41}. Such a B-branch electron transfer is only observed when the relative energy levels of the A- and B-branch cofactors is modulated like for example in Q_A -deficient mutants^{42,43}. Although the non-heme iron is situated between Q_A and Q_B , experiments showed that it is not part of the electron transfer chain to Q_B ^{44,45}. In most bacterial species, an additional carotenoid molecule is bound to the bRC. However, it is also not involved in the electron transfer to Q_B but it has an important role in the self-protection of

the bRC under stress conditions such as high light^{28,46}. In the PSII system several carotenoid molecules are bound^{16,31,47}.

The here described charge transfer to Q_B in the bRC is only slightly altered in PSII RC proteins. Based on several experiments it was suggested that the first charge separation does not occur at the special pair in PSII (formed by Chl_{D1-1} and Chl_{D2-1} ; see Figure 4b) but at the level of the second chlorophyll in the A-branch (Chl_{D1-2} ; see Figure 4)⁴⁸⁻⁵¹. In both the bRC and PSII RC proteins, two electron transfer reactions and the binding of two protons to Q_B lead to the fully reduced and protonated quinol, which leaves the Q_B site and is replaced from the CoQ pool by a fully oxidized CoQ molecule (see Figure 1 and Figure 2) closing the reaction cycle. However, there are differences between the cofactor organization of bRC and PSII RC proteins. First, in each branch of PSII RC proteins an additional peripheral accessory chlorophyll molecule (Chl_z ; see Figure 4b) is bound. The peripheral accessory chlorophyll molecules (Chl_{zD1} and Chl_{zD2}) do not participate in electron transfer from the special pair to Q_B . But experiments showed, that they are likely to be involved into self-protection of the protein against too much light^{32,52} and in the transfer of the excitation energy from the LHC subunits to the special pair^{52,53}. The WOC, which is a central part of the PSII RC proteins, is absent in bRC. In the WOC, water is split into protons and oxygen (see Figure 2). The WOC is formed by a manganese-calcium cluster (four manganese ions and one calcium ion, see Figure 4b), a chloride ion, and a 'special' conserved tyrosine Y_Z (D1-161 in *Thermosynechococcus elongatus*) located near to the manganese-calcium cluster. Tyrosine Y_Z abstracts electrons from the manganese cluster forming a tyrosine radical intermediate⁵⁴. The manganese cluster is then reoxidized by reducing water to oxygen and protons. The protons are transferred to Y_Z and from there via a proton transfer pathway to the lumen⁵⁵. Until now the exact geometry of the manganese cluster is not known, since the geometry of the cluster is not resolved exactly in the existing crystal structures^{16,31,47,56}. But based on mutational studies, several residues were identified influencing the manganese cluster and the oxygen formation⁵⁷⁻⁶⁰. Most of these residues are located in the D1 subunit. However an additional subunit is part of the PSII complex, which is involved in the stabilization and function of the manganese-calcium cluster^{61,62}.

A milestone in photosynthetic research are the first crystallographically determined structures of the bRC proteins of the purple bacteria *Rb. sphaeroides* and *B. viridis* in the eighties of the last century^{63,64}. Later also the three-dimensional structure of the bRC of thermophilic cyanobacterium *Thermochromatium tepidum*⁶⁵ was determined crystallographically. Nowadays more than fifty different crystal structures of the bRC from the named three species are deposited in the RCSB protein data base^{66,67}, including *wild-type* and mutant structures, and structures obtained at different pH values, under different crystallization conditions and with different crystallization

methods. In addition, in the last decade first crystal structures of the complete PSII RC protein were solved crystallographically with reasonable resolution^{16,31,47}. In the past, the bRC was often studied as a model for the more complex PSII protein, since most of the functions (such as electron transfer to CoQ or inhibition by triazine herbicides) are conserved between the proteins. Furthermore the bRC is an ideal protein to study protonation dependent redox reactions, since the electron and proton transfer to Q_B are coupled. A wide variety of different experimental procedures and protocols were developed and used gaining a good overall description of the bRC protein and its function. However details of the reaction mechanisms such as the proton transfer to Q_B remain still unclear.

2 Mechanism of the Bacterial Reaction Center

Despite of the wide variety of different experimental and theoretical studies and the huge amount of information about the reaction and function of bRC, several questions remain unclear. For example, the key residues of proton transfer are well known but the exact organization of the proton transfer network is not determined yet. Since the work presented here is focused on the molecular details of the Q_B site of the bRC and proton and electron transfer reactions to Q_B, the two CoQ binding sites, the catalytic cycle and the proton transfer to Q_B will be summarized and open questions will be emphasized. In the following, the numbering of key residues refers to the one of *Rb. sphaeroides*. Whenever the information is not taken from studies of the bRC of *Rb. sphaeroides*, it is explicitly mentioned.

2.1 The Coenzyme Q Binding Sites

Type II RC proteins all have two CoQ molecules bound (see Figure 3b and Figure 4b), which are crucial for the light-induction reaction. For each of these CoQ molecules a binding site (Q_A and Q_B binding site) exists. In bRC as well as in PSII RC proteins, the two binding sites are connected via a distance of about 15 Å by the non-heme iron with its ligands (HisL190, HisL230, HisM219, HisM266, and GluM234 in the bRC) as it is depicted in Figure 5. However, in PSII RC proteins, the fifth ligand is not a glutamate residue, but a bicarbonate⁶⁸. The two binding sites differ in their structure and in thus, the two CoQ molecules differ in their function. In the following the structural details and the functionality of each binding site will be described for bRC proteins.

Crystal structures of the bRC of *Rb. sphaeroides* revealed two positions of CoQ in the Q_B binding site: distal and proximal to the non-heme iron^{69,70} (see Figure 6a). In these structures, Q_B has to move by 4.5 Å in the binding site to change its position from distal to proximal. Moreover, it was proposed based on crystallographic information, that the head group of Q_B has to undergo a 180° rotation during the movement⁷⁰. Crystallographic experiments at pH 8.5 showed, that

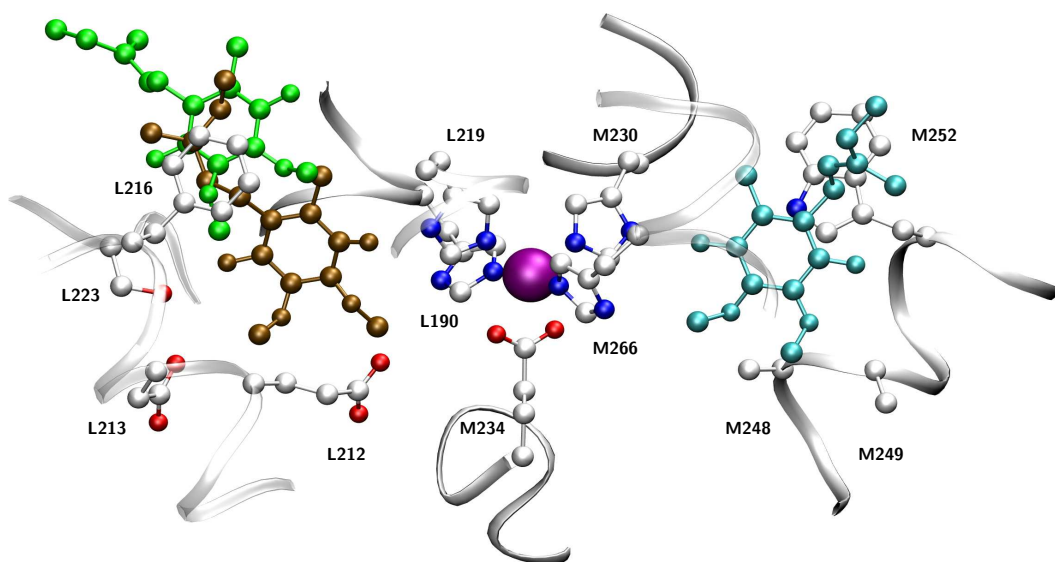


Figure 5: The connection between the two Coenzyme Q binding sites. The Q_B (distal position green; proximal position brown) and the Q_A (cyan) binding sites are connected via the non-heme iron with its ligands (HisL190, HisL230, HisM219, HisM266, and GluM234). In addition some important residues of each binding site are shown. The figure is made using vmd³³ and based on the crystal structure 2C8J³⁴.

the population of the two positions depend on the illumination state of the protein: In the neutral, dark-adapted state, Q_B is bound to 45 % in the proximal position whereas it is bound to 90 % in the proximal position under illumination⁶⁹. Moreover, a theoretical study suggested that the population of the two positions is pH dependent⁷¹. Mutational and crystallographic studies showed, that the amino acid at the position L209 influences the position of Q_B in bRC of *Rb. sphaeroides*^{72,73}. In the ProL209→Tyr mutant, Q_B occupies a single binding position similar to the proximal one of the *wild-type*. In the ProL209→Phe mutant, Q_B occupies a single position, which is intermediate between the proximal and the distal position of the *wild-type*. But the binding position of Q_B in the ProL209→Glu and the ProL209→Thr mutants resembles the *wild-type* situation^{72,73}. In the *wild-type* bRC proteins of *B. viridis*, however, an alanine is found at the L209 position. A recent crystallographic study showed, that in crystals of the bRC from *B. viridis* the proximal position is predominantly occupied by Q_B independent of the illumination state of the protein⁷⁴. A molecular dynamics study simulated spontaneous movement of Q_B from the distal to the proximal position, in both, the $Q_A Q_B$ and the $Q_A^- Q_B$ state. For these calculations they used a Q_B in the distal position which has the same orientation as in the proximal one, i. e. is not rotated by 180°. By further examination using the adiabatic mapping method, they were able to show, that such a rotation of 180° of Q_B during movement from distal to proximal

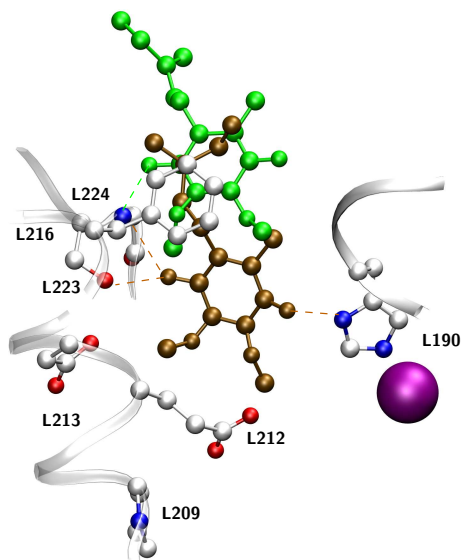
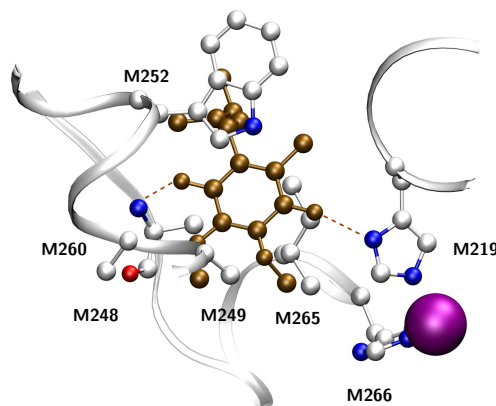
a) Q_B siteb) Q_A site

Figure 6: The two Coenzyme Q binding sites of bacterial reaction center proteins. (a) The Q_B site of bRC is shown with the quinone its proximal (brown) and distal (green) position. (b) The Q_A binding site of bRC with a both CoQ molecule (brown). In both pictures, important residues for binding as well as the non-heme iron (purple) are shown. The figure is prepared with vmd³³ using the crystal structure 2J8C³⁴.

is energetically not feasible. Thus, the authors of this molecular dynamics study concluded, that observed distal position is unproductive⁷⁵. Therefore, the two different Q_B positions are unlikely to be important for the catalytic reaction^{74,76,77}.

In most bRC proteins, the two bound CoQ molecules are chemically identical like for example in both binding sites of the bRC from *Rb. sphaeroides* a ubiquinone molecule is bound. Moreover both sites are symmetrically located in the protein (see Figure 3) and some important residues are equivalent: in both binding sites an aromatic residue forms a $\pi - \pi$ interaction with the CoQ molecule (TrpM252 in the Q_A site and PheL126 in the Q_B site; see Figure 6) and a hydrogen bond is formed to a histidine ligating the non-heme iron (HisM219 in the Q_A site and HisL190 in the Q_B site; see Figure 6) located in the corresponding site. However it has been shown experimentally, that the behavior of Q_A and Q_B differs substantially⁷⁸. Q_A accepts only one electron and no proton whereas Q_B accepts two protons and two electrons. Moreover the Q_B can leave the bRC whereas Q_A remains bound. Experiments with the bRC of *Rb. sphaeroides* showed that the distinct behavior of the proximally bound Q_B and Q_A is not caused by differences in their structure but by differences in the local protein environment like different hydrogen bonding patterns in the two binding sites^{79–81}. The two carbonyl oxygen atoms of a proximally bound Q_B are weakly

and symmetrically bound (by HisL190 and the backbone of L224; see Figure 6a), whereas the two carbonyl oxygen atoms of Q_A are bound with asymmetric strength (by HisM219 and the backbone of M260; see Figure 6b). The asymmetrical hydrogen bonding leads to a change in the electronic structure of the radical Q_A^- compared to the radical in solution making a second reduction of Q_A unlikely^{82,83}. Moreover, electron nuclear double resonance (ENDOR) spectroscopy experiments showed recently, that the hydrogen bonds to Q_A are significantly shorter in the Q_A^- state compared to the neutral quinone state leading to a stabilization of the radical anion⁸¹. The sidechain character at the position M260 (an alanine in the *wild-type*; *Rb. sphaeroides*) is important for the binding of Q_A , even so it is not forming interaction with Q_A . But the mutation of Ala to a bulky aminoacid such as tryptophan leads to an exclusion of Q_A since the mutation makes the Q_A site too small for binding a CoQ molecule^{42,43}. Exclusion of Q_A can also be achieved by mutation of AlaM248→Trp^{42,43}. Moreover, the mutation of IleM265 leads to a changed Q_A binding affinity and to a changed midpoint potential of Q_A indicating that IleM265 is in van-der-Waals contact with Q_A ^{78,84}. Experiments showed, that, also not directly involved in binding of Q_A , the character of the sidechain of M266 (histidine in the *wild-type*; *Rb. sphaeroides*) is important for correct functioning of Q_A ⁸⁵. HisM266 is moreover a ligand of the non-heme iron.

In contrast to Q_A , there are polar and acidic residues near Q_B , in particular SerL223, GluL212 and AspL213. In the *wild-type* bRC proteins of *B. viridis*, an asparagine is found at the position L213. The named three residues (L223, L212 and L213) are crucial for proton transfer to Q_B ^{86,87}. The key residues of proton transfer and proposed proton transfer pathways are discussed later in detail. In contrast to the polar environment of Q_B , Q_A is situated in an unpolar environment making a proton transfer to Q_A rather unlikely. In fact proton transfer to Q_A is never observed. At the corresponding, symmetry-related positions of GluL212 and AspL213, two alanines are located (M246 and M247; see Figure 6) in the Q_A site. The functional importance of the different character of the aminoacids at the position L212/M246 and L213/M247 in the two binding pockets are highlighted in experiments with site-specific mutants: the double mutants GluL212-AspL213→Ala-Ala (AA strain) and AlaM246-AlaM247→Glu-Asp as well as the quadruple mutant GluL212-AspL213-AlaM246-AlaM247→Ala-Ala-Glu-Asp (RQ strain) are not able to grow photosynthetically⁸⁸⁻⁹⁰. A recent study combining neutron scattering and Brownian dynamics experiments, showed that in the AA mutant (with or without an additional mutation at M249) the flexibility of the protein is increased compared to the *wild-type* protein⁹¹. Since rigidity in catalytic sites is thought to be important for the function of the protein, the increased flexibility of the AA mutant might be the reason for the observed inability to grow photosynthetically. However, photoactivity of these mutants could be restored by further, second-site mutations such as the ArgM231→Leu or the AsnM44→Asp mutation in case of the AA mutant⁹². Also for the

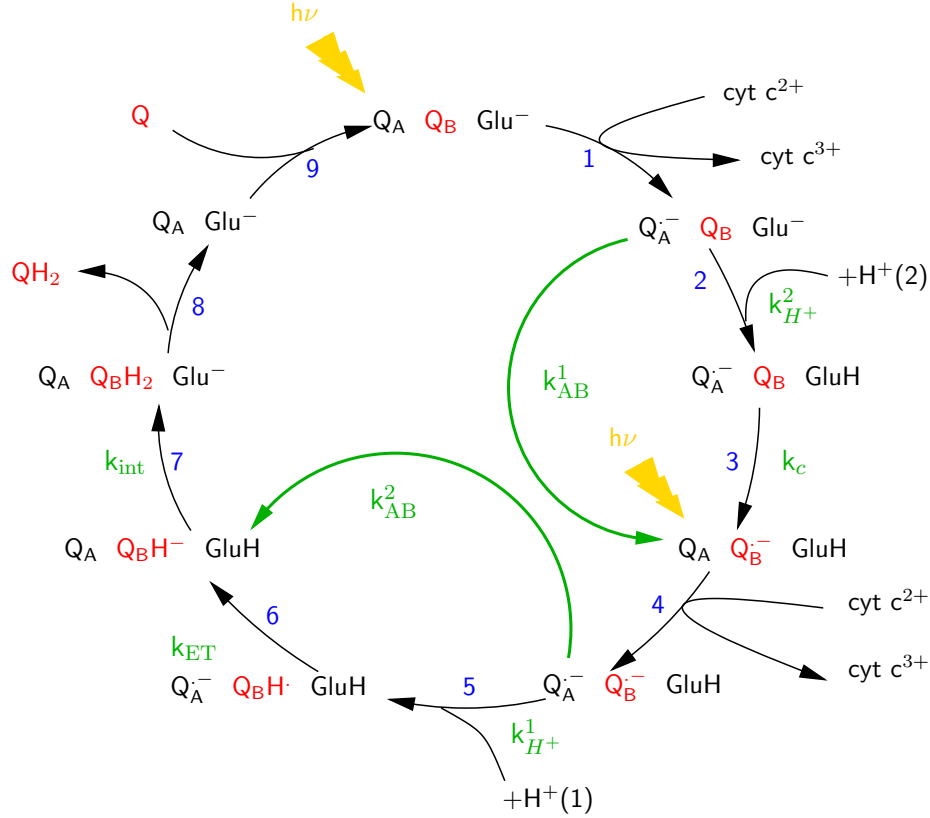


Figure 7: Redox- and protonation cycle of Q_B . The light-induced reaction in the RC of *Rb. sphaeroides* is schematically depicted. During the reaction two electrons and two protons are transferred to Q_B . Nine steps are needed for a complete reaction cycle: step 1/4 - electron transfer to Q_A and re-reduction of the special pair; step 2 - proton uptake from cytoplasm and transfer to GluL212; step 3/6 - electron transfer to Q_B ; step 5 - proton uptake from cytoplasm and transfer to Q_B^- ; step 7 - internal proton transfer to Q_BH^- ; step 8/9 - QH_2 is replaced by Q in the Q_B site. For more details see text.

RQ mutant, photoactivity can be restored, for example by the introduction of additional mutations at the positions L213 and M44 (AlaL213→Tyr; AsnM44→Asp)⁸⁸. Thus, even though the different character of two binding sites seems to be important for the reactivity of the protein, the bRC is very robust, meaning that it can restore its function by second site mutation in case of loss of function mutations in the CoQ binding sites.

2.2 Proposed Catalytic Cycle

In the following, the catalytic cycle of reduction of Q_B is described based on the cycle proposed by Ädelroth *et al.* (2000)⁹³. Many of the catalytic steps are experimentally determined today, but some steps of the here presented reaction cycle are still controversially discussed like for example whether GluL212 is protonated before or after the first reduction of Q_B . Whenever a proposed

step of the catalytic cycle is still debated it will be mentioned in the following.

The light induced reaction cycle (see Figure 7) starts with the electron transfer from the special pair via the A-branch to the primary quinone (Figure 7: step 1). For each of the two reductions of Q_B , photon absorption by illumination of the bRC samples or excitation energy transfer from the LH complexes (see Figure 1) is needed in the bRC for a photo-induced electron transfer. To reactivate the special pair for the next electron transfer step, the oxidized special pair is reduced by electron transfer from the soluble electron carrier cytochrome c_2 (see Figure 1). From the reduced primary semiquinone Q_A^- , the electron is transferred over 15 Å to the quinone Q_B (Figure 7: step 3), which results in the formation of the semiquinone $Q_B^{\cdot-}$ ^{86,87}. Experiments showed, that during the semiquinone formation of either Q_A^- or Q_B^- a proton (H^+) (2) (see Figure 7) is taken up at neutral pH, leading to experimentally measured substoichiometric proton uptake during the formation of Q_A^- and Q_B^- ⁹⁴. The described substoichiometric proton uptake was observed for bRC of different bacteria species⁹⁴⁻⁹⁶. The protonation of GluL212 during the first reduction step was determined experimentally⁹⁷. Moreover it was shown that this proton is later used to protonate the doubly reduced Q_B (see Figure 7: step 7). But whether the protonation of GluL212 occurs after the reduction of Q_A (like it is shown in Figure 7; step 2) or of Q_B is still under debate⁹⁸⁻¹⁰². The proposed proton uptake mechanisms and pathways will be discussed in the next section.

The two CoQ binding sites are connected to each other by the non-heme iron and its ligands (see Figure 5). The 15 Å distance between Q_A and Q_B is bridged by hydrogen bonds through the His- Fe^{2+} -His-connection (Q_B -HisL190- Fe^{2+} -HisM219- Q_A). This connection could be used for the transfer of the electron from Q_A^- to Q_B . But experiments with bRC in which the non-heme iron is depleted or replaced with other divalent metals measured an almost unchanged electron transfer rate: in iron depleted bRC proteins the electron transfer rate is reduced by a factor of about two and in reconstituted bRC proteins the *wild-type* rate is restored⁴⁴. Moreover, in a kinetic X-ray absorption experiment, no oxidation of the non-heme iron was observed during electron transfer¹⁰³. These experiments showed that the non-heme iron is not involved in the electron transfer reaction from Q_A^- to Q_B . A time-resolved Fourier transform Infrared (FTIR) spectroscopy experiment determined that Q_B is reduced previous to the oxidation to Q_A^- during the first reduction step¹⁰⁴. This observation indicates that Q_B cannot be directly reduced by Q_A^- but by an intermediary electron donor, which cannot be the non-heme iron^{103,104}. But the existence of such a intermediary electron donor between Q_A and Q_B is still under debate¹⁰⁵.

The first reduction of Q_B is comprised of the upper described steps (step 1 to 3; see Figure 7). Rates were measured spectroscopically for the first reduction (k_{AB}^1 ; see Figure 7) to be about $4 \cdot 10^3 \text{ s}^{-1}$ at pH 8.5 and to be pH-dependent:⁹⁹ With increasing pH (pH > 8.5), the first electron

transfer rate k_{AB}^1 decreases indicating the involvement of a protonation step (k_{H+}^2). As explained, the first proton uptake does not lead to a protonation of Q_B but to a protonation of the nearby GluL212 (see Figure 7). Measuring the first electron transfer rate k_{AB}^1 in different 'substituted' bRC proteins, in which the naturally bound Q_A is replaced by different CoQ species with varying redox potentials, showed, that the electron transfer rate k_{AB}^1 is not changed independent of the used CoQ species^{106,107}. Such a driving force assay is used to determine, whether a measured electron transfer rate is determined by the electron transfer or by any other (slower) process, like a conformational change. In the case of the bRC, the first measured electron transfer rate k_{AB}^1 is independent on the redox potential of Q_A and thus of the electron driving force. This means, that a slow step with the rate k_c determines the measured electron transfer rate k_{AB}^1 ^{106,107}. The rate determining step could be a conformational change. Such a process is also named conformational gating. Measurements at pH 8.5 estimated the first protonation rate k_{H+}^2 to be about 500 s^{-1} , and the conformational gating to be about $4 \cdot 10^3 \text{ s}^{-1}$. In the past it was thought that the conformational gating is the change of Q_B from the distal (unproductive) to the proximal position. But for bRC proteins, in which Q_B is only bound to the proximal position (like the ProL209→Tyr mutant) a more pronounced gating was observed⁷². Moreover, if the movement of Q_B would be rate-limiting for the gating step, the rate of this gating should depend on the isoprene chain length of Q_B , which is not the case^{108,109}. Thus, the idea of the movement of Q_B between the two positions being the gating mechanism is nowadays thought to be very unlikely^{72,74,76,77}. An ENDOR spectroscopy study of the *wild-type* bRC protein and SerL223→Ala mutant showed, that SerL223 must be involved in the gating. When Q_B is bound, SerL223 is likely to form a hydrogen bond to AspL213, however, when Q_B is reduced, the hydrogen bond is formed between SerL223 and the semiquinone. But in a FTIR experiment the interaction of SerL223 with Q_B or Q_B^- could not be verified¹¹⁰. Thus, it was concluded, that the hydrogen bond cannot be the main factor determining the gating. Until today, the exact molecular mechanisms leading to the observed gating are not determined.

After first reduction of Q_B (step 3; see Figure 7), illumination leads to the second electron transfer to the primary quinone Q_A (step 4). Cytochrome c_2 reactivates the oxidized special pair by rereduction. In contrast to the first electron transfer from Q_A^- to Q_B , the second electron transfer is coupled with the protonation of Q_B^- (Figure 7: step 5). The first, fast step of proton-coupled electron transfer is the protonation of Q_B^- leading to the energetically unfavorable semiquinone intermediate Q_BH^\cdot (see Figure 7: step 6)^{106,111–113}, followed by the second, rate-limiting step, the second electron transfer from Q_A^- , forming the doubly-reduced and singly-protonated quinol Q_BH^- (see Figure 7: step 7). The here described proton-coupled second electron transfer was discussed controversially for a long time^{106,111,112,114–116}. Earlier proposed reaction cycles of

doubly reducing and protonating of Q_B included the formation of the doubly-reduced and deprotonated intermediate, the quinol Q^{2-} . In such a cycle the protonation and reduction steps would be separated. However, Graige *et al.* (1996) were able to rule out models with separated protonation and reduction steps by experiments with a series of 'substituted' bRC proteins in which different naphthoquinone molecules are bound to the Q_A site¹⁰⁶. Additionally, spectroscopic experiments with the bRC of *Rb. capsulatus* and *Rb. sphaeroides* in chromatophores detected the protonated semiquinone species Q_BH^\cdot below pH 7 after the first electron transfer step¹¹³. The second reduction rate k_{AB}^2 is with $3.5 \cdot 10^2 s^{-1}$ (measured at pH 8.5) slower compared to the first reduction rate k_{AB}^1 . Also the second reduction rate k_{AB}^2 describes a two-step process, first the fast protonation of the semiquinone (with the rate k_{H+}^1) and a slow electron transfer from Q_A^- (with the rate k_{ET}) as it is shown in Figure 7. The protonation rate k_{H+}^1 and the electron transfer rate k_{ET} were estimated to be around $350 s^{-1}$ and $3.5 \cdot 10^2 s^{-1}$ at pH 8.5⁹⁹ in the bRC of *Rb. sphaeroides*.

The quinol Q_BH^\cdot takes up a second proton by deprotonating GluL212 (Figure 7: step 7)^{97,117}. The internal proton transfer from GluL212 to Q_BH^\cdot with the rate k_{int} is very fast.⁹⁹ Finally, the quinol Q_BH_2 is released to the CoQ pool of the membrane (Figure 7: step 8)¹¹⁸. Until a new ubiquinone is bound from the CoQ pool of the membrane (Figure 7: step 9), the empty Q_B binding pocket is occupied by a cluster of five or six water molecules¹¹⁹. Photon absorption by illumination of the bRC samples or excitation energy transfer from the LH complexes will initiate a new protonation and oxidation cycle in the bRC.

2.3 Proton Transfer to Q_B

For the bRC, a wide variety of experimental studies using the bRC of the bacteria *Rb. sphaeroides* and *Rb. capsulatus* led to the identification of key residues of proton transfer to Q_B and of possible proton transfer pathways to Q_B .

The key residues of proton transfer are depicted in Figure 8. AspH124, HisH126 and HisH128, TyrM3, AspM240, and AspH224, and AspM17 were proposed to form proton entry points for proton transfer to Q_B ^{86,87}. Inside the protein, GlnH173, GluL212, HisL190, AspL210, AspL213, SerL223, and AsnM44 were identified to be key residues of proton transfer to Q_B . Experiments showed, that the mutation of SerL223 (SerL223→Ala) leads to a reduced proton transfer rate in bRC proteins, highlighting the functional importance of SerL223 for the proton transfer to Q_B ¹²⁰. In the last years, different proton transfer pathways involving some of the key residues named above were proposed. Based on X-ray crystallographically determined structures of the bRC, three possibly interconnected proton transfer pathways (P1, P2, and P3) were proposed^{70,87,121}. The three proton transfer pathways are schematically shown in Figure 8. Starting either at

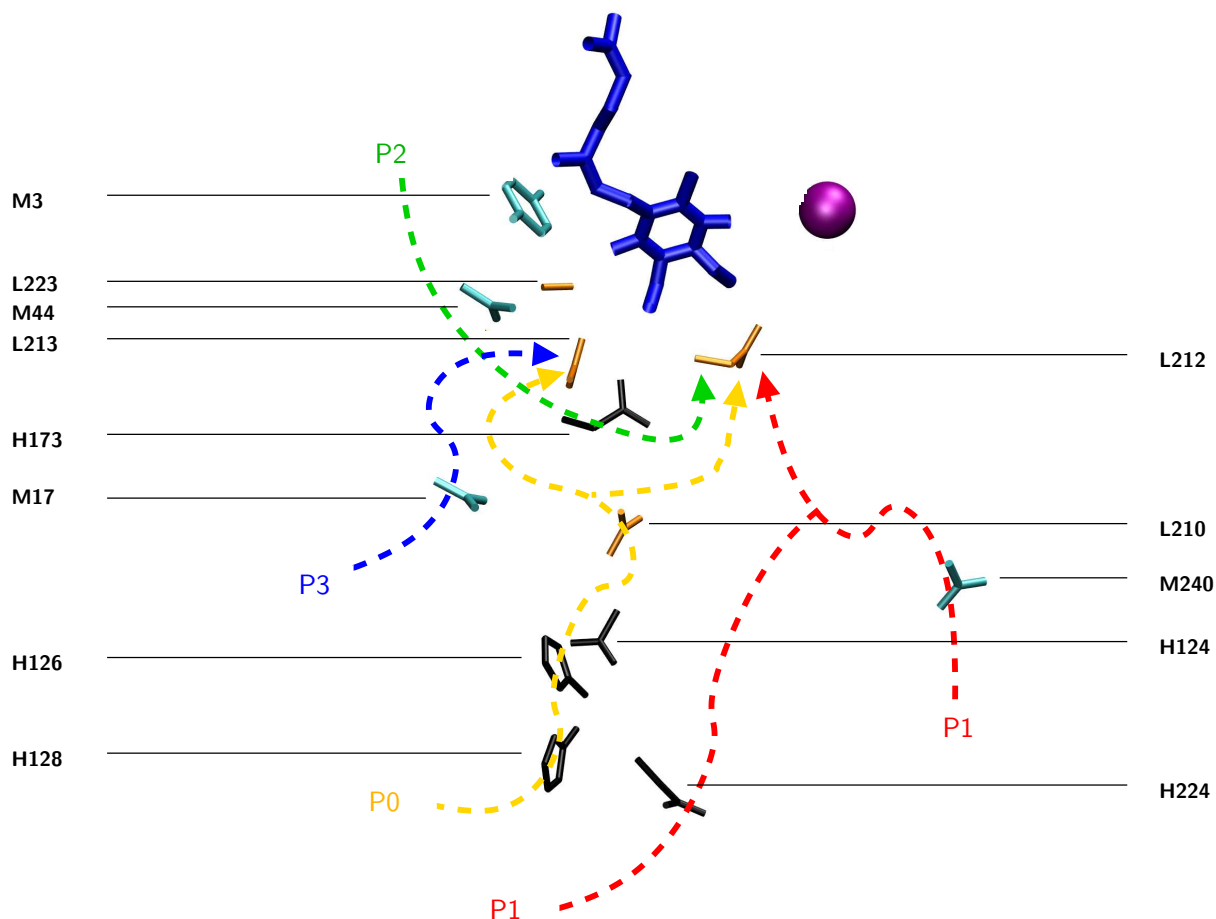


Figure 8: Key residues of proton transfer to Q_B. All residues are colored corresponding to the subunit (M=cyan, L=orange and H=black). Only sidechains are shown. In addition the proposed proton transfer pathways are schematically shown. Additionally the non-heme iron (purple) and the Q_B (blue) are depicted. For the picture the crystal structure 2J8C³⁴ was used.

AspH224 or near to AspM240, the proton transfer pathway P1 transfers the proton over a distance of about 20 Å via a continuous chain of water molecules to GluL212. The second proton transfer pathway P2 connects its proton entry point TyrM3 over a distance of about 20 Å via GluH173 and a pool of water molecules to AspL213. In contrast to P1 and P2, the third proton pathway P3 is short (about 7 Å) and consists of only three residues: the proton entry point AspM17, a water molecule and AspL213. The proton pathways P2 and P3 lead to the ultimate proton donor AspL213, whereas P1 leads to the ultimate proton donor GluL212. In a last proton transfer step, the proton is transferred from these ultimate proton donor residues to Q_B. Even though the pathway P3 is much shorter and thus might be used more frequently, the other transfer pathways P1 and P2 can not be ruled out based on the available experimental data¹²¹. The other two proton transfer pathways P1 and P2 might for example be used to transfer protons to the Q_B

binding site if P3 is inactive due to a mutation.

In contrast to the three possible proton transfer pathways, a single branched proton transfer pathway P0 (shown schematically in Figure 8) was suggested for both protons^{99, 111, 115, 120, 122–126}. The proton transfer pathway P0 is similar to the earlier proposed proton transfer pathway P3. The surface residues HisH126, HisH128 and AspH124 have been identified crystallographically to form a metal binding site of the bRC^{126, 127}. In experiments significantly reduced proton transfer rates were observed when a cadmium, zinc or nickel ion is bound to the bRC⁹⁹ and thus, HisH126, HisH128 and AspH214 were proposed to form the single proton entry point of the proton transfer pathway P0. The protons taken up at AspH124, HisH126, and HisH128 are transferred from there via the cooperating residues AspM17 and AspL210 to either GluL212 or AspL213. Spectroscopic measurements showed significantly lowered reduction rates ($k_{AB}^1 = 5 \cdot 10^2 \text{ s}^{-1}$ and $k_{AB}^2 = 1 \cdot 10^2 \text{ s}^{-1}$ at pH 8.5) in the double mutant HisH126→Ala and HisH128→Ala associated with a much slower proton uptake in both reduction steps⁹⁹. In the double mutant, the proposed proton entry point of the pathway P0 is not able to transfer protons and thus, the pathway P0 might be inactive. The measured slower proton uptake rates highlight the importance of the proton transfer pathway P0 in bRC proteins. However, the fact that even the double mutant is still able to grow photosynthetically⁹⁹ makes a single proton pathway rather unlikely.

Experiments showed that the mutation of AspL213 leads to a drastically reduced proton transfer in bRC proteins and thus to organisms which are not able to grow photosynthetically. But second-site mutations can restore the *wild-type* proton transfer rate and thus lead to revertant organisms, which are again able to grow photosynthetically¹²⁸. Such second-site mutations are for example AspM233→Cys or ArgH177→His^{125, 128}. The positions of AspM233 and ArgH177 are depicted in Figure 9. In the revertant crystal structures a higher flexibility of GluH173 was observed: two positions for GluH173 were found to be populated to nearly the same extent in the mutants¹²⁵. Moreover additional water molecules were found in the structures forming together with the mutation site and GluH173 possibly different proton transfer pathway(s) to Q_B^{125, 128}. Based on a continuum electrostatic calculation, an alternative proton transfer pathway in the revertants was proposed including AspH119, GluH229 and GluH224 as proton entry points and the residues GluH122, AspH170, GluH230, and GluH173 inside the protein (see Figure 9)¹²⁹. The participation in proton transfer was already suggested by a crystallographic study for the residues AspH170 and GluH122 in a pathway similar to the pathway P1, but including additionally GluM236, LysH130, HisH68 and ArgH70 (proton entry points; see Figure 9)⁷⁰. In the same study a second proton transfer pathway was suggested similar to the pathway P3 but involving additionally GlnM11 and SerM8 as proton entry points and GlnH174 inside the protein. Moreover, a mutational study showed that ArgL207 and ArgL217 are important for efficient proton

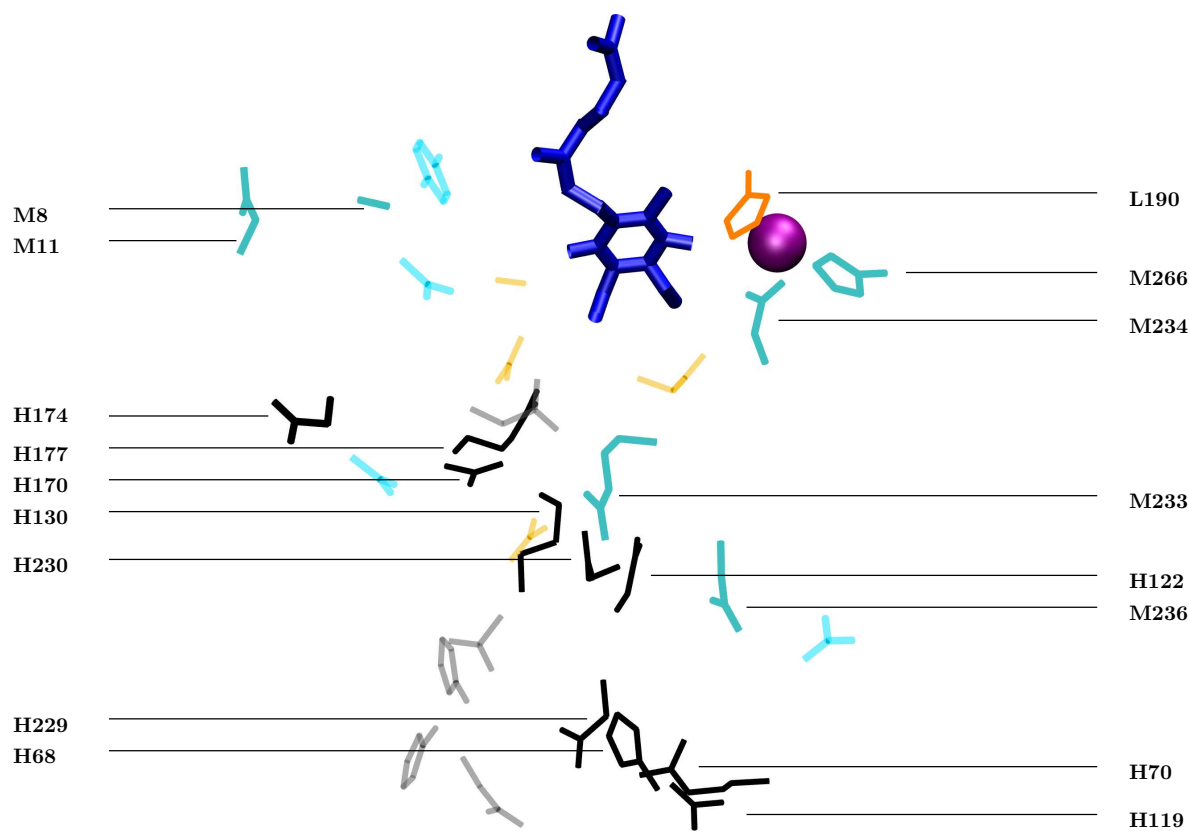


Figure 9: Residues, which might be involved in proton transfer to Q_B . The possibly involved residues are colored corresponding to the subunit (M=cyan, L=orange and H=black). and are shown in addition to the key residues (see Figure 8; M=transparent cyan, L=transparent orange and M=transparent black). The view is slightly altered to the previous figure (Figure 8). The figure is made as described in Figure 8.

transfer¹³⁰. The residues AspL213, AspL210, AspM17, GluH173, AspH170 and AspH124 form an acidic cluster near to Q_B . Since several of these residues are part of the proposed proton pathways, it was also suggested, that this cluster is of functional importance¹²¹. Also for the previously discussed residue ProL209, which might be involved in the regulation of the Q_B binding position, it was shown that mutation of this residues leads to a differentiating proton transfer in the mutants⁷².

However, also the AspM44→Asn mutation restores the photoactivity in the AspL213→Asn mutant of the bRC protein of *Rb. sphaeroides*.

Experimental studies showed a clear correlation of the aminoacid character at the position L213 and M44. In *Rb. sphaeroides* the combination AsnM44/AspL213 is found, whereas the combination AspM44/AsnL213 is the *wild-type* pattern of *B. viridis*. The double mutant AspL213→Asn/AsnM44→Asp of *Rb. sphaeroides* grows photosynthetically while the single mu-

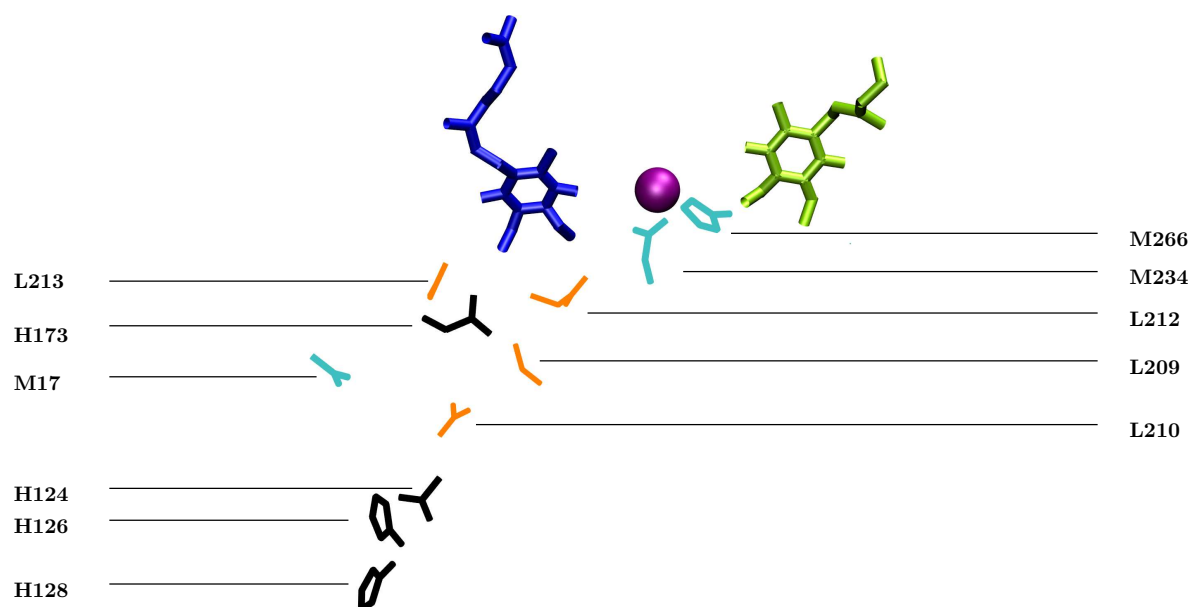


Figure 10: Proton transfer via a proposed anti-cooperative cluster. Residues participating in the proposed anti-cooperative cluster (ProL209, AspL210, GluL212, AspL213, AspM17, GluM234, AlaM249, HisM266) and the proposed proton entry point (AspH124, HisH126, and HisH128) are shown. The figure is made as described in Figure 8 and additionally CoQ_A (colored lime) is shown.

tant AspL213→Asn does not^{128,131}. It seems very likely, that the combination of a polar, unprotonatable (asparagine or glutamine) and of a protonatable residue (aspartate or glutamate) at the positions M44 and L213 is required for proton transfer to Q_B.

Recently, spectroscopic experiments of bRC proteins of *Rb. sphaeroides* mutated at the position HisM266 highlighted the importance of HisM266 for proton transfer⁸⁵: In the HisM266→Ala and the HisM266→Leu mutant the proton uptake stoichiometries after reduction of Q_A and Q_B differs compared to the *wild-type*. Since HisM266 is a ligand of the non-heme iron and experimental evidences exist, that mutation of GluM234 (another ligand of the non-heme iron) also influences the proton transfer to Q_B, it might also be, that the non-heme iron with its ligands is important for proton transfer pathways. Moreover experiments with the AA mutant of *Rb. capsulatus* showed, that by a second site mutation (AlaM247→Tyr) proton transfer to Q_B can be restored to the *wild-type* level⁹⁴, which means, that also residues located in the Q_A site (see Figure 6b) are involved in proton transfer. Also mutation of ProL209 (see Figure 6a) leads to alteration in proton transfer⁹⁴. Thus, it was proposed, that a strong interaction through an extended hydrogen network between the Q_A and Q_B exists, and is necessary to facilitate proton transfer to Q_B^{94,132}. This connection might be altered, when the region around Q_B becomes more

flexible (for example by the mutation of ProL209). In the ProL209, the AA, and the HisM266 mutants the proton uptake stoichiometries after Q_A and Q_B are remarkable similar^{85,132}. All these data led Cheap *et. al*⁸⁵ propose that between the proton entry point of AspH124, HisH126, and HisH128 and the Q_B binding site an anti-cooperative (meaning that the proton binding of one residue in the cluster disfavors additional proton binding to a second residue) cluster formed by the residues ProL209, AspL210, GluL212, AspL213, AspM17, GluM234, AlaM249, HisM266 (see Figure 10) and several water molecules facilitate proton transfer. Such an anti-cooperative cluster would rather suggest an extended proton transfer network than a single proton transfer pathway.

Due to the described facts, that second site mutations can restore photoactivity of previously photoinactive mutants by reactivating proton transfer to Q_B , the different aminoacid combinations of Asn/Asp and Asn/Asp at the positions M44/L213 in different bacterial species and the proposed anti-cooperative cluster, it could be concluded that the concept of distinct and separated single proton pathways to Q_B is unlikely. It could further be assumed, that the bRC acts more like a proton sponge, meaning that a very extended proton transfer network is used to transfer protons to Q_B , in which several residues are more important than other ones and in which certain proton transfer events are more and others are less likely. It is not determined yet which of the two ideas, either distinct pathways or a proton sponge, describes the proton transfer to Q_B in the bRC best.

3 Aim of the Thesis

The aim of this work is to gain further insights into the molecular details of the Q_B binding site and the proton and electron transfer reactions to Q_B . For this purpose several different approaches, such as conservation analysis based on multiple sequence alignment, continuum electrostatics, quantum mechanics and network calculations are combined. To investigate whether the electron transfer to Q_B via the A-branch is determined by the same mechanism in PSII RC and bRC proteins, a conservation study of these proteins is performed. In addition, this conservation study is used to analyze similarities and differences of the bRC and PSII RC proteins. By continuum electrostatic calculations, the population of Q_B in the two possible positions in the bRC of *Rb. sphaeroides* is examined in dependence of the illumination state and the pH. A conservation analysis and a hydrogen bonded network calculation are performed to see whether the proton transfer to Q_B is organized in distinct pathways or in a proton sponge.

Synopsis of the Manuscripts

The central issue of this work is to gain further insights into the molecular details of the Q_B binding site and the proton and electron transfer reactions to Q_B . During the light induced reaction two protons and two electrons are transferred to Q_B . The reaction intermediates (especially the semiquinone $Q_B^{\cdot-}$) are tightly bound whereas the reactant (the quinone Q) and the product (the quinol QH_2) can leave the Q_B binding site. After excitation of the special pair, the electrons are transferred to Q_B . In both, the bRC and the PSII RC proteins, the electrons are transferred via the A-branch. The proton transfer to Q_B differs in bRC and PSII RC proteins. Although the key residues of proton transfer are experimentally determined in bRC, the exact organization of the proton transfer to Q_B is still unknown. Two contradictory ideas exist: first, that the proton transfer follows distinct pathways and second that the cytosolic part of the protein acts like a proton sponge. In the bacterial bRC of *Rb. sphaeroides*, two alternative binding positions (distal and proximal) have been observed in the Q_B binding site, but most likely only the proximal position is thought to be important for the reaction. It was proposed, that for the movement from proximal to distal, Q_B has to change its orientation by 180° .

Combining Quantum chemical calculations with continuum electrostatic followed by Monte Carlo (CE/MC) calculations, it was shown that there is energetically no need for different orientations of Q_B in the two binding positions. Analysis of the electron density of the bRC from *Rb. sphaeroides* showed that the same orientation of Q_B in both binding positions is in agreement with the density. Thus, it was proposed that, since Q_B has the same orientation in the two positions, the 180° rotation of Q_B is not needed during the movement of Q_B from one position to the other. By CE/MC calculations, main factors (the protonation state of the ultimate proton donors GluL212 and AspL213) are identified determining the illumination state dependence and the newly observed pH dependence of the population of Q_B in the two positions (Manuscript A). The described study resulted in the highest resolution crystal structures available today. All further calculations were performed based on these crystal structures. Further investigation of the populations using a CE/MC approach showed, that the main factors determining the population (the protonation state of GluL212 and AspL213) are also important to keep the semiquinone bound in the Q_B site (Manuscript B). Based on a quantum mechanical approach, the energetics of protonation and reduction reactions of different quinone systems were investigated (Manuscript C). This study showed, that the semiquinone state is rather unstable in aqueous solution and it protonates very fast. The common usefulness of the CE/MC approach applied here to investigate redox dependent proteins was highlighted in a review (Manuscript D).

To analyze the functional similarities and differences between the bRC and PSII RC proteins on the sequential level, profile Hidden Markov Models (pHMMs) were constructed (Manuscript

E). The pHMM construction was a prerequisite to examine conservation of the proton transfer key residues in the bRC. In bRC and PSII RC proteins, the mechanisms of directing the electron transfer via the A-branch and of midpoint potential tuning of the special pair was shown to be different in contrast to pigment and cofactor binding. The pHMM alignment was further used to investigate the conservation of residues involved in the water splitting mechanism in PSII RC (Manuscript F). The conservation study (using the built pHMMs) of the experimentally determined key residues of proton transfer to Q_B showed, that putative proton entry points are not conserved, but inner proton transfer key residues are conserved or are only exchanged to functionally similar residues (Manuscript G). By analysis of the hydrogen bonded network from Q_B to the cytoplasm for the bRC of two different purple bacteria (*Rb. sphaeroides* and *Rsp. viridis*), it was determined, that the overall organization is similar, but that the proton entry points differ. Together with the conservation study, these results showed, that the proton transfer from cytoplasm to Q_B is most likely not organized in distinct proton transfer pathways but in a huge network without distinct pathways, in a proton sponge. The work described in manuscript A to G helped to get a more detailed understanding of the molecular details of the reaction in the Q_B site of the bRC and PSII RC proteins.

Manuscript A: PH Modulates the Quinone Position in the Photosynthetic Reaction Center from *Rhodobacter sphaeroides* in the Neutral and Charge Separated States.

The focus of this work was to examine the dependence of the CoQ position in the Q_B site on the pH and the illumination state using both crystallographic experiments and theoretical investigations. The study resulted in the crystal structures with the highest resolution (of 1.87 /Å) available today, revealing a new orientation of CoQ in the distal position of the Q_B binding site. For this orientation no 180° rotation of Q_B is needed compared to the proximal binding position. Due to the high resolution of the structure, it was possible to propose two extended proton transfer pathways to Q_B . By crystallization of the bRC at different pH values, a pH dependence of the Q_B population in the proximal position of the Q_B site was observed in the dark-adapted and illuminated state of the bRC. I could show by quantum chemical and CE/MC calculations, that the new and the old orientation of Q_B in the distal position are almost equal in energy, slightly favoring the new orientation and thus, the 180° rotation of Q_B is not needed. By CE/MC calculations, I was also able to reproduce the measured pH-dependence of the Q_B populations. Also the proton uptake during the first illumination was resembled by the calculations. I identified that the protonation states of GluL212 and AspL213 to be the main factors determining the pH dependence of the populations in the dark-adapted and illuminated states.

Manuscript B: The Role of AspL213 for Stabilizing Semiquinone Binding to the Photosynthetic Reaction Center. The tight binding of the semiquinone $Q\cdot^-$ in the Q_B binding site was investigated by a CE/MC approach using the high resolution structures obtained earlier (Manuscript A). To obtain relative binding curves of the semiquinone, its concentration was included into the calculations by adjusting the protonation state energy to a protonation and binding state energy. I could reproduce the experimentally and theoretically obtained populations of the distal and proximal positions (Manuscript A). Furthermore I identified that the protonation state of GluL212 and AspL213 is coupled to the binding of the semiquinone, meaning that a protonated GluL212 and AspL213 is required to keep the semiquinone bound. Moreover, a protonated AspL213 keeps the semiquinone bound in the proximal, reactive position in the Q_B binding site.

Manuscript C: Calculation of Protonation and Reduction Properties of Different Quinone Systems. The different possible protonation and reduction states of the benzoquinone and the duroquinone system were investigated by quantum chemical calculations. This is not trivial, since the radical states (e. g. the semiquinone state $Q\cdot^-$) and the highly charged states (e. g. the quinol Q^{2-}) are difficult to determine quantum mechanically. Using benzoquinone as a test system, a good optimization and charge determination procedure was found. Therefore I made quantum chemical calculations with different basis sets and used wave-function and density functional theory as well as different charge determination methods for each of the possible protonation and reduction states of the benzoquinone. For validation the calculated reaction energies were compared to experimental values. I used the best protocols for the benzoquinone system to calculate the energetics for duroquinone obtaining also here very good agreement with experiments. The protocol developed here to treat aromatic radicals can be further used for other systems like ubiquinone for example.

Manuscript D: Investigating the Mechanisms of Photosynthetic Proteins Using Continuum Electrostatics. The suitability of continuum electrostatics to gain more insight in the detailed mechanisms and functions of proteins is highlighted in this review. Photosynthetic proteins are used to stress the power of this method to promote understanding of the association of proteins, of the tuning of absorption spectra, of the coupling of electron transfer and proton transfer, of the effect of membrane potentials on the energetics of membrane proteins, and of the kinetics of charge transfer reactions. As an example for coupling of proton and electron transfer I described how continuum electrostatics helped to improve our current view on for the Q_B site of the bRC.

Manuscript E: Profile Hidden Markov Models for Analyzing Similarities and Dissimilarities in the Bacterial Reaction Center and Photosystem II. I investigated similarities and dissimilarities in molecular mechanisms of the bRC and the PSII RC proteins based on their sequences. Such an investigation is nontrivial, because a normal progressive multiple sequence alignment (MSA) fails due to the very low sequence identity shared by the subunits. Such a low sequence identity is also observed comparing random sequences. Using a pHMM approach it was possible to obtain a proper alignment. The analysis of the performed alignment revealed, that cofactor binding is conserved in both, the PSII RC and its evolutionary ancestor, the bRC. Experimentally it was shown, that the addition or the removal of a hydrogen bond to the special pair leads to an alteration of its redox properties in the bRC of *Rb.sphaeroides*. Based on the MSA, I was able to show, that only two different hydrogen bonding pattern of the special pair exist in all bacterial species. The pattern is determined by the existence of the C subunit in bacteria and leads to the same midpoint potential of the special pair in all bacterial species. However, hydrogen bonds to the pigments and the local environment of Q_A and Q_B are not conserved in the bRC and PSII RC proteins. The mechanism of midpoint potential tuning is different in bRC and PSII RC proteins. Previously determined residues favoring the charge transfer via the A-branch are conserved in bRC proteins but differ in PSII RC proteins. This means, that regulation of the charge transfer via the A-branch differs in bRC and PSII RC proteins. Thus, although the bRC and PSII RC proteins catalyze the same biological reaction, they differ in details such as midpoint potential of the special pair, or the regulation of the A-branch electron transfer.

Manuscript F: Conservation Analysis of Functional Important Residues of the Oxygen Evolving Mechanism Located in the D1 Subunit of Photosystem II. The previously built pHMM model for the L/D1 subunit (Manuscript E) was used to investigate functionally important residues for oxygen evolution in PSII RC proteins. This reaction takes place at the water oxidizing complex formed by a manganese cluster and a nearby tyrosine. Although the exact coordination of the manganese cluster is not yet determined, several residues were found to participate in coordination and in the water oxidizing mechanism. I showed, that even if most of the residues are strictly conserved, for some residues exchange to aminoacids known to prevent oxygen formation in mutants of the cyanobacteria *Syncheocystis* is observed.

Manuscript G: Proton Transfer Pathways in Photosynthetic Reaction Centers Analyzed by profile Hidden Markov Models and Network Calculations. The proton transfer from cytoplasm to Q_B was investigated combining two different approaches: a conservation analysis of experimentally determined key residues of proton transfer to Q_B and an analysis of the hydrogen bonded network of the bRC of two different bacterial species (*Rb. sphaeroides* and

B. viridis).

I made the conservation analysis using the pHMM models built before for the L/D1 and for the M/D2 subunit (Manuscript E) and a newly constructed pHMM for the H subunit showing that the proton entry points are not conserved throughout the species. However, the inner key residues are conserved or only exchanged to other aminoacids with the similar properties. The bRC proteins of *Rb. sphaeroides* and of *B. viridis* have a huge network connecting the cytoplasm with Q_B. Both networks include the inner key residues but differ in the proton entry points. Thus, the results of the two independent approaches (conservation analysis and hydrogen bonded network calculation) make the idea of distinct pathways rather unlikely. Therefore, it seems more likely, that the proton transfer is mediated through a huge network without distinct pathways, a proton sponge.

Bibliography

- [1] Lockhart PJ, Larkum AWD, Steel MA, Waddell PJ, and Penny D (1996): Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence alignment. *Proc. Natl. Acad. Sci. USA* 93, 1930–1934
- [2] Hillier W and Babcock GT (2001): Photosynthetic reaction centers. *Plant Physiol.* 125, 33–37
- [3] Fajer J (2004): Chlorophyll chemistry before and after crystals of photosynthetic reaction centers. *Photosyn. Res.* 80, 165–172
- [4] Tavano CL and Donohue TJ (2006): Development of the bacterial photosynthetic apparatus. *Curr. Opt. Microbiol.* 9, 625–631
- [5] Verméglio A and Joliot P (2002): Supramolecular organisation of the photosynthetic chain in anoxygenic bacteria. *Biochim. Biophys. Acta* 1555, 60–64
- [6] Joliot P, Joliot A, and Verméglio A (2005): Fast oxidation of the primary electron acceptor under anaerobic conditions requires the organization of the photosynthetic chain of *Rhodobacter sphaeroides* in supercomplexes. *Biochim. Biophys. Acta* 1706, 204–214
- [7] Joliot P, Verméglio A, and Joliot A (1989): Evidence for supercomplexes between reaction centers, cytochrome *c2* and cytochrome *bc1* complex in *Rhodobacter sphaeroides* whole cells. *Biochim. Biophys. Acta* 975, 336–345
- [8] Qian P, Hunter CN, and Bullough PA (2005): The 8.5 Å Projection Structure of the Core RC-LH1-PufX Dimer of *Rhodobacter sphaeroides*. *J. Mol. Biol.* 349, 948–960
- [9] Roszak AW, Howard TD, S J, Gardiner AT, Law CJ, Isaacs NW, and Cogdell RJ (2003): Crystal Structure of the RC-LH1 Core Complex from *Rhodopseudomonas palustris*. *Science* 302, 1969 – 1972
- [10] Scheuring S, Seguin J, Marco S, Lévy D, Robert B, and Rigaud JL (2003): Nanodissection and high-resolution imaging of the *Rhodopseudomonas viridis* photosynthetic core complex in native membranes by AFM. *Proc. Natl. Acad. Sci. USA* 100, 1690–1693
- [11] Scheuring S, Rigaud JL, and Sturgis JN (2004): Variable LH2 stoichiometry and core clustering in native membranes of *Rhodospirillum rubrum*. *EMBO J.* 23, 4127–4133

- [12] Scheuring S, Sturgis JN, Prima V, Bernadac A, Lévy D, and Rigaud JL (2004): Watching the photosynthetic apparatus in native membranes. *Proc. Natl. Amer. Soc. USA* 101, 11293–11297
- [13] Goncalves RP, Bernadac A, Sturgis JN, and Scheuring S (2005): Architecture of the native photosynthetic apparatus of *Phaeospirillum molischianum*. *J. Struc. Biol.* 152, 221–228
- [14] Scheuring S, Lévy D, and Rigaud JL (2005): Watching the components of photosynthetic bacterial membranes and their in situ organization by atomic force microscopy. *Biochim. Biophys. Acta* 1712, 109–127
- [15] Kargul J and Barber J (2007): Photosynthetic acclimation: structural reorganization of light harvesting antenna—role of redox-dependent phosphorylation of major and minor chlorophyll a/b binding proteins. *FEBS J.* 275, 1056–1068
- [16] Ferreira KN, Iverson TM, Maghlaoui K, Barber J, and Iwata S (2004): Architecture of the photosynthetic oxygen-evolving center. *Science* 303, 1831–1838
- [17] Amunts A, Drory O, and Nelson N (2007): The structure of a plant photosystem I super-complex at 3.4 Å resolution. *Nature* 447, 58–63
- [18] Renger G (2001): Photosynthetic water oxidation to molecular oxygen: apparatus and mechanism. *Biochim. Biophys. Acta* 1503, 210–228
- [19] Kern J and Renger G (2007): Photosystem II: structure and mechanism of the water:plastoquinone oxidoreductase. *Photosyn. Res.* 94, 183–202
- [20] Breyton C (2000): The cytochrome b_6f complex: structural studies and comparison with the bc_1 complex. *Biochim. Biophys. Acta* 1459, 467–474
- [21] Allen JP and Williams JC (1998): Photosynthetic reaction centers. *FEBS Lett.* 438, 5–9
- [22] Allen JF and Martin W (2007): Evolutionary biology. Out of thin air. *Nature* 445, 610–612
- [23] Vermass WFJ (1994): Evolution of heliobacteria: implications for photosynthetic reaction center complexes. *Photosyn. Res.* 41, 285–294
- [24] Olson JM and Blankenship RE (2004): Thinking about the evolution of photosynthesis. *Photosyn. Res.* 80, 373–386
- [25] Xiong J and Bauer CE (2002): Complex evolution of photosynthesis. *Annu. Rev. Plant. Biol.* 53, 503–521

- [26] Blankenship RE (1992): Origin and early evolution of photosynthesis. *Photosyn. Res.* 33, 91–111
- [27] Nitschke W and Rutherford AW (1991): Photosynthetic reaction centres: variations on a common structural theme? *TIBS* 16, 241–245
- [28] Blankenship RE (1994): Protein structure, electron transfer and evolution of prokaryotic photosynthetic reaction centers. *Antonie Van Leeuwenhoek* 65, 311–329
- [29] Chen IP, Mathis P, Köpke J, and Michel H (2000): Uphill electron transfer in the tetraheme cytochrome subunit of the *Rhodospseudomonas viridis* photosynthetic reaction center: evidence from site-directed mutagenesis. *Biochemistry* 39, 3592–3602
- [30] Becker T, Ullmann RT, and Ullmann GM (2007): Simulation of the electron transfer between the tetraheme subunit and the special pair of the photosynthetic reaction center using a microstate description. *J. Phys. Chem. B* 111, 2957–2968
- [31] Loll B, Kern J, Saenger W, Zouni A, and Biesiadka J (2005): Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. *Nature* 438, 1040–1044
- [32] Tracewell CA, Vrettos JS, Bautista JA, and Frank HA (2001): Carotenoid photooxidation in photosystem II. *Arch. Biochem. Biophys.* 385, 61–69
- [33] Humphrey W, Dalke A, and Schulten K (1996): VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38
- [34] Köpke J, Krammer EM, Klinge AR, Sebban P, Ullmann GM, and Fritzsche G (2007): PH Modulates the Quinone Position in the Photosynthetic Reaction Center from *Rhodobacter Sphaeroides* in the Neutral and Charge Separated States. *J. Mol. Biol.* 371, 396–409
- [35] Heathcote P, Fyfe PK, and Jones MR (2002): Reaction centres: the structure and evolution of biological solar power. *Trends Biochem. Sci.* 27, 79–87
- [36] Michel H and Deisenhofer J (1988): Relevance of the photosynthetic reaction center from purple bacteria to the structure of photosystem II. *Biochemistry* 27, 1–6
- [37] Rost B (1997): Protein structures sustain evolutionary drift. *Folding Des.* 2, S19–S24
- [38] Breton J, Wakeham MC, Fyfe PK, Jones MR, and Navedryk E (2004): Characterization of the bonding interactions of Q_B upon photoreduction via A-branch or B-branch electron transfer in mutant reaction centers from *Rhodobacter sphaeroides*. *Biochim. Biophys. Acta* 1656, 127–138

- [39] Wakeham MC, Breton J, Navedryk E, and Jones MR (2004): Formation of a semiquinone at the Q_B site by A- or B-branch electron transfer in the reaction center from *Rhodobacter sphaeroides*. *Biochemistry* 43, 4755–4763
- [40] Paddock ML, Chang C, Xu Q, Abresch EC, Axelroth HL, Feher G, and Okamura MY (2005): Quinone (Q_B) reduction by B-branch electron transfer in mutant bacterial reaction centers from *Rhodobacter sphaeroides*: quantum efficiency and X-ray structure. *Biochemistry* 44, 6920–6928
- [41] Laible PD, Kirmaier C, Udawatte CSM, Hofman SJ, Holten D, and Hanson DK (2003): Quinone reduction via secondary B-branch electron transfer in mutant bacterial reaction centers. *Biochemistry* 42, 1718–1730
- [42] McAuley KE, Fyfe PK, Ridge JP, Cogdell RJ, Isaacs NW, and Jones MR (2000): Ubiquinone binding, ubiquinone exclusion, and detailed cofactor conformation in a mutant bacterial reaction center. *Biochemistry* 39, 15032–15043
- [43] Ridge JP, van Brederode ME, Goodwin MG, van Grondelle R, and Jones MR (1999): Mutations that modify or exclude binding of Q_A ubiquinone and carotenoid in the reaction center from *Rhodobacter sphaeroides*. *Photosyn. Res.* 59, 9–26
- [44] Debus RJ, Feher G, and Okamura MY (1986): Iron-depleted reaction centers from *Rhodopseudomonas sphaeroides* R-26.1: characterization and reconstitution with Fe^{2+} , Mn^{2+} , Co^{2+} , Cu^{2+} and Zn^{2+} . *Biochemistry* 25, 2276–2287
- [45] Hermes S, Bremm O, Garczarek F, Derrien V, Liebisch P, Sebban P, Gewert K, and Haumann M (2006): A time-resolved iron-specific X-ray absorption experiment yields no evidence for an $Fe^{2+} \rightarrow Fe^{3+}$ electron transfer in the photosynthetic reaction center. *Biochemistry* 45, 353–359
- [46] Laible PD, Chynwat V, Thurnauer MC, Schiffer M, Hanson DK, and Frank HA (1998): Protein modifications affecting triplet energy transfer in bacterial photosynthetic reaction centers. *Biophys. J.* 74, 2623–2637
- [47] Zouni A, Witt HT, Kern J, Fromme P, Krauss N, Saenger W, and Orth P (2001): Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature* 409, 739–743
- [48] Groot ML, Pawlowicz NP, van Wilderen LJGW, Breton J, van Stokkum IHM, and van Grondelle R (2005): Initial electron donor and acceptor in isolated photosystem II

- reaction centers identified with femtosecond mid-IR spectroscopy. *Proc. Natl. Acc. Soc.* 102, 13087–13092
- [49] Holzwarth AR, Müller MG, Reus M, Nowaczyk M, Sander J, and Rögner M (2006): Kinetics and mechanism of electron transfer in intact photosystem II and in the isolated reaction center: pheophytin is the primary electron acceptor. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6895–6900
- [50] Renger T and Schlodder E (2008): The Primary Electron Donor of Photosystem II of the Cyanobacterium *Acaryochloris marina* Is a Chlorophyll *d* and the Water Oxidation Is Driven by a Chlorophyll *a*/Chlorophyll *d* Heterodimer. *J. Phys. Chem. B* 112, 7351–7354
- [51] Raszewski G, Saenger W, and Renger T (2005): Theory of Optical Spectra of Photosystem II Reaction Centers: Location of the Triplet State and the Identity of the Primary Electron Donor. *Biophys. J.* 88, 986–998
- [52] Ruffle SV, Wand J, Johnston HG, Gustafson TL, Hutschison RS, Minagawa J, Crofts A, and Sayre RT (2001): Photosystem II peripheral accessory chlorophyll mutants in *chlamydomonas reinhardtii*. Biochemical characterization and sensitivity to photo-inhibition. *Plant Physiol.* 127, 633–644
- [53] Johnston HG, Wang J, Ruffle SV, Sayre RT, and Gustafson TL (2000): Fluorescence decay kinetics of wild type and D2-H117N mutant photosystem II reaction centers isolated from *Chlamydomonas reinhardtii*. *J. Phys. Chem. B* 104, 4777–4781
- [54] Nugent JHA, Rich AM, and Evans MCW (2001): Photosynthetic water oxidation: towards a mechanism. *Biochim. Biophys. Acta* 1503, 138–146
- [55] Tommos C and Babcock GT (2000): Proton and hydrogen currents in photosynthetic water oxidation. *Biochim. Biophys. Acta* 1458, 199–219
- [56] Yano J, Sauer K, Latimer MJ, Pushkar Y, Biesiadka J, Loll B, Saenger W, Messinger J, Zouni A, and Yachandra VK (2006): Where water is oxidized to dioxygen: structure of the photosynthetic Mn₄Ca cluster. *Science* 321, 821–825
- [57] Chu HA, Hillier W, and Debus RJ (2004): Evidence that the C-terminus of the D1-polypeptide of photosystem II is ligated to the manganese ion that undergoes oxidation during the S₁ to S₂ transition: an isotope-edited FTIR study. *Biochemistry* 43, 3152–3166

- [58] Debus RJ, Campbell KA, Pham DP, Hays AMA, and Britt RD (2000): Glutamate 189 of the D1 polypeptide modulates the magnetic and redox properties of the manganese cluster and Tyrosine Y_Z in photosystem II. *Biochemistry* 39, 6275–6287
- [59] H-A Chu APN and Debus RJ (1995): Amino acid residues that influence the binding of manganese or calcium to photosystem II. 1. The lumenal interhelical domains of the D1 polypeptide. *Biochemistry* 34, 5839–5858
- [60] Nixon PJ and Diner BA (1992): Aspartate 170 of the Photosystem II Reaction Center Polypeptide D1 Is Involved in the Assembly of the Oxygen-Evolving Manganese Cluster? *Biochemistry* 31, 942–948
- [61] Svensson B, Tiede DM, Nelson DR, and Barry BA (2004): Structural studies of the manganese stabilizing subunit in photosystem II. *Biophys. J.* 86, 1807–1812
- [62] Wyman AJ, Popelkova H, and Yocum CF (2008): Site-directed mutagenesis of conserved C-Terminal tyrosine and tryptophan residues of PsbO, the Photosystem II manganese-stabilizing protein, alters its activity and fluorescence properties. *Biochemistry* 47, 6490–6498
- [63] Deisenhofer J, Epp O, Miki K, Huber R, and Michel H (1985): Structure of the protein subunits in the the photosynthetic reaction center from *Rhodospseudomonas viridis* at 3 Å resolution. *Nature* 318, 618–624
- [64] Allen JP, Feher G, Yeates TO, Komiya H, and Rees DC (1987): Structure of the reaction center from *Rhodobacter sphaeroides* R-26: the cofactors. *Proc. Natl. Acad. Sci. USA* 84, 5730–5734
- [65] Nogi T, Fathir I, Kobayashi M, Nozawa T, and Miki K (2000): Crystal structures of photosynthetic reaction center and high-potential iron-sulfur protein from *Thermochromatium tepidum*: thermostability and electron transfer. *Proc. Natl. Acad. Sci. USA* 97, 13561–13566
- [66] RSCB Protein Data Base. <http://www.rcsb.org/pdb/>
- [67] Bernstein FC, Koetzle TF, Williams GJB, Meyer Jr EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, and Tasumi M (1977): The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112, 535–542
- [68] Wang X, Cao J, Maròti P, Stilz HU, Finkle U, Lauterwasser C, Zinth W, Oesterheit D, Govindjee, and Wraight CA (1992): Is bicarbonate in Photosystem II the equivalent of the

- glutamate ligand to the iron atom in bacterial reaction centers? *Biophys. Biochim. Acta* 1100, 1–8
- [69] Fritzsche G, Köpke J, Diem R, Kuglstatter A, and Baciou L (2002): Charge separation induces conformational changes in the photosynthetic reaction centre of purple bacteria. *Acta Cryst.* 58, 1660–1663
- [70] Stowell MHV, McPhillips TM, Rees DC, Soltis SM, Abresch E, and Feher G (1997): Light-induced structural changes in photosynthetic reaction center: Implications for mechanism of electron-proton transfer. *Science* 276, 812–816
- [71] Taly A, Sebban P, Smith JC, and Ullmann GM (2003): The position of Q_B in the photosynthetic reaction center depends on pH: a theoretical analysis of the proton uptake upon Q_B reduction. *Biophys. J.* 84, 2090–2098
- [72] Kuglstatter A, Ermler U, Michel H, Baciou L, and Fritzsche G (2001): X-ray structure analysis of photosynthetic reaction center variants from *Rhodobacter sphaeroides*: structural changes induced by point mutations at position L209 modulate electron and proton transfer. *Biochemistry* 40, 4253–4260
- [73] Navedryk E, Breton J, Sebban P, and Baciou L (2003): Quinone (Q_B) binding site and protein structural changes in photosynthetic reaction center mutants at Pro-L209 revealed by vibrational spectroscopy. *Biochemistry* 42, 5819–5827
- [74] Baxter RHG, Seagle BL, Ponomarenko N, and Norris JR (2005): Cryogenic structure of the photosynthetic reaction center of *Blastochloris viridis* in the light and dark. *Acta Cryst.* D61, 605–612
- [75] Zachariae U and Lancaster CRD (2001): Proton uptake associated with the reduction of the primary quinone Q_A influences the binding site of the secondary quinone Q_B in *Rhodospseudomonas viridis* photosynthetic reaction centers. *Biochim. Biophys. Acta* 1505, 280–290
- [76] Breton J, Boullais C, Mioskowski C, Sebban P, Baciou L, and Navedryk E (2002): Vibrational spectroscopy favors a unique Q_B binding site at the proximal position in wild-type reaction centers and in the Pro-L209→ Tyr mutant from *Rhodobacter sphaeroides*. *Biochemistry* 41, 12921–12927
- [77] Pokkuluri PR, Laible PH, Crawford AE, Mayfield JF, Yousef MA, Ginsell SL, Hanson DK, and Schiffer M (2004): Temperature and cryoprotectant influence secondary quinone binding position in bacterial reaction centers. *FEBS Lett.* 570, 171–174

- [78] Takahashi E, Wells TA, and Wraight CA (2001): Protein control of the redox potential of the primary quinone acceptor in reaction centers from *Rhodobacter sphaeroides*. *Biochemistry* 40, 1020–1028
- [79] Remy A, Boers RB, Egorova-Zachernyuk T, Gast P, Lugtenburg J, and Gerwert K (2003): Does different orientation of the methoxy groups of ubiquinone-10 in the reaction centre of *Rhodobacter sphaeroides* cause different binding at Q_A and Q_B? *Eur. J. Biochem.* 270, 3603–3609
- [80] Nonella M, Mathias G, Eichinger M, and Tavan P (2003): Structures and vibrational frequencies of the quinones in *Rb. sphaeroides* derived by a combined density functional/molecular mechanics approach. *J. Phys. Chem. B* 107, 316–322
- [81] Flores M, Isaacson R, Abresch E, Calvo R, Lubitz W, and Feher G (2006): Protein-cofactor interactions in bacterial reaction centers from *Rhodobacter sphaeroides* R-26: II. Geometry of the hydrogen bonds to the primary quinone Q_A^{•-} by ¹H and ²H ENDOR spectroscopy. *Biophys. J.*
- [82] Brudler R, de Groot HJM, van Liemt WBS, Steggerda WF, Esmeijer R, Gast P, Hoff AJ, Lugtenburg J, and Gerwert K (1994): Asymmetric binding of the 1- and 4-C=O groups of Q_A in *Rhodobacter sphaeroides* R26 reaction centres monitored by Fourier transform infra-red spectroscopy using site-specific isotopically labelled ubiquinone-10. *EMBO J.* 13, 5523–5530
- [83] Flores M, Isaacson R, Abresch E, Calvo R, Lubitz W, and Feher G (2007): Protein-cofactor interactions in bacterial reaction centers from *Rhodobacter sphaeroides* R-26: II. Geometry of the hydrogen bonds to the primary quinone Q_A^{•-} by ¹H and ²H ENDOR spectroscopy. *Biophys. J.* 92, 671–682
- [84] Wells TA, Takahashi E, and Wraight CA (2003): Primary quinone (Q_A) binding site of bacterial photosynthetic reaction centers: mutations at residue M265 probed by FTIR spectroscopy. *Biochemistry* 42, 4064–4074
- [85] Cheap H, Tandori J, Derrien V, Benoit M, de Oliveira P, Köpke J, Lavergne J, Maróti P, and Sebban P (2007): Evidence for delocalized anticooperative flash induced proton binding as revealed by mutants at the M266His iron ligand in bacterial reaction centers. *Biochemistry* 46, 4510–4521
- [86] Paddock ML, Feher G, and Okamura MY (2003): Proton transfer pathways and mechanism in bacterial reaction centers. *FEBS Lett.* 555, 45–50

- [87] Okamura MY, Paddock ML, Graige MS, and Feher G (2000): Proton and electron transfer in bacterial reaction center. *Biochim. Biophys. Acta* 145, 148–163
- [88] Hanson DK and Schiffer M (1998): Symmetry-related mutants in the quinone binding sites of the bacterial reaction center - the effects of changes in charge distribution. *Photosynth. Res.* 55, 275–280
- [89] Laible PD, Zhang Y, Morris AL, Snyder SW, Ainsworth C, Greenfield SR, Wasielewski MR, Parot P, Schoepp B, Schiffer M, Hanson DK, and Thurnauer MC (1997): Spectroscopic characterization of quinone-site mutants of the bacterial photosynthetic reaction center. *Photosynth. Res.* 52, 99–103
- [90] Pokkuluri PR, Laible PD, Deng YL, Wong TN, Hanson DK, and Schiffer M (2002): The structure of a mutant photosynthetic reaction center shows unexpected changes in main chain orientations and quinone position. *Biochemistry* 41, 5998–6007
- [91] Saquin-Mora S, Sebban P, Derrien V, Frick B, Lavery R, and Alba-Simionesco C (2007): Probing the flexibility of the bacterial reaction center: the wild-type protein is more rigid than two site-specific mutants. *Biochemistry* 46, 14960–14968
- [92] Hanson DK, Baciou L, Tiede DM, Nance SL, Schiffer M, and Sebban P (1992): In bacterial reaction centers protons can diffuse to the secondary quinone by alternative pathways. *Biochim. Biophys. Acta* 1102, 260–265
- [93] Ädelroth P, Paddock ML, Sagle LB, Feher G, and Okamura MY (2000): Identification of the proton pathway in bacterial reaction centers: both protons associated with reduction of Q_B to Q_BH_2 share a common entry point. *Proc. Natl. Acad. Sci. USA* 97, 13086–13091
- [94] Tandori J, Maróti P, Alexov E, Sebban P, and Baciou L (2002): Key role of proline L209 in connecting the distant quinone pockets in the reaction center of *Rhodobacter sphaeroides*. *Proc. Natl. Acad. Sci. USA* 99(10), 6702–6706
- [95] Miksovská J, Schiffer M, Hanson DK, and Sebban P (1999): Proton uptake by bacterial reaction centers: the protein complex responds in a similar manner to the reduction of either quinone acceptor. *Proc. Natl. Acad. Sci. USA* 96, 14348–14353
- [96] Tandori J, Miksovská J, Valerio-Lepiniec M, Schiffer M, Maróti P, Hanson DK, and Sebban P (2002): Proton uptake of *Rhodobacter sphaeroides* reaction center mutants modified in the primary quinone environment. *Photochem. Photobiol.* 75, 126–133

- [97] Mezzetti A and Leibl W (2005): Investigation of ubiquinol formation in isolated photosynthetic reaction centers by rapid-scan Fourier transfer IR spectroscopy. *Eur. Biophys. J.* 34, 921–936
- [98] Nabedryk E, Breton J, Hienerwadel R, Fogel C, Mäntele W, Paddock ML, and Okamura MY (1995): Fourier transform infrared difference spectroscopy of secondary quinone acceptor photoreduction in proton transfer mutants of *Rhodobacter sphaeroides*. *Biochemistry* 34, 14722–14732
- [99] Ädelroth P, Paddock ML, Tehrani A, Beatty JT, Feher G, and Okamura MY (2001): Identification of the proton pathway in bacterial reaction centers: decrease of proton transfer rate by mutation of surface histidines at H126 and H128 and chemical rescue by imidazole identifies the initial proton donors. *Biochemistry* 40, 14538–14546
- [100] Paddock ML, Ädelroth P, Feher G, Okamura MY, and Beatty JT (2002): Determination of proton transfer rates by chemical rescue: Application to bacterial reaction centers. *Biochemistry* 41, 14716–14725
- [101] Hienerwadel R, Grzybek S, Fogel C, Kreutz W, Okamura MY, Paddock ML, Breton J, Nabedryk E, and Mäntele W (1995): Protonation of Glu L212 following Q_B^- formation in the photosynthetic reaction center of *Rhodobacter sphaeroides*: evidence from time-resolved infrared spectroscopy. *Biochemistry* 34, 2832–2843
- [102] Miksovská J, Maróti P, Tandori J, Schiffer M, Hanson DK, and Sebban P (1996): Distant electrostatic interactions modulate the free energy level of Q_A^- in the photosynthetic reaction center. *Biochemistry* 35, 15411–15417
- [103] Hermes S, Stacknik JM, Onidas D, Remy A, Hofmann E, and Gewert K (2006): Proton uptake in the reaction center mutant L210DN from *Rhodobacter sphaeroides* via protonated water molecules. *Biochemistry* 45, 13741–13749
- [104] Remy A and Gewert K (2003): Coupling of light-induced electron transfer from proton uptake in photosynthesis. *Nat. Struct. Biol.* 10, 637–644
- [105] Breton J (2007): Steady-state FTIR spectra of the photoreduction of Q_A and Q_B in *Rhodobacter sphaeroides* reaction centers provide evidence against the presence of a proposed transient electron acceptor X between the two quinones. *Biochemistry* 46, 4459–4465
- [106] Graige MS, Paddock ML, Bruce JM, Feher G, and Okamura MY (1996): Mechanism of proton-coupled electron transfer for quinone (Q_B) reduction in reaction centers of *Rb. sphaeroides*. *J. Am. Chem. Soc.* 118, 9005–9016

- [107] Graige MS, Feher G, and Okamura MY (1998): Conformational gating of the electron transfer reaction $Q_A^- \cdot Q_B \rightarrow Q_A Q_B^-$ in bacterial reaction centers of *Rhodobacter sphaeroides* determined by a driving force assay. *Proc. Natl. Acad. Sci. USA* 95, 11679–11684
- [108] Xu Q, Baciou L, Sebban P, and Gunner MR (2002): Exploring the energy landscape for Q_A^- to Q_B electron transfer in bacterial photosynthetic reaction centers: effect of substrate position and tail length on the conformational gating step. *Biochemistry* 41, 10021–10025
- [109] McComb JC, Stein RR, and Wraight CA (1990): Investigation on the influence of head-group substitution and isoprene side-chain length in the function of primary and secondary quinones of bacterial reaction centers. *Biochim. Biophys. Acta* 1015, 156–171
- [110] Nabadryk E, Paddock ML, Okamura MY, and Breton J (2005): An isotope-edited FTIR investigation of the role of Ser-L223 in binding quinone (Q_B) and semiquinone (Q_B^-) in the reaction center of *Rhodobacter sphaeroides*. *Biochemistry* 44, 14519–14527
- [111] Paddock ML, Feher G, and Okamura MY (2000): Identification of the proton pathway in bacterial reaction centers: replacement of Asp-M17 and Asp-L210 with Asn reduces the proton transfer rate in the presence of Cd^{2+} . *Proc. Natl. Acad. Sci. USA* 97, 1548–1553
- [112] Graige MS, Paddock ML, Feher G, and Okamura MY (1999): Observation of the protonated semiquinone intermediate in isolated reaction centers from *Rhodobacter sphaeroides*: Implications for the mechanism of electron and proton transfer in proteins. *Biochemistry* 38, 11465–11473
- [113] Lavergne J, Matthews C, and Ginet N (1999): Electron and proton transfer on the acceptor side of reaction center in chromatophores of *Rhodobacter capsulatus*: evidence for direct protonation of the semiquinone state of Q_B . *Biochemistry* 38, 4542–4552
- [114] Sebban P, Maróti P, and Hanson DK (1995): Electron and proton transfer to the quinones in bacterial photosynthetic reaction centers: insight from combined approaches of molecular genetics and biophysics. *Biochemistry* 77, 677–694
- [115] Takahashi E and Wraight CA (1992): Proton and electron transfer in the acceptor quinone complex of *Rhodobacter sphaeroides* reaction centers: characterization of site-directed mutants of the two ionizable residues, Glu^{L212} and Asp^{L213}, in the Q_B binding site. *Biochemistry* 31, 855–866
- [116] Paddock ML, Rongley SH, Feher G, and Okamura MY (1989): Pathway of proton transfer in bacterial reaction centers: Replacement of glutamic acid 212 in the L subunit by glutamine inhibits quinone (secondary acceptor) turnover. *Proc. Natl. Acad. Sci. USA* 86, 6602–6606

- [117] Ädelroth P and Brzezinski P (2004): Surface-mediated proton-transfer reactions in membrane-bound proteins. *Biochim. Biophys. Acta* 1655, 102–115
- [118] McPherson PH, Okamura MY, and Feher G (1990): Electron transfer from the reaction center of *Rhodobacter sphaeroides* to the quinone pool: doubly reduced Q_B leaves the reaction center. *Biochim. Biophys. Acta* 1016, 289–292
- [119] Lancaster CRD and Michel H (1997): The coupling of light-induced electron transfer and proton uptake as derived from crystal structures of reaction centers from *Rhodobacter viridis* modified at the binding site of the secondary quinone, Q_B . *Structure* 5, 1339–1359
- [120] Paddock ML, McPherson PH, Feher G, and Okamura MY (1990): Pathway of proton transfer in bacterial reaction centers: replacement of serine-L223 by alanine inhibits electron and proton transfers associated with reduction of quinone to dihydroquinone. *Proc. Natl. Acad. Sci. USA* 87, 6803–6807
- [121] Abresch EC, Paddock ML, Stowell MHB, McPhillips TM, Axelrod HL, Soltis SM, Okamura MY, and Feher G (1998): Identification of proton transfer pathways in the X-ray crystal structure of the bacterial reaction center from *Rhodobacter sphaeroides*. *Photosynth. Res.* 55, 119–125
- [122] Paddock ML, Ädelroth P, Chang C, Abresch EC, Feher G, and Okamura MY (2001): Identification of the proton pathway in bacterial reaction centers: cooperation between Asp-M17 and Asp-L210 facilitates proton transfer to the secondary quinone (Q_B). *Biochemistry* 40, 6893–6902
- [123] Paddock ML, Graige MS, Feher G, and Okamura MY (1999): Identification of the proton pathway in bacterial reaction centers: inhibition of proton transfer by binding of Zn^{2+} or Cd^{2+} . *Proc. Natl. Acad. Sci. USA* 99, 6183–6188
- [124] Navedryk E, Breton J, Okamura MY, and Paddock ML (2001): Simultaneous replacement of Asp-L210 and Asp-M17 with Asn increases proton uptake by Glu-L212 upon first electron transfer to Q_B in reaction centers from *Rhodobacter sphaeroides*. *Biochemistry* 40, 13826–13832
- [125] Xu Q, Axelrod HL, Abresch EC, Paddock ML, Okamura MY, and Feher G (2004): X-ray structure determination of three mutants of the bacterial photosynthetic reaction centers from *Rb. sphaeroides*: altered proton transfer pathways. *Structure* 12, 703–716

- [126] Utschig LM, Poluektov O, Schlesselman SL, Thurnauer MC, and Tiede DM (2001): Cu²⁺ site in photosynthetic bacterial reaction centres from *Rhodobacter sphaeroides*, *Rhodobacter capsulatus* and *Rhodopseudomonas viridis*. *Biochemistry* 40, 6132–6141
- [127] Axelrod HL, Abresch EC, Paddock ML, Okamura MY, and Feher G (2000): Determination of the binding sites of the proton transfer inhibitors Cd²⁺ and Zn²⁺ in bacterial reaction centers. *Proc. Natl. Acad. Sci. USA* 97, 1542–1547
- [128] Paddock ML, Senft ME, Graige MS, Rongey SH, Turanchik T, Feher G, and Okamura MY (1998): Characterization of second site mutations show that fast proton transfer to Q_B⁻ is restored in bacterial reaction centers of *Rhodobacter sphaeroides* containing the Asp-L213 → Asn lesion. *Photosynth. Res.* 58, 281–291
- [129] Ishikita H and Knapp EW (2005): Energetics of proton transfer pathways in reaction centers from *Rhodobacter sphaeroides*. *J. Bio. Chem.* 280, 12446–12450
- [130] Cherepanov DA, Bibikov SI, Bibikova MV, Bloch DA, Drachev LA, Gupta OA, Oesterhelt D, Semenov AY, and Mulikidjanian AY (2000): Reduction and protonation of the secondary quinone acceptor of *Rhodobacter sphaeroides* photosynthetic reaction center: kinetic model based on a comparison of wild-type chromatophores with mutants carrying Arg → Ile substitution at sites 207 and 217 in the L-subunit. *Biochim. Biophys. Acta* 1459, 10–34
- [131] Rongley SH, Paddock ML, Feher G, and Okamura MY (1993): Pathway of proton transfer in bacterial reaction centers: Second-site mutation Asn-M44 → Asp restores electron and proton transfer in reaction centers from the photosynthetically deficient Asp-L213 → Asn mutant of *Rhodobacter sphaeroides*. *Proc. Natl. Acad. Sci. USA* 90, 1325–1329
- [132] Tandori J, Baciou L, Alexov E, Maróti P, Schiffer M, Hanson DK, and Sebban P (2001): Revealing the involvement of extended hydrogen bond networks in the cooperative function between distant sides in bacterial reaction centers. *J. Bio. Chem.* 276, 45513–45515

List of Abbreviations

<i>B.</i>	<i>Blastochloris</i>
<i>Rb.</i>	<i>Rhodobacter</i>
AA	GluL212-AspL213→Ala-Ala mutant of the bRC
ATP	Adenosintriphosphat
Bcl	Bacteriochlorophyll
Bph	Bacteriopheophytin
bRC	Bacterial reaction center
CE/MC	Continuum electrostatic/Monte Carlo
Chl	Chlorophyll
CoQ	Coenzyme Q
ENDOR	electron nuclear double resonance
FNR	Ferredoxin-NADP oxidoreductase
FTIR	Fourier transform infrared
LH1	Light harvesting protein 1
LH2	Light harvesting protein 2
LHC	Light harvesting complex
MSA	Multiple sequence alignment
NADP ⁺	Nicotinamide adenine dinucleotide phosphate
NADPH	reduced form of NADP ⁺
P680	Special pair of the PSII RC
P700	Special pair of the PSI RC
P870	Special pair of the bRC
Phe	Pheophytin

pHMM	profile Hidden Markov Model
PSI	Photosystem I
PSII	Photosystem II
Q_A	Coenzyme Q bound in the Q_A binding site
Q_B	Coenzyme Q bound in the Q_B binding site
RC	Reaction center
RQ	GluL212-AspL213-AlaM246-AlaM247→Ala-Ala-Glu-Asp mutant of the bRC
WOC	Water oxidizing complex
Y_Z	a special tyrosine in the vicinity of the manganese-calcium cluster; abstracts electrons and protons from the manganese-calcium cluster

List of the Manuscripts

- A** Jürgen Köpke, Eva-Maria Krammer, Astrid R. Klingen, Pierre Sebban, G. Matthias Ullmann and Günter Fritzsche (2007): PH Modulates the Quinone Position in the Photosynthetic Reaction Center from *Rb. sphaeroides* in the Neutral and Charge Separated States. *J. Mol. Biol.*, 371, 396-409.

Jürgen Köpke and Günter Fritzsche solved the crystal structures reported in this manuscript. The corresponding parts of the manuscript were written by Jürgen Köpke, Günter Fritzsche and Pierre Sebban. I performed all reported calculations with support and help from Astrid R. Klingen. The results were interpreted by myself, Astrid R. Klingen, Pierre Sebban and G. Matthias Ullmann. The corresponding parts of the manuscript were prepared by myself, Astrid R. Klingen, Pierre Sebban and G. Matthias Ullmann.

- B** Eva-Maria Krammer and G. Matthias Ullmann (2007): The Role of AspL213 for Stabilizing Semiquinone Binding to the Photosynthetic Reaction Center. *Energy of the sun*, Edt.: J. Allen, E. Gantt, J. Golbeck and B. Osmond Springer, 127-131.

All calculations presented in the manuscript were performed by myself. The resulting data were analyzed by myself together with G. Matthias Ullmann. The manuscript was written by myself together with G. Matthias Ullmann.

- C** Eva-Maria Krammer, G. Matthias Ullmann and Timm Essigke (2008): Calculation of Protonation and Reduction Properties of Different Quinone Systems. *To be submitted*.

All calculations presented in the paper were done by myself with support and help of Timm Essigke. The results were interpreted by Timm Essigke, myself and G. Matthias Ullmann. The manuscript was written by myself together with G. Matthias Ullmann and Timm Essigke.

- D** G. Matthias Ullmann, Edda Kloppmann, Timm Essigke, Eva-Maria Krammer, Astrid R. Klingen, Torsten Becker and Elisa Bombarda (2008): Investigating the Mechanisms of Photosynthetic Proteins using Continuum Electrostatics. *Photosyn. Res.*, 97, 33-53.

How continuum electrostatic calculations helped to understand the coupled proton and electron transfer to Q_B was summarized by myself and the corresponding part of the manuscript were prepared by G. Matthias Ullmann, Elisa Bombarda and myself. The remaining parts of the manuscript were prepared by G. Matthias Ullmann, Edda Kloppmann, Timm Essigke, Astrid R. Klingen, Torsten Becker, and Elisa Bombarda.

- E** Eva-Maria Krammer, Pierre Sebban and G. Matthias Ullmann (2008): Profile Hidden Markov Models for Analyzing Similarities and Dissimilarities in the Bacterial Reaction Center and Photosystem II. *Submitted to Biochemistry*.

All calculations were performed by myself. The results were interpreted by myself together with G. Matthias Ullmann and Pierre Sebban. The manuscript was prepared by myself together with G. Matthias Ullmann and Pierre Sebban.

- F** Eva-Maria Krammer and G. Matthias Ullmann (2008): Conservation Analysis of Functional Important Residues of the Oxygen Evolving Mechanism Located in the D1 Subunit of Photosystem II. *From Computational Biophysics to Systems Biology (CBSB08)*, Edt.: U. H. E. Hansmann, J. Meinke, S. Mohanty, W. Nadler, O. Zimmermann, NIC Series, 265-268.

All calculations were performed by myself. The results were interpreted by myself together with G. Matthias Ullmann. The manuscript was prepared by myself together with G. Matthias Ullmann.

- G** Eva-Maria Krammer, Mirco S. Till, Pierre Sebban and G. Matthias Ullmann (2008), Proton Transfer Pathways in Photosynthetic Reaction Centers Analyzed by profile Hidden Markov Models and Network Calculations. *To be submitted*.

The conservation analysis was performed by me. The structure preparation and the network calculations were done by Mirco S. Till and myself. The necessary programs for the network calculations were written by Mirco S. Till. The results were interpreted by myself, Mirco S. Till and G. Matthias Ullmann. The manuscript was prepared by myself together with Mirco S. Till and G. Matthias Ullmann.

Manuscript A

PH Modulates the Quinone Position in the Photosynthetic Reaction Center from *Rb. sphaeroides* in the Neutral and Charge Separated States.

Jürgen Köpke, Eva-Maria Krammer, Astrid R. Klingen, Pierre Sebban, G. Matthias Ullmann and Günter Fritzsche, *J. Mol. Biol.*, 371, 396-409, 2007.

Manuscript B

The Role of AspL213 for Stabilizing Semiquinone Binding to the Photosynthetic Reaction Center.

Eva-Maria Krammer and G. Matthias Ullmann, *Energy of the sun*, Edt.: J. Allen, E. Gantt, J. Golbeck and B. Osmond Springer, 127-131, 2007.

Manuscript C

Calculation of Protonation and Reduction Properties of Different Quinone Systems.

Eva-Maria Krammer, G. Matthias Ullmann and Timm Essigke, *to be submitted*, 2008.

Calculation of Protonation and Reduction Properties of Different Quinone Systems

Eva-Maria Krammer¹, G. Matthias Ullmann¹ and Timm Essigke¹

¹ Structural Biology/Bioinformatics, University of Bayreuth, Universitätsstr. 30, BGI, Bayreuth, 95447, Germany.

Key words

benzoquinone,
charge calculation,
duroquinone,
reduction potential,
semiquinone,
pK_a,
quantum chemistry,
quinone,
quinol.

Abstract

Quinones play an important role in industry and nature due to their ability to be doubly reduced and doubly protonated involving charged and radical intermediate states. In the study presented here, the energetics of all possible reduction and protonation reactions for two quinone systems, the benzoquinone and the duroquinone, are examined by quantum chemical calculations. Correct relative energy levels of the uncharged reactant and product state can be calculated easily. In contrast, the charged and radical states are difficult to describe correctly by quantum chemical methods. Our investigations lead to a protocol optimizing the coordinates of each states at the B3LYP/6-311++G(2d,2p) level of theory followed by single point calculations at the CCSD/6-311G level. Solvent effects are modeled by the CPCM method. Based on the resulting electron distribution, partial charges are obtained by popular methods as Mulliken, CHELPG, NPA and MK for Poisson-Boltzmann electrostatic calculations. By this procedure we succeed in obtaining correct energy levels for the uncharged, charged and radical states. The resulting reaction energies agree within 1.74 kcal/mol for the best charge model with the most reliable experimental results in literature on benzoquinone. Interestingly, our calculations lead to pK_a values for the protonation reaction of the semiquinone 3-4 units lower than in the experiment. This discrepancy can be explained, assuming that the experiments have measured the disproportionation reaction, which was discussed before as a possible side reaction of the used setup. Thus, the protonation of the semiquinone is less likely than proposed before. A complete and self-consistent description of the energetics of all microstates in a quinone system is given for the first time.

Abbreviations

PB - Poisson-Boltzmann

PCM - polarizable continuum model

CPCM - conductor-like PCM

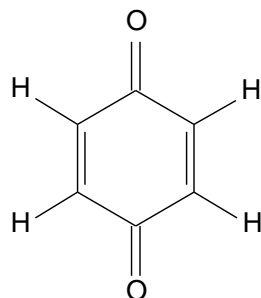
Introduction

Members of the chemical class of the quinones, which can be characterized as aromatic diketones, are important in a lot of different reactions and purposes. Quinones are used in medicine, as an antioxidant, in paints, varnishes, motor fuels and oils. Moreover, quinones can be used as polymerization inhibitors, as cross-linking agent and for the fabrication of dyes and agrochemicals. In industry, very widely used quinones are *p*-benzoquinone (Figure 1a) and duroquinone (Figure 1b). Also in nature, quinones are needed for central reactions like photosynthesis, mitochondrial respiration, extra-mitochondrial electron transfer, and for the regulation of the physicochemical properties of cell membranes.

The enormous importance of the quinones in industry and in nature can be explained by their chemical properties. Each quinone molecule can be doubly reduced and doubly protonated, passing through several intermediate reaction steps as it is shown in Figure 2. The reaction starts with the reduction of the quinone Q to the deprotonated semiquinone $Q^{\cdot-}$ (e1; Figure 2), which can be either protonated leading to the protonated semiquinone QH^{\cdot} (p1; Figure 2) or be reduced to the doubly deprotonated quinol $Q^{2\cdot-}$ (e2; Figure 2). The reduction of QH^{\cdot} (e3; Figure 2) or the protonation of $Q^{2\cdot-}$ (p2; Figure 2) leads to the deprotonated quinol state QH^- . A last protonation reaction (p3; Figure 2) leads to the reaction product, the quinol QH_2 . Theoretically, also positively charged intermediates (QH^+ , QH_2^+ , or QH_2^{2+} ; Figure 2) could be part of the reaction, which have been characterized in acidic solutions of aprotic organic solvents [1–9]. However, due to their low pK_a [10–12] it is not likely that these intermediates occur in aqueous solution. Usually, it is not possible in experiments to measure each reaction independently of the other reactions. In addition, experimental determination of the energetics of each reduction or protonation state can be difficult, since alternative reactions can occur such as the disproportionation reaction (k4; Figure 2) or the quinone-hydroxide addition reaction (kc; Figure 2). Also these alternative reactions have been examined experimentally [13–16].

The energetics of the reduction and protonation reactions of the benzo- and duroquinone

a) Benzoquinone



b) Duroquinone

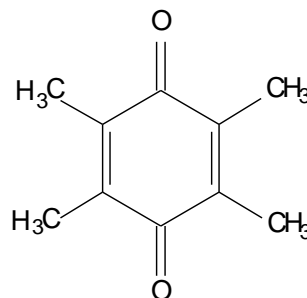
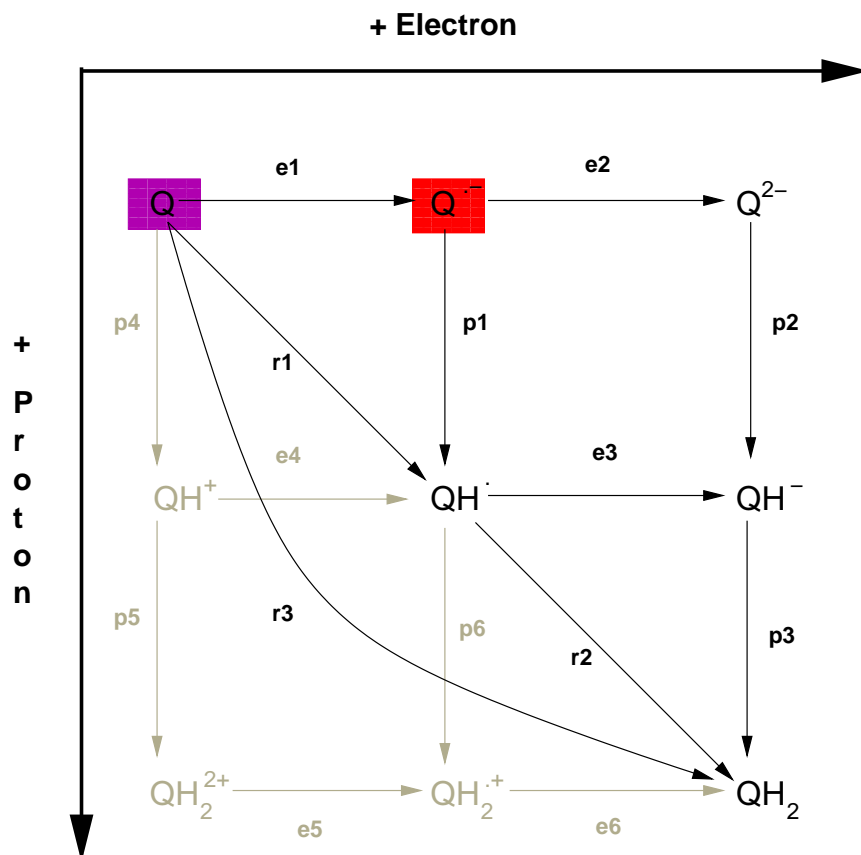


Figure 1: Different quinone molecules: (a) benzoquinone and (b) duroquinone.

system were examined by several experimental [13,15–21,21–29] and theoretical [30,31] studies in the last decades. However, until now, not all energetic parameters are known for the two systems. The easiest reaction to measure is the the complete reaction from the quinone state Q to the quinol state QH₂ (r3; Figure 2), because the pK_a values for the reactions p2 and p3 are rather high. The midpoint potential of the total reaction r3 is experimentally determined to be 0.669 V [16–25] and 0.480 V [21,27] for the benzo- and duroquinone system, respectively. The experiments were done in acidic solution (0.01 M to 0.1 M HCl) containing 50% ethanol in case of duroquinone due to its poor solubility. Baxendale and Hardy [26] have measured the pK_a values of the reaction p2 and p3 of benzoquinone to be 14.61 and 12.15 at 298 K by changes of the optical spectra as function of the pH. The same authors determined the pK_a values of duroquinone to be 13.30 and 10.75 [13,26]. Later Bishop and Tong [14] noticed that at high pH a significant amount of quinone-hydroxide adduct was formed. They corrected the pK_a values of the reactions p2/p3 to be 11.4/9.85 and 13.2/11.25 for the benzo- and duroquinone system, respectively, which will be used here. Slightly higher values of 11.9 and 9.9 for benzoquinone were reported more recently by Bailey and Ritchie [16]. Further, the reaction constants for the disproportionation reaction k4 could be determined to be 4.2 and 1.3 for benzo- and duroquinone, respectively. Also the pK_a values of the reaction p1 have been determined to be 4.1 and 5.1 for benzo- and duroquinone, respectively [15,28,29]. Therefore, pH dependent pulse radiolysis experiments were done with the quinone in acetone and isopropanol (1 M) solution. The generated radicals scavenged by the acetone/isopropanol solution were stabilized on the quinones. The pH dependent decrease of a characteristic absorption band in the Q^{•-} spectrum

a) Reduction and Protonation Scheme



b) Additional Important Reactions

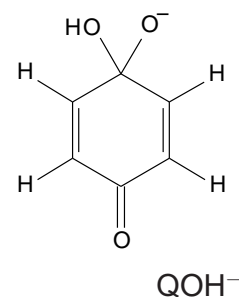
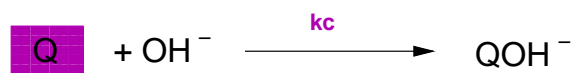


Figure 2: a) Reduction and protonation scheme of a quinone system. In addition to states possibly occurring in aqueous solution (colored black) also positively charged states are shown (colored grey), which are partially characterized in organic solvents. b) Important reactions which can occur in addition to the reduction and protonation reactions are shown in red.

was attributed to the formation of QH[•]. Using the determined midpoint potentials and pK_a values, Ilan *et al.* (1976) calculated the missing standard reduction potentials for the reactions r1/r2 to be 0.339/1.041 V and 0.065/0.895 V for the benzo- and duroquinone system, respectively [32]. The available information is just enough to calculate the neutral and negatively charged states in Figure 2.

The work presented here aims at determining the energetic properties of the reduction and protonation reactions of benzoquinone and duroquinone. Therefore, a suitable quantum chemical approach was developed using benzoquinone as test system and applied to the duroquinone system. Since the data of this study should be the basis for future force field developments and electrostatic calculations, it was required to map the charge distribution obtained by the quantum mechanical calculations onto a set of partial charges. Hence, a choice of a suitable procedure for determining partial charges was similarly important as the choice of a suitable quantum chemical method. The quality of each combination of quantum chemical method and set of partial charges was benchmarked against the experimental energetics of the reduction and protonation reactions of benzoquinone and validated for duroquinone. This is to our knowledge the first time that the energetics of all reduction and protonation steps in the benzo- and duroquinone system are described.

Material and Methods

Quantum Chemical Calculations. For the optimization of all reduction and protonation states, the quinol QH₂ state was created for benzoquinone and duroquinone using the Maestro package. Starting with the quinol QH₂ state, all other reduction and protonation states were modeled by removing hydrogens from the starting quinol structure and adjusting the charge and the spin of the system. Benzoquinone was used as a test system for optimizing the calculation protocol. For geometry optimizations and the vibrational energy calculations of the reduction and protonation states several different methods were used as implemented in Gaussian 03 [33]: Hartree Fock (HF) [34,35], the hybrid HF/density functional theory method B3LYP [36,37] and the Møller-Plesset perturbation theory of second order method (MP2) [38,39]. For all these calculations the basis-set 6-31G was used. In addition the basis-set 6-311G including no, single (+) or double (++) diffusion combined with additional polarization functions on no, 1p, 1p1d, and 2p2d orbitals [40–42] was used for the B3LYP calculations. For all obtained geometries, partial charges were obtained in vacuum using different approaches: orbital based approaches

namely Mulliken [43–46] and natural population analysis (NPA) [47–50], and potential based approaches namely the CHELPG [51] and the Merz-Kollmann [52, 53] (MK) approach. The commonly used Bader charges (atoms in molecules, AIM) were not used due to stability problems especially for the charged species in the implementation available with Gaussian 03. Recent articles have pointed out the limits of this method for charge calculations, i.e. the very ionic charge distributions obtained for water [54] or HCN [55]. Voronoi deformation density charges [55] or Hirshfeld charges [56] are not available in the Gaussian package. Also a restrained electrostatic potential method (RESP) [52, 57], was used combining an orbital and a potential based charge determination method weighting the two methods by a adjustable factor q . Additionally, partial charges were obtained by these charge determination methods after a single point calculation using the coupled cluster method CCSD [58, 59]. For the CCSD calculations the basis-set 6-311G without polarization and diffusion functions was used as concession to the computational cost of the method. Moreover the polarization of partial charges due to the solvent was included by a single point calculations with the conductor-like polarizable continuum model CPCM [60, 61], an implicit solvent model, at the B3LYP and CCSD level. For the duroquinone system, only the procedures resulting in the best fit to the experimental values of the test system benzoquinone were used for the geometry optimization, vibrational energy and partial charge calculation.

Electrostatic Calculations. The solvation energy $\Delta G_s^{\text{solv}, \text{PB}}$ was calculated using the MEAD package [62, 63] as the electrostatic energy difference of a molecule in vacuum and solution. The electrostatic energies were calculated by solving the linearized Poisson-Boltzmann (PB) equation numerically on a grid. Two focussing steps were performed on a cubic grid of 75.5 Å and 15.125 Å per dimension with a grid spacing of 0.5 Å and 0.125 Å, respectively. For all calculations an ionic strength of 100 mM was used. Vacuum and solution were represented by a dielectric constant of 1 and 80, respectively.

Calculation of the Reaction and State Energies. For each protonation and reduction states of the quinone systems, the standard energy G_s° was calculated as:

$$G_s^\circ = G_s^{\text{QM}} + \Delta G_s^{\text{solv}, \text{PB}} + n_{\text{ub } \text{H}^+} \cdot G_{\text{H}^+} + n_{\text{ub } \text{e}^-} \cdot G_{\text{e}^-}, \quad (1)$$

where G_s^{QM} is the quantum chemical energy of the state s , $\Delta G_s^{\text{solv}, \text{PB}}$ is the difference in energy, when the state s is transferred from vacuum to solution with MEAD as described before, G_{H^+} and G_{e^-} are the energies of the unbound proton and electron. The number of unbound protons

and the number of unbound electrons compared to a reference state are given by $n_{\text{ub } \text{H}^+}$ and $n_{\text{ub } \text{e}^-}$. Here, this reference state is set to the fully reduced and protonated QH_2 state. The quantum chemical energy G_{s}^{QM} is directly calculated with Gaussian 03 [33] and consists of two or three terms $\left(G_{\text{s}}^{\text{QM}} = H_{\text{s}}^{\text{vac}} + G_{\text{s}}^{\text{vib}} + H_{\text{s}}^{\text{strain}}\right)$: the vacuum energy $H_{\text{s}}^{\text{vac}}$, the vibrational energy $G_{\text{s}}^{\text{vib}}$ and the strain energy $H_{\text{s}}^{\text{strain}}$ in case the CPCM method was used to polarize the charge distribution for the solvent environment. The CPCM energy, $G_{\text{s}}^{\text{CPCM}}$, as given by Gaussian has two contributions additionally to the vacuum energy without the CPCM model: the CPCM solvation energy, $\Delta G_{\text{s}}^{\text{solv,CPCM}}$, and the strain energy $H_{\text{s}}^{\text{vac}}$. The strain energy is calculated as $H_{\text{s}}^{\text{strain}} = G_{\text{s}}^{\text{CPCM}} - H_{\text{s}}^{\text{vac}} - \Delta G_{\text{s}}^{\text{solv,CPCM}}$. The solvation energy is not taken from the CPCM calculation, since we aimed to calculate the solvation energy based on partial charges as stated in the introduction. Therefore, the solvation energy $\Delta G_{\text{s}}^{\text{solv,PB}}$ is used instead of the solvation energy $\Delta G_{\text{s}}^{\text{solv,CPCM}}$.

The singly protonated states (QH^+ , QH^\cdot and QH^-) have two indistinguishable microscopic states, being protonated at either one of the two oxygen atoms, while their deprotonated counterparts (Q , Q^\cdot and Q^{2-}) have only one microstate. To be comparable with macroscopic (i.e. measurable) pK_a values, the energies of the protonated microstates QH^+ , QH^\cdot and QH^- were adjusted by adding $\ln(2)$ [64]. The energy of an unbound proton is given by

$$G_{\text{H}^+} = H_{\text{H}^+}^{\text{vac}} + \Delta G_{\text{H}^+}^{\text{solv}} + H_{\text{H}^+}^{\text{trans}} + \Delta(pV) - T \cdot [S_{\text{H}^+}], \quad (2)$$

where $H_{\text{H}^+}^{\text{vac}}$ is the energy of formation of a proton, $\Delta G_{\text{H}^+}^{\text{solv}}$ is the solvation energy and $H_{\text{H}^+}^{\text{trans}}$ is the translation energy of a proton, $\Delta(pV)$ is the energy change due to the volume change in the gas-phase reaction and $-T \cdot [S_{\text{H}^+}]$ is the entropic portion of the gas-phase free energy of a proton [65]. The energy of formation of a proton $H_{\text{H}^+}^{\text{vac}}$ is defined as zero in the Gaussian package. The solvation energy of proton $\Delta G_{\text{H}^+}^{\text{solv}}$ is -260.5 kcal/mol as calculated from the experimentally measured potential of the standard hydrogen electrode [66]. The translational energy $H_{\text{H}^+}^{\text{trans}}$ of a proton is $\frac{3}{2}k_{\text{B}}T$, the energy change due to the volume change $\Delta(pV)$ is $k_{\text{B}}T$, and the entropic portion of the gas-phase free energy of a proton $-T \cdot [S_{\text{H}^+}]$ is -7.8 kcal/mol as derived from the Sackur-Tetrode equation [67]. The energy of a free electron G_{e^-} is given by the Faraday constant ($F = 23.06$ kcal/mol) and the standard potential of the hydrogen electrode ($\Delta \text{SHE} = -4.43$ V) [66]:

$$G_{\text{e}^-} = F \cdot \Delta \text{SHE} \quad (3)$$

Two kinds of binding reactions are described in this paper: protonation and reduction re-

actions. The energy corresponding to a binding reaction $\Delta G^{\text{reaction}}$ is given by the energy difference between the bound and the unbound state. The standard energies for both states are calculated using eq. 1 as explained above. To compare the calculated protonation and reduction energies to available experimentally determined pK_a values and the midpoint potentials E_{mid} of the reactions, all quantities are converted to kcal/mol ($\Delta G^{\text{prot}} = pK_a \cdot RT \ln 10$ and $\Delta G^{\text{red}} = F \cdot E_{\text{mid}}$).

Results and Discussion

In this work we examined the protonation and reduction energies between all microstates of two different quinone systems, benzoquinone and duroquinone, using a quantum chemical approach. In a previous theoretical study on quinone systems [30], only the overall reaction was investigated. Further, since partial atomic charges were not of interest to the authors, the solvation energies were directly obtained by a PCM method. Since partial atomic charges are needed for several purposes, e.g. electrostatic calculations, we present here a quantum chemical procedure including partial charge determination, which describes the protonation and reduction energetics of different quinone systems well. Because a lot of experimental data exists for benzoquinone [14, 21, 26, 28, 32], we used this system to search an optimal protocol for the energy calculations by using different quantum chemical methods, basis-sets and charge determination procedures (Table 1). The protocol is then applied to duroquinone, because it is possible to challenge the procedure with experimental values. The resulting reaction energies of the benzo- and duroquinone system are listed in Table 3.

Energetics of the Different Reduction and Protonation States of Benzoquinone.

Towards an Optimal Calculation Protocol. For each reduction and protonation state of the benzoquinone system, the state energy ΔG_s^o (see eq. 1) is calculated by a range of quantum chemical methods. To judge the correctness of the calculations several criteria are used. First, the calculated reaction energy r3 of the double reduction and double protonation of Q to QH₂ (Figure 2) has to be close to the experimentally obtained value. As experimental values we relied on those cited by Ilan *et al.* (1976) [32]. Second, the mean energy difference $\Delta \Delta G_{p2,p3,r3}$ between the calculated and the experimental energies for the reactions p2, p3 and r3 should be low, since these are the most reliable experimental values. Further, in experiments the protonation reaction of QH⁻ (p3) was assigned the lower pK_a value and the protonation reaction of Q²⁻ (p2) the higher [14, 26]. This is what one would expect from theoretical considerations, because

Method/Basis Set	Charge Set	Reaction Energies in kcal/mol			
		r3	$\Delta p2, p3, r3$	p2	p3
experiment	-	-32.24	-	-15.62	-13.49
HF/6-31G	NPA	-3.22	19.19	-25.92	-26.02
	Mulliken	0.63	23.44	-30.23	-32.33
	MK	-2.55	21.66	-30.17	-31.21
	CHELPG	-1.34	22.22	-30.00	-31.39
MP2/6-31G	NPA	2.88	21.47	-25.79	-20.22
	Mulliken	8.15	25.76	-28.86	-27.04
	MK	3.80	23.20	-29.43	-24.70
	CHELPG	5.14	23.90	-29.18	-25.04
B3YLP/6-31G	NPA	-27.95	8.99	-26.88	-23.36
	Mulliken	-22.02	14.13	-30.38	-30.14
	MK	-25.66	12.43	-30.37	-27.72
	CHELPG	-24.72	12.65	-30.25	-27.97
B3YLP/6-311G	NPA	-28.27	3.32	-17.94	-16.93
	Mulliken	-28.17	8.41	-22.83	-25.47
	MK	-28.98	6.72	-23.00	-21.88
	CHELPG	-27.71	7.02	-22.79	-22.20
B3YLP/6-311++G(1d,1p)	NPA	-30.13	4.71	-7.78	-14.30
	Mulliken	-16.08	18.17	8.89	-2.13
	MK	-18.68	4.85	-9.24	-30.37
	CHELPG	-27.01	5.53	-9.32	-18.45
B3YLP/6-311++G(2d,2p)	NPA	-28.83	3.38	-10.01	-14.50
	Mulliken	-30.68	3.93	-11.85	-18.02
	MK	-30.49	3.59	-12.72	-17.51
	CHELPG	-28.89	4.62	-11.53	-18.21
B3YLP/6-311++G(2d,2p)//CPCM	NPA	-30.57	5.30	-12.99	-4.86
	Mulliken	-11.45	17.99	-4.27	6.76
	MK	-29.31	4.29	-15.16	-6.69
	CHELPG	-27.33	4.81	-15.00	-6.79
B3LYP/6-311++G(2d,2p)//CCSD/6-311G	NPA	-30.90	4.19	-9.22	-13.08
	Mulliken	-28.82	5.29	-14.21	-22.42
	MK	-28.48	2.97	-14.03	-18.07
	CHELPG	-27.24	3.53	-13.85	-18.29
B3LYP/6-311++G(2d,2p)//CCSD/6-311G//CPCM	NPA	-28.09	6.94	-12.52	-2.64
	Mulliken	-31.00	1.74	-15.29	-10.76
	MK	-30.80	4.15	-17.44	-6.69
	CHELPG	-28.83	4.44	-17.28	-6.80

Table 1: Performance of different quantum chemical methods, basis-sets and charge determination methods to reproduce experimental data on benzoquinone. Reduction (r3) and protonation (p2 and p3) energies as defined in Figure 2 are given in kcal/mol. $\Delta p2, p3, r3$ is the mean energy difference between the calculated and experimental energies for the reactions p2, p3 and r3. The experimental data are taken from Ilan *et al.* (1976) [32].

it should be easier to abstract the first proton from QH_2 than the second from the negatively charged QH^- . Thus, the third criteria is the trend of the p2 and p3 reaction energies. Table 1 lists some representative results of our quantum chemical calculations.

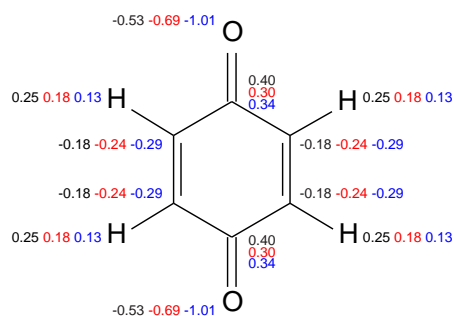
To judge the different quantum chemical methods (HF, MP2, and B3LYP) and different basis-sets (6-31G, 6-311G, 6-311++G(1d,1p), and 6-311++G(2d,2p)) we compare in the following the results of the NPA calculations. However, similar trends can also be seen for the MK and CHELPG calculations. The results of the Mulliken calculations, however, show no steady convergence with increasing quality of the quantum chemical method as criticized often before. Both, the HF/6-31G and the MP2/6-31G calculations lead to poor agreement with experiment. The calculated reaction energy for r3 differs by about 30 kcal/mol, while with a B3LYP/6-31G method/basis-set only a difference of about 5 kcal/mol is obtained. Also previous quantum chemical calculations for the Q and QH_2 state showed, that a good fit for the reaction energy r3 can be obtained using the B3LYP method [30, 31]. With a higher basis-set (6-311G) and with addition of diffusion and polarization (6-311++G(1d,1p) and 6-311++G(2d,2p)) functions the calculated energy r3 approaches the experimental value. As previously described, also in our calculations the B3LYP/6-311++G(1d,1p) method/basis-set leads already to a very good description of the total reaction energy r3 [30]. In contrast to the earlier study, we do not calculate the solvation energy with the PCM model but use partial charges to calculate the PB solvation energy. But this does not change the quality of the results on the reaction energy r3 significantly. However, although the overall reaction r3 is calculated in good agreement with experiment, another criteria is not fulfilled: the trend of the reaction energies of p2 and p3 is not correct, i.e. second deprotonation step requires less energy than the first step. Our data shows (Table 1), that the trend is wrong for all charge determination methods performed in vacuum and independent on the used quantum chemical method or used basis-set. Using CPCM to include the polarization of the electron distribution due to the polar solvent environment a correct trend of p2 and p3 is observed. But in these calculations, the error in the energy obtained for reactions p2, p3 and r3 increases by nearly 2 kcal/mol compared to the results without CPCM (Table 1). The error originates from a very bad agreement with experiment of the protonation reaction p3. A reason could be, that the HOMOs of the highly charged species Q^{2-} are still not sufficiently represented by the basis-set and the results suffer from basis-set incompleteness. After geometry optimization at the B3LYP/6-311++G(2d,2p) level of theory we therefore did single point cal-

culations with the CCSD method to minimize the effects of basis-set incompleteness especially in the valence orbitals. The CPCM method was used again to model the polarization due to the solvent. This procedure leads to worse results if the partial charges are determined using the NPA method, but using the CHELPG or the Mulliken method the results improve significantly (Table 1). Especially the Mulliken method leads to very good agreement with experimental energies. For the reaction r3 an energy of -31.00 kcal/mol is obtained which is close to the experimentally determined value ($\Delta\Delta G_{r3} = -32.24$ kcal/mol). Also the trend of the energies of reactions p2 and p3 is calculated properly and the calculated energies are closer than before to the experimental values ($\Delta G_{p2} = 0.33$ kcal/mol and $\Delta G_{p3} = 2.73$ kcal/mol). This leads to the lowest energy difference $\Delta\Delta G_{p2,p3,r3}$ (1.74 kcal/mol) observed in the calculations made here. Additionally for all three methods, the determined partial charges (Figure 3) are symmetric for the Q, Q^- , Q^{2-} and QH_2 and not symmetric for the QH and QH^- in agreement with the symmetry of the structures. Thus, the best results for the benzoquinone system were obtained for structures geometry optimized with the B3LYP/6-311++G(2d,2p) method and single point energies obtained with CCSD/6-311G and the CPCM solvent model followed by fitting partial charges with the Mulliken method.

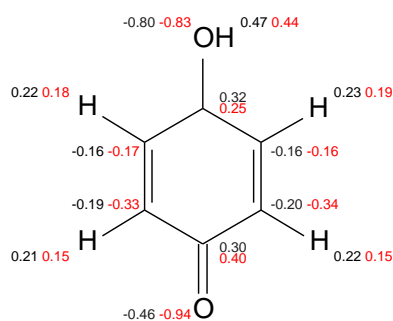
Based on the previously performed calculations, we used the RESP charge fit approach to obtain the standard energies G_s° for all considered reduction and protonation states. Therefore, all possible combinations of the used orbital and potential based charge determination methods and different values of the weighting factor q were tested for the RESP charge calculations. The resulting reaction energies for the reactions p2, p3 and r3 and the mean energy difference $\Delta\Delta G_{p2,p3,r3}$ are listed in Table 2 for the different RESP calculations. Most of the RESP calculations have a reaction energy r3 comparable to the experimental result. However, none of the RESP calculations performs better or comparably well in calculating the mean energy difference $\Delta\Delta G_{p2,p3,r3}$ as the Mulliken charge determination approach alone. Thus, we do not examine the RESP calculations further, but state, that for the benzoquinone system the Mulliken partial charges calculations show the best fit to the experiments.

Calculation of Energy Levels. To obtain the reaction energies, we calculated the total energies G_s° of each state. It is possible to compare the energy levels of the different protonation and reduction states relative to a reference state, the reactant state, quinone Q. For the Q^{2-} , QH^- , and QH_2 state relative energy levels can be calculated using the experimental values of p2,

a) unprotonated



b) singly protonated



c) protonated

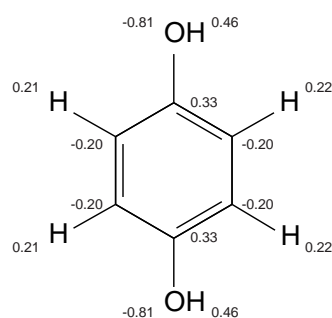


Figure 3: Partial charges of the different benzoquinone states. Mulliken charges are calculated after the B3LYP/6-331++G(2d,2p)//CCSD/6-311G//CPCM procedure. (a) Partial charges for Q (black), Q^{·-} (red) and Q²⁻ (blue); (b) partial charges for QH[·] (black), QH⁻ (red); (c) partial charges for QH₂ (black).

Input charge sets		RESP weight	Reaction Energies in kcal/mol			
Potential based	Orbital based	q	r3	p2	p3	$\Delta p2.p3.r3$
MK	Mulliken	0.001	-31.11	-17.42	-6.72	4.10
MK	NPA	0.001	-30.96	-17.42	-6.69	4.13
CHELPG	Mulliken	0.001	-28.85	-17.28	-6.80	4.44
CHELPG	NPA	0.001	-28.91	-17.28	-6.80	4.42
MK	Mulliken	0.010	-32.36	-17.36	-6.84	3.97
MK	NPA	0.010	-31.61	-17.31	-6.67	4.08
CHELPG	Mulliken	0.010	-28.99	-17.29	-6.81	4.40
CHELPG	NPA	0.010	-28.91	-17.28	-6.80	4.42
MK	Mulliken	0.100	-34.00	-17.53	-7.31	3.87
MK	NPA	0.100	-32.03	-17.07	-6.43	4.17
CHELPG	Mulliken	0.100	-29.91	-17.38	-6.88	4.18
CHELPG	NPA	0.100	-29.36	-17.33	-6.75	4.34
MK	Mulliken	0.300	-34.72	-17.69	-7.68	3.84
MK	NPA	0.300	-31.93	-16.76	-6.07	4.34
CHELPG	Mulliken	0.300	-30.79	-17.50	-6.99	4.00
CHELPG	NPA	0.300	-29.74	-17.37	-6.68	4.31
MK	Mulliken	0.500	-34.86	-20.83	-5.86	3.82
MK	NPA	0.500	-31.81	-16.51	-5.80	4.48
CHELPG	Mulliken	0.500	-31.23	-17.57	-7.08	3.91
CHELPG	NPA	0.500	-29.88	-17.36	-6.63	4.31

Table 2: Performance of different orbital and potential based methods to determine partial charges combined by the RESP procedure on benzoquinone. These results of the B3LYP/6-311++G(2d,2p)//CCSD/6-311G//CPCM procedure can be directly compared with the results of the pure partial charge sets and experimental data given in Table 1.

p3 and r3 for comparison. In Figure 4a, the calculated energy levels are plotted, showing again that the results with the Mulliken charge set agree best with the experimental data. Moreover, it can be seen, that the energy difference between the neutral charge states QH^\cdot and QH_2 for different charge determination methods is small. In contrast, the charged states Q^- , QH^- and especially the Q^{2-} state have different energies depending on the charge set used. Due to the very accurate quantum chemical description, the simple Mulliken procedure is able to yield charges representing the charged states much better than the other charge determination methods. We consider it to be important, that the geometry of the nuclear coordinates is well optimized as we obtained it by the B3LYP/6-311++G(2d,2p) method. Further the orbitals, in particular the valence orbitals, have to be described well, as it can be achieved by the CCSD method. Mulliken analysis is known to have problems with polarization and diffusion functions, thus the 6-311G basis-set is optimal in the CCSD calculation step. Additionally, the polarization of the charge distribution due to the high dielectric of the solvent improves the Mulliken results a lot. Thus, the calculation protocol we chose is particularly well suited for the Mulliken method,

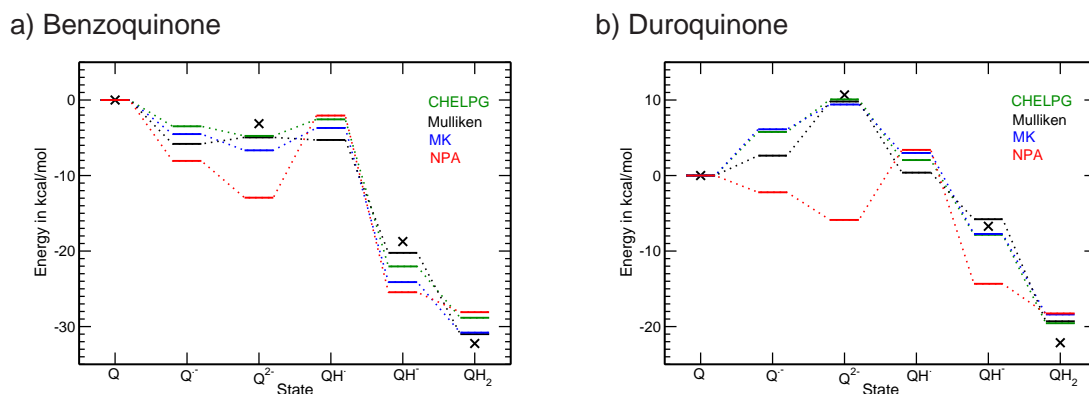


Figure 4: Energy levels of quinone states compared to experiment. Energies were calculated by the B3LYP/6-331++G(2d,2p)//CCSD/6-311G//CPCM procedure and solvation energies were calculated with point charges fitted by CHELPG, Mulliken, MK or NPA. The energies are shown for the Q, Q⁻, Q²⁻, QH, QH⁻ and QH₂ states of benzoquinone (a) and duroquinone (b). The experimental values for benzo- and duroquinone are taken from Ilan *et al.* (1976) [32].

albeit the other charge determination methods generally also improve. As it can be seen from Figure 4a, the reactant state, the quinone Q, is the highest in energy for benzoquinone. The semiquinone Q⁻ is lower in energy, whereas the doubly deprotonated quinol Q²⁻ is higher in energy compared to this semiquinone. Thus, already here all other charge determination methods fail, showing lower energies for Q²⁻ than Q⁻, which contradicts chemical intuition expecting a doubly charged species to be most unfavorable. The protonated semiquinone QH has about the same energy compared to the semiquinone Q⁻ for the Mulliken charge set, while higher energies are obtained by the other charge determination methods. The other quinol states, the singly and doubly protonated quinol QH⁻ and QH₂ are lower in energy.

Calculation of Reaction Energies. In addition, the here calculated total energies G_s° can be used to calculate the reaction energies for each protonation and reduction step (Figure 2). In Table 3 these reaction energies are given for the B3LYP/6-311++G(2d,2p)//CCSD/6-311G//CPCM calculations. As it was already observed for the reaction energies for p2, p3 and r3, the energies of the possible reduction and protonation reactions differ between the methods. However, in the following only the calculated reaction energies for the best solution obtained using the Mulliken charges will be discussed. Ilan *et al.* (1976) [32] calculated the reaction energies r1 and r2 based on experimental values for the reactions p2, p3, r3 and the disproportionation

System	Charge Set	Reaction Energies in kcal/mol								
		e1	e2	e3	r1	r2	r3	p1	p2	p3
BQ	experimental	-	-	-	-7.54	-24.74	-32.24	(-5.62)	-15.62	-13.49
	Mulliken	-5.81	0.87	-14.95	-5.29	-25.71	-31.00	0.52	-15.29	-10.76
	CHELPG	-3.48	-1.27	-19.47	-2.56	-26.27	-28.83	0.92	-17.28	-6.80
	NPA	-4.51	-2.16	-20.41	-3.70	-27.10	-30.80	0.81	-17.44	-6.69
	MK	-8.07	-4.87	-23.39	-2.06	-26.03	-28.09	6.01	-12.52	-2.64
DQ	experimental	-	-	-	-1.29	-20.85	-22.14	(-6.99)	-17.40	-15.41
	Mulliken	2.63	7.17	-6.17	0.38	-19.68	-19.30	-2.25	-15.59	-13.52
	CHELPG	5.77	4.31	-9.90	2.05	-21.61	-19.56	-3.72	-17.93	-11.71
	NPA	-2.22	-3.66	-17.73	3.39	-21.65	-18.26	5.61	-8.45	-3.93
	MK	6.12	3.28	-10.72	2.99	-21.38	-18.39	-3.14	-17.14	-10.66

Table 3: Reaction energies obtained by the four charge determination methods with the B3LYP/6-311++G(2d,2p)//CCSD/6-311G//CPCM procedure. All reaction energies for benzoquinone (BQ) and duroquinone (DQ) are given in kcal/mol. The experimental data for benzoquinone and duroquinone is taken from Ilan *et al.* (1976) [32].

Charge Set	Reaction Energies in kcal/mol					
	p4	p5	p6	e4	e5	e6
Mulliken	16.58	26.39	-1.63	-21.87	-49.90	-24.08
CHELPG	15.20	26.52	-1.37	-17.76	-45.65	-24.90
NPA	19.14	28.88	1.28	-21.20	-48.81	-27.31
MK	15.26	26.58	-1.00	-18.96	-46.54	-26.10

Table 4: Energies for the reactions including positively charged states obtained by the four chargefit methods with the B3LYP/6-311++G(2d,2p)//CCSD/6-311G//CPCM procedure for benzoquinone.

reaction k4. The energies calculated here for r1 and r2 are relatively close to the experimental value (2.3 and 1 kcal/mol, respectively). We calculated a pK_a value for the protonation of the semiquinone $Q^{\cdot-}$ of -0.38 (0.5 kcal/mol), meaning, that a protonated semiquinone QH^{\cdot} is not even likely to be formed at pH 0. The significant discrepancy to the experimental pK_a value of 4.1 will be discussed in a separate section. In addition, if a protonated semiquinone occurs, it is fastly reduced to the singly protonated quinol state QH^- as it can be seen from the low reaction energy calculated for e3 (-14.95 kcal/mol). Also the reduction of the reactant state Q to the semiquinone $Q^{\cdot-}$ is associated with a low energy of -5.81 kcal/mol. In contrast, the subsequent reduction of the semiquinone is not as favorable since the reaction energy is slightly positive (0.87 kcal/mol). As stated above, the doubly reduced and unprotonated quinol Q^{2-} is not stable in aqueous solution due to its high pK_a value and is therefore protonated in the reaction following steps.

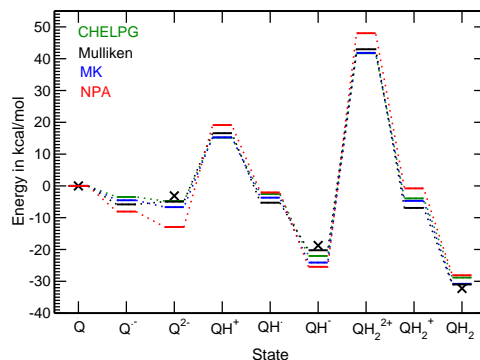


Figure 5: Energy levels of all nine benzoquinone states compared to experiment. Energies were calculated by the B3LYP/6-331++G(2d,2p)//CCSD/6-311G//CPCM procedure and solvation energies were calculated using the Mulliken, CHELPG, npa, or mk charges. The energies are shown for the Q, Q^- , Q^{2-} , QH^+ , QH^- , QH_2^{2+} , QH_2^{2+} and QH_2 states of benzoquinone. The experimental values for benzo- and duroquinone are taken from Ilan *et al.* (1976) [32].

Calculation of Positively Charged States. Additionally, we calculated the positively charged states to test if they are as unlikely in aqueous solution as they are generally assumed. For better comparison with the results discussed above, the energies of all nine states are shown relative to Q in Figure 5. The additional reaction energies for the protonation reactions p4, p5, and p6, and the reduction reactions e4, e5 and e6 are shown in Table 4.

Both, the QH^+ and the QH_2^{2+} state are very high in energy and therefore not accessible in a aqueous solution. Interestingly, the state energy for QH_2^{2+} is similar to the energies of Q^- and QH^- . For all determined charge sets except NPA, QH_2^{2+} is even the most stable semiquinone state at pH 0. This finding is in agreement with experiment, where in dichloromethane at -50°C with strong acids the hydroquinone cation radical was found to be stable [9]. However, the pK_a value for the reaction p6 is for three charge determination methods around 1 leading to an easy first deprotonation followed by an exergonic deprotonation of the second proton (Table 4), because the asymmetric QH^- state is less stable than Q^- . The energies for the other two protonation reactions, p4 and p5 (from Q to QH^+ and from QH^+ to QH_2^{2+} , respectively) are very high leading to strongly negative pK_a values for both reactions. The trend is in the right direction making the first protonation more likely than the second.

Energetics of the Different Reduction and Protonation States of Duroquinone. Since we obtained very good agreement for benzoquinone with the experiments we consider most

reliable, we used the B3LYP/6-311++G(2d,2p)//CCSD/6-311G//CPCM procedure and the four mentioned methods to determine partial charge sets to calculate energies for duroquinone. Also experimental values exist for the reduction and protonation steps of duroquinone (Figure 2). We followed again the data in Ilan *et al.* (1976). Using different charge sets, again significantly different results were obtained (Table 3). Once more the best agreement with the experimental values (with a mean energy difference $\Delta\Delta G_{p2,p3,r3}$ of 2.23 kcal/mol) is obtained by the Mulliken method followed by the CHELPG approach (with a mean energy difference of $\Delta\Delta G_{p2,p3,r3}$ of 2.63 kcal/mol). Also the trend calculated for p2 and p3 is correct in our calculations. In contrast to the benzoquinone system, the reactant state Q is not the state with the highest state energy (Figure 4b). Especially the doubly deprotonated quinol state Q^{2-} is much higher in energy. The relative state energy of Q^{2-} calculated here is close to the value calculated from experimental data for all charge determination methods except NPA, giving confidence in our protocol. Both semiquinone states $Q^{\cdot-}$ and $QH^{\cdot-}$ are significantly lower in energy compared to the unprotonated quinol state Q^{2-} .

Duroquinone is well known for its long living and relatively stable semiquinone state especially at high pH. This stability can be understood by the calculated energy levels. If duroquinone is reduced, but the protonated states $QH^{\cdot-}$ and QH_2 are not accessible due to the high pH of the solution, the reaction can be trapped at reduction potentials not low enough to form Q^{2-} . For benzoquinone instead, the energy level of the Q^{2-} state is not much higher than the $Q^{\cdot-}$ state, resulting in a full reduction of Q to Q^{2-} without a stable radical intermediate (Figure 4a and b). The differences in the energy levels of the between duro- and benzoquinone can also be seen in the reaction energies (Table 1). For the reaction p1 we calculated a pK_a value of -0.38 (0.52 kcal/mol) for benzoquinone and a pK_a value of 1.67 (-2.25 kcal/mol) for duroquinone. Thus a protonated radical is more likely to be observed for the duroquinone system.

The Mulliken calculations show a good agreement with experiments for the reactions p1, p2, and r3 with trustable experimental data and the explanation of the erroneous assignment of experimental data to p1 given in the following section holds true for both systems (Table 3). Thus, the developed protocol is also able to reproduce the experimental results for the duroquinone system, which makes it applicable to other quinone systems.

Energetics of the Protonation Reaction of the Semiquinone $Q^{\cdot-}$. For both, the benzo- and the duroquinone system, the calculated reaction energies are generally in good agreement

with experimental data. However, the calculated protonation energy of the semiquinone $Q^{\cdot-}$ clearly differs from experimental results assigned to this reaction. For reaction p1 in the benzoquinone and duroquinone system a high energy difference is obtained between calculations and experiment of about > 6 kcal/mol and > 4 kcal/mol, respectively. The reason for this discrepancy could be that not the intended reaction is actually measured in the experiment. The pulse radiolysis experiments of Willson [28] monitored the reaction of the benzoquinone system by absorption spectroscopy at 430 nm. The intense color of the $Q^{\cdot-}$ at pH 7 decayed towards lower pH, which was attributed to the protonation of the semiquinone (reaction p1). We assume, that instead the measured reaction is the disproportionation of the semiquinone k4 and subsequent protonation of Q^{2-} (Figure 2).

In fact it was stated that 'At pH = 2, however, an absorption similar to that observed in acid t-butanol solutions of benzoquinone was apparent'. Since he measured at pH 2 - 7 the doubly deprotonated quinol state Q^{2-} is not stable but immediately protonates. It is very likely, that Willson measured a mixture of the following two reactions:



and



Based on our B3LYP/6-311++G(2d,2p)//CCSD/6-311G//CPCM//Mulliken calculations, we obtain for the benzoquinone system an reaction energy of -4.31 kcal/mol and -9.69 kcal/mol for k4a and k4b, respectively. The experimental value attributed to p1 before is -5.61 kcal/mol and lies in between the energies of k4a and k4b. Analogously, for the duroquinone system we obtain an reaction energy of -5.53 kcal/mol and -12.28 kcal/mol for k4a and k4b, respectively. The experimental value of -6.99 kcal/mol lies again in between the energies of k4a and k4b. Thus, it is likely, that the occurrence of radical-radical reactions, Willson considered as a possible source of errors in his experiments, plays actually a dominant role for both, the benzo- and the duroquinone system and that the energy determined here for p1 is correct instead. In summary, our calculations could not only be validated by their good agreement with reliable experimental data and provide energies for reactions, which could not be measured and have not been calculated so far, but also misinterpretations of experimental results could be detected and corrected.

Conclusions

The protonation and reduction scheme of the benzo- and duroquinone system is complex and contains highly charged and radical intermediate states. In contrast to the uncharged reactant and product state, the intermediate states are not trivial to characterize by quantum-chemical methods. Standard methods and basis-sets like B3LYP/6-311G fail to represent these intermediate states. Convincing results for all microstates could be obtained combining geometry optimization at the B3LYP/6-311++G(2d,2p) level and single point calculations at the CCSD/6-311G level. CPCM was used to optimize the electron distribution in an implicit solvent. For each state charge distributions were obtained by orbital-based approaches (Mulliken and natural population analysis) as well as potential-based approaches (CHELPG and Merz-Kollman). Good agreements with experimental results were achieved using Mulliken or CHELPG charges for the calculation of protonation and reduction energies. Noticeably, for all tested sets of partial charges our results for the protonation reaction of the semiquinone $Q^{\cdot-}$ differ significantly from the experimentally determined values. For benzoquinone, the experiments assigned a pK_a value of about 4.1 to this reaction while our calculations indicate a pK_a value below zero. Also for duroquinone, the pK_a value is 2.7 or less dependent on the charge set, but definitively lower than the experimental value of 5.1. This discrepancy can be explained, assuming that instead of the intended protonation reaction the disproportionation of the semiquinone to Q and Q^{2-} with subsequent protonation of Q^{2-} to QH^- and QH_2 was measured.

To our knowledge, we give for the first time a complete and self-consistent description of the energetics of all microstates in a quinone system, which agrees well with the experimental data on the overall reaction (Q to QH_2) and the two pK_a values for deprotonating QH_2 . The newly determined pK_a value for the semiquinone protonation as well as energies of reactions which have not been characterized so far might facilitate both experimental and theoretical studies of quinone systems.

Acknowledgments

This work was supported by the DFG grant UL174/7-1

References

- [1] Eggins BR and Chambers JQ (1969): Electrochemical oxidation of hydroquinone in acetonitrile. *J. Chem. Soc. Chem. Commun.* 232–233
- [2] Eggins BR (1969): Interpretation of electrochemical reduction and oxidation waves of quinone-hydroquinone system in acetonitrile. *J. Chem. Soc. Chem. Commun.* 1267–1268
- [3] Parker VD (1969): The hydroquinone-quinone redox behaviour in acetonitrile. *J. Chem. Soc. Chem. Commun.* 716–717
- [4] Bessard J, Cauquis G, and Serve D (1970): Propriétés électrochimiques du couple quinone-hydroquinone en milieu organique. Intervention de la quinone protonée. *Tetrahedron Lett.* 35, 3103–3106
- [5] Parker VD and Ebersson L (1970): On the anodic oxidation of hydroquinone in acetonitrile. *J. Chem. Soc. Chem. Commun.* 1289–1290
- [6] Eggins BR (1972): Evidence for a one-electron intermediate in the anodic oxidation of hydroquinone in acetonitrile. *J. Chem. Soc. Chem. Commun.* 427–428
- [7] Eggins BR (1973): Detection of electrochemically generated intermediates by variable sweep rate cyclic voltammetry. *Faraday Discuss. Chem. Soc.* 56, 276–284
- [8] Parker VD (1973): The anodic oxidation of hydroquinone in acetonitrile on the question of a possible one electron intermediate. *Electrochim. Acta* 18, 519–524
- [9] Hammerich O and Parker VD (1982): The oxidation of hydroquinone by protonated quinone. *Acta Chem. Scand.* 1, 63–64
- [10] Rich PR (1984): Electron and proton transfers through quinones and cytochrome *bc* complexes. *Biochim. Biophys. Acta* 768, 53–79
- [11] Brandt U (1996): Bifurcated ubiquinol oxidation in the cytochrome *bc₁* complex by proton-gated charge transfer. *FEBS Lett.* 387, 1–6
- [12] Brandt U (1997): Proton-translocation by membrane-bound NADPH:ubiquinone-oxidoreductase (complex I) through redox-gated ligand conduction. *Biochim. Biophys. Acta* 1318, 79–91
- [13] Baxendale JH and Hardy R (1953): The formation constant of durosemiquinone. *Trans. Faraday Soc.* 49, 1433–1437
- [14] Bishop CA and Tong LKJ (1965): Equilibria of substituted semiquinones at high pH values. *J. Amer. Chem. Soc.* 87, 501–505
- [15] Adams BE and Michel BD (1967): Pulse radiolysis of benzoquinone and hydroquinone. *Trans. Faraday Soc.* 63, 1171–1180
- [16] Bailey SI and Ritchie IM (1985): A cyclic voltammetric study of the aqueous electrochemistry of some quinones. *Biochem. Biophys. Acta* 30, 3–12
- [17] Biilmann E and Krarup I (1924): CCL. - The temperature coefficient of the quinhydrone electrode. *J. Chem. Soc. Trans.* 125, 1954–1956

- [18] Hovorka F and Dearing WC (1935): The "salt error" and normal electrode potential of the quinhydrone electrode at 25⁰. *J. Am. Chem. Soc.* 57, 446–453
- [19] Stonehill HI (1943): The salt error of the quinhydrone electrode in aqueous nitric acid, and the potentials of the hydroquinhydrone and quino-quinhydrone electrodes. *Trans. Faraday Soc.* 39, 67–72
- [20] Harned HS and Wright DD (1935): A study of the cell, Pt/Quinhydrone, Hcl (0,01*M*)/AgCl/Ag, and the normal electrode potential of the quinhydrone electrode from 0 to 40⁰. *J. Am. Chem. Soc.* 55, 4849–4857
- [21] Clark WM (1960): *Oxidation-reduction potentials of organic systems*. Baltimore, The Williams & Wilkins Company
- [22] Hale JM and Parsons R (1963): Reduction of p-quinones at a dropping mercury electrode. *Trans. Faraday Soc.* 59, 1430–1437
- [23] Kano K and Uno B (1993): Surface-redox reaction mechanism of quinones adsorbed on basal-plane pyrolytic graphite electrodes. *Anal. Chem.* 65, 1088–1093
- [24] Driebergen RJ, Den Hartigh J, Holthuis JJM, Hulshoff A, van Oort WJ, Kelder SJP, Verboom W, Reinhoudt DN, Bos M, and van der Linden WE (1990): Electrochemistry of potentially bioreductive alkylating quinones. Part I. Electrochemical properties of relatively simple quinones, as model compounds of mitomycin- and ayridinylquinone-type antitumour agents. *Analytica Chimica Acta* 233, 251–268
- [25] LaMer VK and Baker LE (1922): The effect of substitution on the free energy of oxidation-reduction reactions. I. benzoquinone derivatives. *J. Am. Chem. Soc.* 44, 1954–1964
- [26] Baxendale JH and Hardy R (1953): The ionization constants for some hydroquinones. *Trans. Faraday Soc.* 49, 1140–1144
- [27] Conant JB and Fieser LF (1923): Reduction potentials of quinones I. The effect of the solvent of the potentials of certain benzoquinones. *J. Am. Chem. Soc.* 44, 2194–2218
- [28] Willson RL (1971): Pulse radiolysis studies of electron transfer in aqueous quinone solutions. *Trans. Faraday Soc.* 3020–3029
- [29] Patel KP and Willson RL (1973): Semiquinone free radicals and oxygen. *J. Chem. Soc., Faraday Trans. I* 69, 814–825
- [30] Wass JRTJ, Ahlberg E, Panas I, and Schiffrin DJ (2006): Quantum chemical modelling of the rate determining step for oxygen reduction on quinones. *Phy. Chem. Chem. Phys.* 8, 4189–4199
- [31] Namazian M (2003): Density functional theory response to the calculation of electrode potentials of quinones in non-aqueous solution of acetonitrile. *J. Mol. Stru.* 664-665, 273–275
- [32] Ilan YA, Czapski G, and Meisel D (1976): The one-electron transfer redox potentials of free radicals I. The oxygen/superoxide system. *Biochim. Biophys. Acta* 430, 209–224

- [33] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, and Pople JA. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004
- [34] Wiliams Q and Weatherly TL (1953): Nuclear quadrupole resonances in some chlorine compounds. *J. Chem. Phys.* 22, 572
- [35] McWeeny R and Diercksen G (1961): Self-consistent perturbation theory. II. Extension to open shells. *J. Chem. Phys.* 49, 4852–4856
- [36] Becke AD (1992): Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652
- [37] Becke AD (1992): Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652
- [38] Frisch MJ, Head-Gordon M, and Pople JA (1990): A direct MP2 gradient method. *Chem. Phys. Lett.* 166, 275–280
- [39] Head-Gordon M, Pople JA, and Frisch MJ (1988): MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* 153, 503–506
- [40] Clark T, Chandrasekhar J, Spitznagel GW, and von Ragué Schleyer P (1982): Efficient diffuse function-augmented basis sets for anion calculations. III.* The 3-21+G basis set for first-row elements, Li-F. *J. Comp. Chem.* 1982, 294–301
- [41] Krishnan R, Binkley JS, and Pople JA (1980): Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* 72, 650–654
- [42] Woon DE and Jr THD (1992): Gaussian basis sets for use in correlated molecular calculations. III: The atoms aluminum through argon. *J. Chem. Phys.* 98, 1358–1371
- [43] Mulliken RS (1955): Electronic population analysis on LCAO-MO molecular wave function. *J. Chem. Phys.* 23, 1833–1840
- [44] Mulliken RS (1955): Electronic population analysis on LCAO-MO molecular wave functions. II. Overlap populations, bond orders, and covalent bond energies. *J. Chem. Phys.* 23, 1841–1846
- [45] Mulliken RS (1955): Electronic population analysis on LCAO-MO molecular wave functions. III. Effects of hybridization on overlap and cross AO populations. *J. Chem. Phys.* 23, 2338–2342
- [46] Mulliken RS (1955): Electronic population analysis on LCAO-MO molecular wave functions. IV. Bonding and antibonding in LCAO and valence-bond theories. *J. Chem. Phys.* 23, 2343–2346
- [47] Foster JP and Weinhold F (1980): Natural hybrid orbitals. *J. Am. Chem. Soc.* 102, 7211–7218

- [48] Reed AE and Weinhold F (1983): Natural bond orbital analysis of near-Hartree-Fock water dimer. *J. Chem. Phys.* 78, 4066–4073
- [49] Reed AE, Weinstock RB, and Weinhold F (1985): Natural population analysis. *J. Chem. Phys.* 83, 735–746
- [50] Reed AE, Curtiss LA, and Weinhold F (1988): Intermolecular interactions from natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* 88, 899–926
- [51] Breneman CM and Wilberg KB (1990): Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comp. Chem.* 11, 361–373
- [52] Besler BH and Merz Jr KM (1989): Atomic charges derived from semiempirical methods. *J. Comp. Chem.* 11, 431–439
- [53] Singh UC and Kollmann PA (1984): An approach to computing electrostatic charges for molecules. *J. Comp. Chem.* 5, 129–145
- [54] Martin F and Zipse H (2004): Charge distribution in the water molecule - a comparison of methods. *J. Comput. Chem.* 26, 97–105
- [55] Guerra CF, Handgraaf JW, Baerends EJ, and Bickelhaupt FM (2004): Voronoi deformation density (VDD) charges: assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD methods for charge analysis. *J. Comput. Chem.* 25, 189–210
- [56] Hirschfeld FL (1977): Bonded-Atom fragments for describing molecular charge densities. *Theoret. Chim. Acta* 44, 129–138
- [57] Bayly CI, Cieplak P, Cornell WD, and Kollmann PA (1993): A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 97, 10269–10280
- [58] Puris III GD and Bartlett RJ (1982): A full coupled-cluster singles and doubles model: the inclusion of disconnected triples. *J. Chem. Phys.* 76, 1910–1918
- [59] Scuseria GE, Janssen CL, and Schaefer III HF (1988): An efficient reformulation of the closed-shell coupled cluster single and double excitation (CCSD) equations. *J. Chem. Phys.* 89, 7382–7387
- [60] Klamt A and Schüürmann G (1993): COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2*, 799–805
- [61] Barone V and Cossi M (1998): Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* 102, 1995–2001
- [62] Bashford D and Karplus M (1990): pK_a 's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* 29, 10219–10225
- [63] Bashford D and Gewert K (1992): Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* 224, 473–486
- [64] Lopez X, Schaefer M, Dejaegere A, and Karplus M (2002): Theoretical evaluation of pK_a in phosphoranes: implications for phosphate ester hydrolysis. *J. Am. Chem. Soc.* 124, 5010–5018

- [65] Ullmann GM, Noodleman L, and Case DA (2002): Density functional calculation of pK_a values and redox potentials in the bovine Rieske iron-sulfur protein. *J. Biol. Inorg. Chem.* 7, 632–639
- [66] Reiss H and Heller A (1985): The absolute potential of the standard hydrogen electrode: a new estimate. *J. Phys. Chem.* 89, 4207–4213
- [67] Hill TL (1960): *An introduction to statistical thermodynamics*. Dover, New York

Manuscript D

Investigating the Mechanisms of Photosynthetic Proteins using Continuum Electrostatics.

G. Matthias Ullmann, Edda Kloppmann, Timm Essigke, Eva-Maria Krammer, Astrid R. Klingen, Torsten Becker and Elisa Bombarda, *Photosyn. Res.*, 97, 33-53, 2008.

Manuscript E

Profile Hidden Markov Models for Analyzing Similarities and Dissimilarities in the Bacterial Reaction Center and Photosystem II.

Eva-Maria Krammer, Pierre Sebban and G. Matthias Ullmann, *Biochemistry*, 48, 1230-1243, 2009.

Manuscript F

Conservation Analysis of Functional Important Residues of the Oxygen Evolving Mechanism Located in the D1 Subunit of Photosystem II.

Eva-Maria Krammer and G. Matthias Ullmann *From Computational Biophysics to Systems Biology (CBSB08)*, Edt.: U. H. E. Hansmann, J. Meinke, S. Mohanty, W. Nadler, O. Zimmermann, NIC Series, 265-268, 2008.

Manuscript G

Proton-Transfer Pathways in Photosynthetic Reaction Centers Analyzed by profile Hidden Markov Models and Network Calculations.

Eva-Maria Krammer, Mirco S. Till, Pierre Sebban and G. Matthias Ullmann, *J. Mol. Biol.*, 388, 631-643, 2009.

Appendix

Profile Hidden Markov Models

Proteins of a protein family like the Type II RCs maintain normally the same or a related function. But during evolution, their primary sequence may have diverged from each other. Thus, finding and aligning distant related proteins is often not possible with a simple pairwise or a progressive multiple sequence alignment (MSA). Often residues involved in the catalytic reaction, residues determining the protein structure, and residues binding certain cofactors are highly conserved, whereas the rest of the sequence is more variable leading to a specific conservation pattern for a protein family. Thus, an alternative approach to perform a MSA is to use such statistical features of the conservation pattern of a certain protein family as it is done in the profile Hidden Markov Model (pHMM) method. Before explaining the principle of a pHMM a short introduction to the underlying methods of the Markov chains and Hidden Markov Models (HMM) will be given in the following.

a Exemplary Sequences

0	1	2	3	4	5	6	Position i of the state path Q ($i \in \{0 \dots I\}$)
Start	H	H	H	C	C	Stop	State Path Q
	H	H	H	C	C		Observation Sequence O
	1	2	3	4	5		Position l in the observation sequence O ($l \in \{1 \dots L\}$)

b Markov Chain

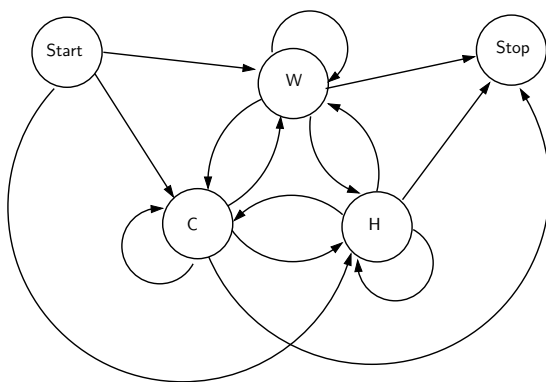


Figure A1: An example of a Markov chain. (a) A weather observation sequence (O) and the corresponding state path (Q) are depicted. Additionally the position in the state path (i) and in the observation sequence (l) is shown. (b) A Markov chain for a weather observation sequence. The states are shown in nodes (circles) and the transitions as edges between the nodes (arrows).

1 Markov Chains and Hidden Markov Models

1.1 Markov Chains

To generate or determine the probability of a sequence O of observations o ($O = \{o_1, \dots, o_l, \dots, o_L\}$) with the length L observing for each position l , a certain observation o_l out of a pool of possible symbols s_o , Markov chains can be used. An example for such a observation sequence would be the weather per day in a particular month as it is depicted in Figure A1a. The length L of the observation sequence would be the days of the month and each observation o_l is a certain weather conditions like *Hot*, *Cold* and *Warm*. The set of K possible symbols s_o is given by $s_o = \{H, C, W\}$ for observations of *Hot*, *Cold* and *Warm* weather.

The statistical model used to create or calculate the probability of a weather observation sequence is the Markov Chain shown in Figure A1b. In the Markov Chain, a set of N states and transition probabilities a_{nm} between them are used for the representation of the observation sequence. In the weather example, the N states are given by a state for each of the observation symbols $s_o = \{H, C, W\}$ and a special start and stop state, leading to a total of five states (see Figure A1b). Thus, the set of state symbols s_q is given by ($s_q = \{H, C, W\}$). In the Markov Chain the sets of observation symbols s_o and of state symbols s_q are equivalent. However, they are introduced here as different variables, since they differ from each other in the following statistical models (HMM and pHMM). Transitions are allowed from the initial, the Start state to all other states apart from the final, the Stop state, from each state apart from the Start state to the Stop state, and from all states representing any symbol of the set of symbols s_o to all states representing any symbol of the set of symbols s_o (see Figure A1b). Other transitions (like from the State to the Stop state) are not allowed by the model.

By the statistical model, a state path Q ($Q = q_0, \dots, q_i, \dots, q_I$) through the model can be created with the length I . Since transitions from a state representing any symbol of the set of symbols s_o to itself are allowed (see Figure A1b), the length of the state path Q can be higher than the number of states N in the model. The first obtained state q_0 and the last obtained state q_I in the state path Q are the Start and Stop states, respectively. The other obtained states q_i can be any of the possible state symbols s_q ($s_q = \{H, C, W\}$). Although both, the state n in the statistical model and the obtained state q_i are named state, they describe different properties: q_i is one of the obtained states in the output of the model, the state path, whereas n is one of the internal states of the model. However, each observation sequence O can be represented by a state path Q through the model (see Figure A1a).

For the each of the previously described transitions, transition probabilities a_{nm} are defined, giving the probability to get from state n to state m in the model (i. e. that a Hot day is followed

by a Cold day). Thus, a Markov chain is given by a set of $N - 2$ states (in the used example Hot, Cold and Warm), a special Start and Stop state and a transition probability matrix A :

$$A = \begin{bmatrix} a_{12} & \dots & a_{1N} \\ \vdots & a_{nm} & \vdots \\ a_{N2} & \dots & a_{NN} \end{bmatrix} \quad (1)$$

where each a_{nm} represents the probability of moving from state n to state m in the Markov chain. The probabilities a_{n1} from each state n back to the Start state ($n=1$) are not included in the transition probability matrix, since it is by definition not allowed in the model to go back to the Start state. The transitions a_{Nm} and a_{NN} are zero since the transition from the Stop state to each other state n and the transition from the Stop to the Stop state are by definition not allowed in the model. Moreover the sum of all transition probabilities ($\sum_{m=2}^N a_{nm}$) from a certain state n (n is not the Stop state) to all states must be one¹:

$$\sum_{m=2}^N a_{nm} = 1. \quad (2)$$

The statistical Markov chain can be used to create a state path through the model, which resembles an observation sequence O and is named state path Q . In a Markov chain the observation sequence O and the state path Q are the same, but in the later described later described HMM and pHMM approaches, Q is different to the observation sequence O . Thus, the state path Q is already used here to explain the concepts of a Markov chain. The probability $P(Q)$ of an certain state path Q ($Q = \{q_0, \dots, q_i, \dots, q_I\}$), which resembles an existing observation sequence O , can be calculated as:

$$P(Q) = P(q_I, q_{I-1}, \dots, q_i, \dots, q_0) \quad (3)$$

$$= P(q_I | q_{I-1}, \dots, q_i, \dots, q_0) P(q_{I-1} | q_{I-2}, \dots, q_i, \dots, q_0) \dots P(q_0) \quad (4)$$

where $P(a)$ stands for the probability of the event a and $P(a|b)$ defines the probability of observing a in dependence of b . The first-order Markov chain assumes that the obtained symbol q_i at the position i of the state sequence Q (like for example the symbol H at position two in the state path in Figure A1a) just depends on the previous obtained state $q_{(i-1)}$ at the position $i - 1$ and not on the obtained state at any other position q_i in the state path. In other words, the probability of obtaining the state n at the position i leading to the obtained state q_i depends not on all other obtained states ($P(q_i | q_I \dots q_0)$) but only on the previous symbol q_{i-1} ($P(q_i | q_{i-1})$).

$$P(q_i | q_I \dots q_0) = P(q_i | q_{i-1}) \quad (5)$$

This assumption is also named first-order Markov assumption, since it means that the probability of a particular state is only dependent on the previous state. In the weather example, the first-order Markov assumption means, that the weather of a certain day only depends on the weather the day before, but not on the weather of all other days of the month. The first-order Markov assumption is a key property of Markov chains, HMMs and pHMMs. Using the first-order Markov assumption (see eq. 5), the probability of a certain observation sequence ($P(Q)$; see eq. 3) can be easily calculated.

Each transition probability a_{nm} of the transition probability matrix A can be estimated by dividing the number of all transitions from state n to state m through the number of all transitions from state n to all.

$$a_{nm} = \frac{\text{expected number of transitions from state } n \text{ to state } m}{\text{expected number of transitions from state } n} \quad (6)$$

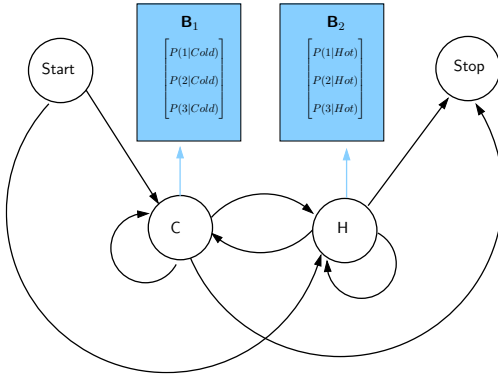
1.2 Components of Hidden Markov Models

A Markov Chain is useful to describe an observable system like the weather in the used example. But an observable system can also depend on a not-observable system. Extending the previously used weather example, one could imagine, that a ice-man selling ice-cream inside a shopping center (for example one, two or three hundred ice-cream portions per day) and is not able to see the weather outside (could be Hot or Cold in the example used here). The observation sequence O with the length L is now given by the amount of sold ice-cream portions observation o_l at each day/position l in the sequence, where each o_l is one of the K possible symbols \mathbf{s}_o ($\mathbf{s}_o = \{1, 2, 3\}$). Since the amount of ice-cream portions sold per day does depend on the weather, the hidden state path Q ($Q = \{q_0, \dots, q_i, \dots, q_I\}$) of the length I is given by the corresponding weather conditions (Hot and Cold), where at each position i in the hidden state path a certain state q_i is obtained out of a set of N states inside the statistical model. The N states of the model are a special Start and Stop state, and a state for Hot and a state for Cold weather. Thus the set of state symbols s_q is given by $\{C, H\}$ ¹. An observation sequence O and state path Q as well as the HMM for the weather/ice-cream example is depicted in Figure A2.

Thus, the statistical model of a HMM is determined by a set of $(N - 2)$ hidden states, a special hidden Start and hidden Stop state, a transition or emission probability matrix A , and for each hidden state n except the start and stop state, a probability vector B_n ($B_n = \{b_{n1}, \dots, b_{nk}, \dots, b_{nK}\}$) consisting of the probability for each of the K different possible observation symbols in the set of observation symbols s_o . All emission vectors are stored in a matrix, the

a Exemplary Sequences

0	1	2	3	4	5	6	Position i of the state path Q ($i \in \{0 \dots I\}$)
Start	H	H	H	C	C	Stop	Hidden State Path Q
<hr/>							
	↓	↓	↓	↓	↓		Observation Sequence O
	3	3	2	1	2		
	1	2	3	4	5		Position l in the observation sequence O ($l \in \{1 \dots L\}$)

b Hidden Markov Model**c Training Sequences**

Month	1.	2.	3.	4.	5.	...
February	1	1	1	2	1	...
March	1	1	2	1	1	...
April	1	3	3	2	1	...
May	3	1	1	1	3	...
June	1	3	2	3	3	...
July	3	3	1	3	3	...
August	2	3	3	3	3	...
September	3	2	3	1	1	...

Figure A2: An Hidden Markov Model for the weather/ice-cream example. (a) An observation sequence O and the corresponding hidden state path Q are depicted. (b) The HMM for the weather/ice-cream example; observation probability vectors B_1 and B_2 are shown as filled light blue squares; the hidden states and the transitions between them are shown as nodes and arrows. (c) A training set for the construction of a HMM is depicted.

observation likelihood or emission probability matrix B :

$$B = \begin{bmatrix} b_{11} & \dots & b_{1K} \\ \vdots & b_{nk} & \vdots \\ b_{(N-1)1} & \dots & b_{(N-1)K} \end{bmatrix} \quad (7)$$

where b_{nk} is the emission probability corresponding to the hidden state n and the observation with the symbol k of the possible K symbols. Similar to the estimation of the transition probabilities a_{nm} , the observation probabilities b_{nk} can be estimated by dividing the number of times observing the symbol k when being in the hidden state n through the number of times being in the hidden state n :

$$b_{nk} = \frac{\text{expected number of times in state } n \text{ and observing symbol } k}{\text{expected number of times in state } n} \quad (8)$$

As for the Markov chain, the HMM is based on the first-order Markov assumption (see eq. 5). In

addition a second assumption, the outcome independence assumption, is made in a HMM¹:

$$P(o_l|q_1...q_i, ...q_I, o_1...o_l...o_L) = P(o_l|q_i). \quad (9)$$

This outcome independence assumption means that the probability for an observation o_l does not depend on all states in the state path and not on all observations in the observation sequence ($P(o_l|q_1...q_i, ...q_I, o_1...o_l...o_L)$) but just depends on the actual hidden obtained state q_i ($P(o_l|q_i)$). For the used example this assumption means, the amount of ice-cream portions sold just depend on the weather today, but not on the weather yesterday or on the weather any day before.

1.3 Construction of a Hidden Markov Model.

A HMM is constructed from a known set of observation sequences S ($S = \{O_1, \dots, O_x, \dots, O_X\}$) consisting of X observation sequences and a defined symbol set (s_o and s_q) for the observations and hidden states in the construction or training procedure. In the here used example, the set of observation sequences S would be the amount of ice-cream portions sold per day for several months X (for each month x an observation sequence O_x is given; see Figure A2c), the hidden states would be the different weather conditions ($s_q = \{C, H\}$) and the observations would be the different amounts ice-cream portions sold per day ($s_o = \{1, 2, 3\}$). Using this information, the transition and emission probability matrices can be calculated. The standard algorithm is the forward-backward or Baum-Welch Algorithm², which trains both, the transition (A ; eq. 1) and emission (B ; eq. 7) probabilities. This algorithm uses a set model λ^p , which includes an initial transition and emission matrix and an initial state distribution π to calculate iteratively the accurate probabilities of A and B . After the training, the resulting model λ with accurate probabilities for A and B can be used to compute the likelihood of a new sequence of observations as well as finding the best hidden state sequence Q to a given sequence of observations O . The training procedure using the Baum-Welch algorithm as well as the other methods are not described in more detail here, since other algorithms are preferred for the construction and utilization of a pHMM.

2 A Profile Hidden Markov Model

HMMs can also be used to perform a MSA of a protein family. For this purpose special HMMs were developed, the pHMMs. The idea behind a pHMM is, that the specific information about each position in the multiple alignment of a complete protein family is stored in the emission and transition probabilities using HMM theory. In the following the statistical model of a pHMM is explained based on an existing alignment as it is shown in Figure A3. Several different implementations of pHMMs exist. In the work presented here, the program package HMMER⁴ is used and

Partial Alignment		N-Terminus	
L subunit	{	gi 146278067	-----.....EM-RTP-DHEDT--FFRDLVG--YSI---GTL-GI ² HRLGL...
		gi 77463826	-----.....EM-RTP-DHEDT--FFRDLVG--YSI---GTL-GI ² HRLGL...
		gi 6690736	-----.....EM-RTP-DHEDT--FFRDLVG--YSI---GTL-GI ² HRLGL...
		gi 27530687	-----.....EM-RTP-DHEDT--FFRDLVG--YSI---GTL-GI ² HRLGL...
		gi 132180	-----.....TM-RTP-DHEDT--YFRDLVG--YSV---GTL-GI ² HRLGL...
		gi 85703519	-----.....EM-KTP-DHEDT--FFRDFIG--YSV---GTL-GI ² HRLGL...
		gi 149202073	-----.....EM-KTP-DHEDT--FFRDFIG--YSV---GTL-GI ² HRLGL...
		gi 21328651	-----.....TI-GTP-DHEDT--YFRDLIG--YSI---GPL-GI ² HRLGL...
D1 subunit	{	gi 150251436	-----.....GETYNI-VAAHG--YFGRLLFQYASF--NNSR-SLHFFLA...
		gi 119492284	-----.....EETYNI-VAAHG--YFGRLLFQYASF--NNSR-SLHFFLA...
		gi 126660350	-----.....EETYNL-VAAHG--YFGRLLFQYASF--NNSR-SLHFFLA...
		gi 77744799	-----.....EETYNI-VAAHG--YFGRLLFQYASF--NNSR-SLHFFLA...
		gi 15529753	-----.....EETYNI-VAAHG--YFGRLLFQYASF--NNSR-SLHFFLA...
		gi 12620335	-----.....EETYNI-VAAHG--YFGRLLFQYASF--NNSR-SLHFFLA...
		gi 12620386	-----.....EETYNF-VAAHG--YFGRLLFQYASF--NNSR-SLHFFLA...
		gi 113954773	MSWIFPEETYNI-VAAHG--YFGRLLFQYASF--NNSR-SLHFFLA...
gi 37521275	-----.....EETYNI-SAAHG--YISRLVFQYAFW ³ CANSR-SLHFIMA...		
<hr/>			
Position in the Alignment i		123456.....1234567890 ² 1234567890 ³ 1234567890 ⁴ 1234567890 ⁵ ...	
Example Hidden State Path Q		IIIIII.....MMDMMIMMMMIIMMMMMMIIMMMIIMMMMDMMMMMM...	
Position in the Hidden State Path i		123456.....1234567890 ² 1234567890 ³ 1234567890 ⁴ 1234567890 ⁵ ...	

Figure A3: Part of a multiple sequence alignment. The depicted partial MSA is taken from an alignment of the L and D1 subunits of type II RC proteins. The position i in the depicted MSA does not represent the numbering of the complete MSA. An example of an ungapped part, an insertion, a deletion, and an insertion before the N-Terminal region are framed red, black, green and underlined in yellow, respectively. Each sequence can be identified by the shown GenBank Identifier³. In addition to the partial alignment, an exemplary hidden state path Q and the position i in the hidden state path is depicted.

thus, in addition to the main concepts of pHMMs, the description will focus on the algorithms used by this program package.

2.1 Definition of the profile Hidden Markov Model.

A MSA S consists of X aligned sequences and can thus be described by an set of observation sequences as it was done in the weather/ice-cream example (see Figure A2). Each row of the MSA S ($S = \{O_1, \dots, O_x, \dots, O_X\}$) is an observation sequence O_X and represents a 'single aligned sequence'. Such a 'single aligned sequence' would be for example the aligned sequence of the L subunit with the GenBank Identifier³ gi|146278067 in Figure A3. Each observation o_i can be one of the twenty ($K=20$) aminoacids represented in their one letter code. Thus, the set of observation symbols s_o is given by the one letter code of the aminoacids and the symbol (-) for

a gap ($s_q = \{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y, -\}$)^{1,4}. The hidden states, which were in the previous example the different possible weather events, are now named Match, Deletion, and Insertion state and are represented with the symbols $s_q = \{M, D, I\}$. Additionally a special Start and Stop state exist in the pHMM approach. In the previously explained Markov chains and HMM each symbol of the set of symbols s_q (like Hot for the weather/ice-cream example; see Figure A2) forms a single state occurs only once in the build model. In contrast, in a pHMM the three symbols $\{M, D, I\}$ for of hidden states (Match, Insertion and Deletion) occur more than once in the model. The statistical model thus consists of N hidden states and T positions, where at the first and the last positions, the Start and Stop state is placed, respectively. At any position t ($t \neq \{0, T\}$) in the model all the three hidden state types can occur apart from the first (the Start state) and the last (the Stop state). In the following, the three types of hidden states, the building of a pHMM and the utilization of a scoring matrix for the emission probabilities will be explained.

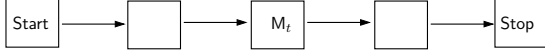
2.2 Organization of a profile Hidden Markov Model

Each position in a sequence O_x of a MSA S , such as the MSA shown in Figure A3, can be described by either a match or a gap in respect to the same position in the complete alignment. Gaps can than be subdivided into insertions and deletions. A match is defined as a position where in all sequences an aminoacid is found. A deletion in a certain sequence occurs when the aminoacid sequence is shifted by one, meaning that this aminoacid is missing compared to the other sequences. An insertion means that a portion of the sequence O_x does not match to anything in the model. Examples for deletion, insertion and matches are marked in Figure A3.

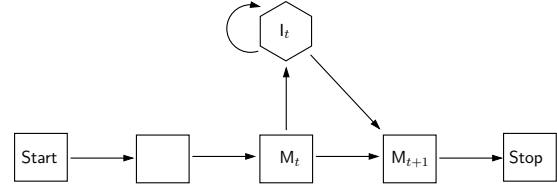
In the pHMM approach, to each of these three different types of positions in an MSA a state symbol is assigned. Gaps tend to line up with each others leading to ungapped blocks, without any deletions or insertions (as marked in Figure A3 at the position l 25-31). Each position in such an ungapped regions is described by the hidden state symbol M and is named Match state. Thus, the pHMM for an ungapped region with the length p would consist of p Match states and transitions $M_t \rightarrow M_{t+1}$ between them. Thus, the length T of an existing pHMM would increase due to p new Match states. A pHMM consisting of Match states is shown exemplary in Figure A4a.

For an insertion area with the length p in the alignment (as marked in Figure A3 at the position l 37/38), a single Insertion state is added in the pHMM as it is depicted in Figure A4b. In principle, in the MSA each match could be followed by an insertion and thus to each Match state an Insertion state is assigned in the model. Consequently, the Insertion states do not change the length T of previously existing statistical model. For an Insertion state I_t in the statistical

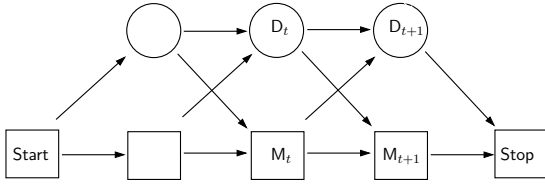
a Match States



b Match and Insert States



c Match and Deletion States



d Complete pHMM

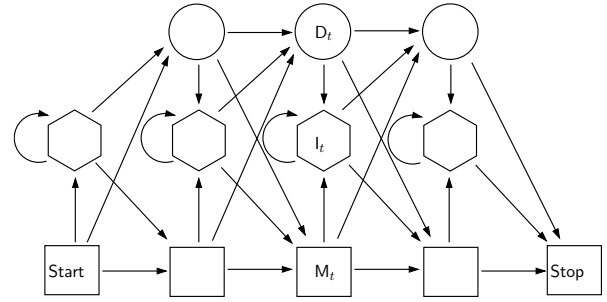


Figure A4: Stepwise construction of a profile Hidden Markov Model. (a) The first model consists of only Match states. Either an Insertion state (b) or Deletion states (c) are added. (d) The complete pHMM consisting of Match, Insertion and Deletion states. The Match, Insertion and Deletion states are shown as squares, diamonds, and circles, respectively. For clarity, only the hidden states and the transition probabilities are shown. Additional to each Match and Insertion state, emission probabilities for each aminoacid exist. The picture is modified from Reference A4.

model additional transitions for going to the Insertion state $M_t \rightarrow I_t$ and for going from the Insertion state to the next Match state $I_t \rightarrow M_{t+1}$ are needed. Since the insertion region can be more than a single aminoacid (for example in Figure A3 at the positions l 37/38, two aminoacids form an insertion), there must also be the possibility to get from an Insertion state into the same Insertion state. Thus, an additional transition ($I_t \rightarrow I_t$) is needed for looping the Insertion state⁵.

In the pHMM approach deletions of a MSA (see Figure A3 at the position l 3) are represented by Deletion states. In the MSA each match could be principally be followed by a deletion, and thus a Deletion state has to be assigned to each Match state in the model (see Figure A4c). Such an assignment does not lead to an increase of the length T . Also for the Deletion state, additional transitions are needed, namely for going from a Match state to a Deletion state ($M_t \rightarrow D_t$), for going from a Deletion state to the next Deletion state ($D_t \rightarrow D_{t+1}$) and from a Deletion state to the next Match state ($D_t \rightarrow M_{t+1}$).

It is also possible that a sequence of aminoacids is fused to the N-Terminus of the protein

of a certain species during evolution. Such a fusion would be represented by an Insertion state in the pHMM. This special Insertion state cannot be assigned to any Match state and is thus assigned to the Start state. Consequently an additional transition from the Start state to the Insertion is needed ($Start \rightarrow I_0$). It is also possible, that the C-Terminus of a certain protein is missing (represented by a Deletion state in the pHMM) or the protein's chain is elongated (represented by an Insertion state in the pHMM). Consequently, also transitions from the last Deletion and Insertion state to the Stop state ($I_{T-1} \rightarrow Stop$ and $D_{T-1} \rightarrow Stop$) are needed. In the resulting pHMM the Insertion states are added between the Deletion and Match states and also transitions $D_t \rightarrow I_t$ between them, even they are very improbable. The resulting complete pHMM is depicted in Figure A4d.

2.3 Construction of a profile Hidden Markov Model

The construction of a pHMM using an existing MSA is not trivial, since the length T of the model λ can differ, depending on which columns in the alignment are assigned to Insert states. An example for a MSA and different assignments of columns are shown exemplary in Figure A5. Here four different models (λ_1 , λ_2 , λ_3 and λ_4) are depicted. Model λ_2 has the highest number of positions ($T=7$), since all positions in the alignment are assigned to Match states in the model. In contrary, Model λ_3 is the shortest model since the position two, three and four in the alignment are assigned to an Insertion state and thus, the pHMM consists of just four positions. Model λ_1 and λ_4 have the same number of positions ($T=5$), however, the assignment of Insertion and Match states differs between the two models. Using model λ_1 would lead to other probabilities for a certain set of observation sequences compared to model λ_4 . But model λ_1 might represent the existing MSA better than the model λ_4 . Thus, the most important task for constructing a pHMM is to find the optimal model λ^* , which is the model, best representing the existing MSA.

For an MSA with the length L there are in principle 2^L possible pHMMs, and thus the main problem is to find the optimal model⁵. Based on an existing MSA, there are in principle three approaches to find the optimal model: manual construction, heuristic construction and *MAP* (maximum *a posteriori*) construction⁵. Manual construction means, that the user marks each of the columns in a MSA as Match, Deletion or Insertion state by hand. In a heuristic construction, a pre-set rule determines whether a column should be marked as a Match state. Such a rule could be for example, that the portion of gaps in the column is under a certain threshold. The best way is to use the *MAP* algorithm, which uses the Bayes rule to find the optimal model λ^* . The Bayes rule defines, that the probability $P(\lambda|S)$ of a certain model λ in dependence of the used MSA S is given by the probability $P(S|\lambda)$ that all sequences are build by the model divided by

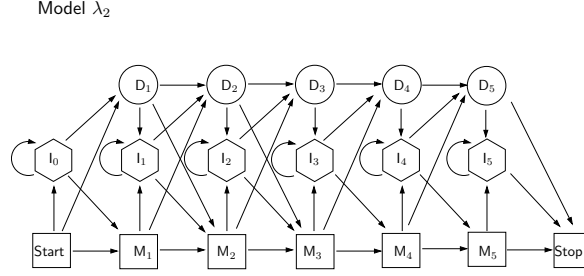
a Multiple sequence alignment

```

gi|146278067      I - - - G
gi|77463826       I - - - G
gi|21328651       F - - N N
gi|119492284      F - - N N
gi|37522225       F W G A N
Position l        1 2 3 4 5

```

c Architecture of the models



b Different Possible Models

Position l	1	2	3	4	5
Model λ_1	M	I	i	M	M
Position t_1	1		2	3	
Model λ_2	M	D	D	M	M
Position t_2	1	2	3	4	5
Model λ_3	M	I	i	i	M
Position t_3	1				2
Model λ_4	M	D	I	i	M
Position t_4	1	2		3	

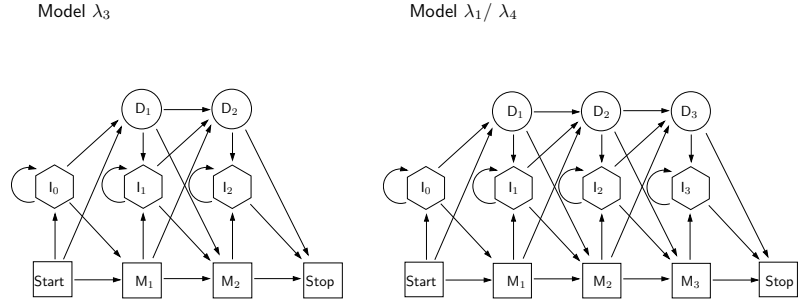


Figure A5: Finding the optimal model. A part of the model construction of a pHMM for the L/D1 subunit of the Type II RC is shown. (a) An part of the MSA is shown. The position l in the MSA are marked. (b) The assignment of the Match, Deletion and Insertion states to the given MSA in four different models. Additionally to each model, the position t of the model is marked. (c) The architecture of different possible models (λ_1 , λ_2 , λ_3 , and λ_4) are depicted. Model 1 and 4 have the same architecture.

the probability of the MSA ($P(S)$) and the model ($P(\lambda)$).

$$P(\lambda|S) = \frac{P(S|\lambda)P(\lambda)}{P(S)} \quad (10)$$

The probability $P(S|\lambda)$, also named likelihood indicates how good the actual model λ is. It defines the probability, that each sequence O_x of a set of observations sequences S (which would be any MSA) consisting of X sequences ($O_1, \dots, O_x, \dots, O_X$) is generated by the model λ ⁶.

$$P(S|\lambda) = \prod_{x=1}^X P(O_x|\lambda) \quad (11)$$

meaning that the probability ($P(S|\lambda)$) that all sequences of the set S are build by the used model λ is given the probabilities ($P(O_x|\lambda)$) that each of the 'single aligned sequences' O_x is build by the used model λ . The probability $P(S)$ in eq. 10 can also be seen as the probability that all sequences of a MSA occur randomly. This probability is a normalization constant and in case of finding the optimal model λ^* it can be left out, changing eq. 11 to:

$$P(\lambda|S) = P(S|\lambda)P(\lambda). \quad (12)$$

The probability $P(\lambda|S)$ is also named the *posterior* probability, whereas the probability $P(\lambda)$ is named *prior* probability. The *prior* probability is used to sort out models, which are wrong or uninteresting. The *prior* probability is determined by the probability of the model structure ($P(\lambda_S)$) and by the probability of the adjustable parameters ($P(\lambda_A|\lambda_s)$):

$$P(\lambda) = P(\lambda_S)P(\lambda_A|\lambda_s). \quad (13)$$

The probability $P(\lambda_S)$ is used to penalize model with many positions (with a high T). The adjustable parameters λ_A are given by the transition and emission matrices A and B ^{7,8}. To estimate the transition and emission matrices in order to calculate the probability $P(\lambda_A)$, different methods exists. One possibility is be to assume that each transition and emission have same probability. Another possibility is to determine the transition probability a_{nm} by transforming eq. 6 to:

$$a_{nm} = \frac{c_{nm} + \alpha_{nm}}{\sum_1^m c_{nm} + \alpha_{nm}} \quad (14)$$

where c_{nm} is the number of counted transitions from the hidden state n to the state m and α_{nm} is the Dirichlet prior term. The Dirichlet *prior* terms (used commonly in Bayesian statistics) for emission probabilities were determined experimentally for for the transitions between the possible states (the Match, Insertion, Deletion, Start and Stop state) using existing pHMMs of a wide variety of protein sequences in the past. Moreover, Dirichlet *prior* terms were also determined experimentally for each possible aminoacid from existing MSAs and pHMMs of a wide variety of protein sequences in the past^{5,6,9–11}. Since the Dirichlet *prior* terms are determined from existing MSAs and pHMMs, they include already information about how probable it is to find a certain aminoacid at a certain position in the MSA. By using Dirichlet terms also the possibility of overfitting the model is reduced⁵.

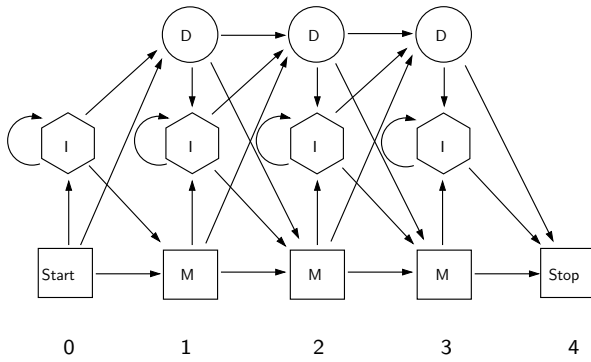
For the optimal model λ^* , the *posterior* probability $P(\lambda|S)$ is maximal. The *MAP* algorithm starts at an arbitrary starting point and tries to find the local maximum of $P(\lambda|S)$ by iteratively re-estimating the model. Once having found the optimal model, the emission and transition probabilities could be calculated as for the HMM (see eq. 6 and eq. 8) by simply counting up the number of times each transition or emission is used (see Figure A6). However, the calculation of these probabilities is done more sophisticated and will be explained in the next following section.

In addition to the method presented here to construct a pHMM from a existing accurate alignment, there is also an approach to build the model using the unaligned sequences. This is done with the Baum-Welch algorithm². However, this way of constructing a pHMM is not used in the work presented here and will thus not be described in further detail.

a Multiple sequence alignment

gi 146278067	I	-	-	-	G
gi 77463826	I	-	-	-	G
gi 21328651	F	-	-	N	N
gi 119492284	F	-	-	N	N
gi 37522225	F	W	G	A	N
Position l	1	2	3	4	5
Position l	1	1	1	2	3

b pHMM architecture



c Observed emission/transition counts

		model positions			
		0	1	2	3
match emissions	A	-	0	1	0
	I	-	2	0	0
	F	-	3	0	0
	G	-	0	0	2
	N	-	0	2	3
insert emissions	G	0	1	0	0
	W	0	1	0	0
	M-M	5	2	3	5
state transitions	M-I	0	1	0	0
	M-D	0	2	2	0
	I-M	0	1	0	0
	I-I	0	1	0	0
	I-D	0	0	0	0
	D-M	-	0	2	0
	D-I	-	0	0	0
	D-D	-	0	0	0

Figure A6: Calculation of the emission and transition probabilities. A part of the model construction of a pHMM for the L/D1 subunit of RC proteins is shown. (a) In the shown partial MSA, the three columns marked, which are assigned as Match states in the model. (b) The assignment of transition and emission counts based on the MSA. (c) The used pHMM architecture. The picture is modified from Reference A4.

2.4 Calculation of Emission Probabilities: Scoring Matrix and Gap Penalties

Several of the explained approaches to construct and utilize pHMMs are implemented as Dynamic programming algorithms. Commonly, dynamic programming algorithms are used to solve optimization problems over a N-dimensional search space. Therefore the large problem of finding the optimum is broken down into several sub-problems which can be solved. After solving a sub-problem incrementally, the next sub-problem can be solved without changing the solution of the other sub-problems leading in the end to a solution of the large problem. Dynamic programming algorithms are also used in pairwise sequence alignment approaches such as for example in the Needleman-Wunsch algorithm¹² or in the Smith-Waterman algorithm¹³ or in progressive MSA approaches such as ClustalW[?]. In pairwise alignment, progressive MSA and pHMM methods, the dynamic programming algorithms are based on a scoring matrix and a gap penalty function $\gamma(g)$.

To define a scoring matrix for emission probabilities, first one needs to define the probability

$P(O_x|\lambda)$ of a ungapped block of a sequence O_x of the length p (all match states) starting at the position p_o in the pHMM λ as⁵:

$$P(O_x|\lambda) = \prod_{t=p_o}^p b_{tk} \quad (15)$$

where b_{tk} is the emission probability of a certain aminoacid k out of a set of observation symbols s_o at the position t in the model λ . Normally one determines the probability of this sequence O_x in relation to a random model R_k , since it ensures that the emission probabilities resemble characteristics of the protein family and not the probability of the aminoacid to occur in any protein sequence. The random model R_k assumes, that the probabilities of the aminoacid at the same position in two sequences is just the product of the probabilities of each aminoacid⁵. Thus, the probability $P(O_x|\lambda)$ can be rewritten as:

$$P(O_x|\lambda) = \prod_{t=1}^r \frac{b_{tk}}{R_k}. \quad (16)$$

For $P(O_x|\lambda)$ the so called log-odd score $s(t, k)$ can be defined:

$$s(t, k) = \sum_{t=r_0}^r \log \frac{b_{tk}}{R_k}. \quad (17)$$

Instead of probabilities Log-odd scores are used for the scoring matrix and later for the gap penalty function $\gamma(g)$, since logarithmic values reduce the run-time of the used algorithms significantly by replacing multiplications by additions. The scores $\log \frac{b_{tk}}{R_k}$ are defined for each position t and each aminoacid k and fill a matrix. This matrix is very similar to the scoring matrices used normally in progressive MSA like BLOSUM¹⁴ and PAM¹⁵. These usual scoring matrix are given by the aminoacid exchange probabilities estimated from a huge variety of different protein sequences with a certain identity. In contrast the scoring matrix $S(t, k)$ of the pHMM, is given by the probabilities of observing the aminoacid k at position t in the model λ . Thus the scoring matrix is specific for the a certain protein family. Such a scoring matrix is named position specific score matrix (PSSM). The different scoring matrices in progressive MSA and pHMM are a central reason, why pHMMs can align sequences with very small identities where progressive MSA algorithms fail^{5, 16, 17}.

However, a large number of sequences is required for constructing the pHMM. Otherwise some transitions or emissions might not be observed in the training alignment. The probabilities of these transitions and emissions would therefore set to zero, meaning they would not be allowed for further alignments. Thus pseudo counts are added to the observed frequencies⁵. Pseudo counts can be calculated using different methods such as simple pseudo counts, Dirichlet mixtures, substitution matrix mixtures and estimation based on an ancestor. The simplest method is to use the Laplace's rule to determine the pseudo counts. As a Simple Pseudo count method, in the

Laplace's rule a pseudo count of 1 is used for each aminoacid⁵. The PSSM is than calculated as described above.

For an insert region of the alignment, it is assumed, that the probability for each aminoacid is described appropriate by a random model, in which for example the probability of each amino acid is set to $\frac{1}{20}$ ⁹. Normally the previous described random model R_k is used here as well⁵. The costs of an insertion with the length r starting at the position i in the state path Q and at the corresponding position t in the model can be defined as:

$$\gamma(g) = \log a_{M_t I_t} + \log a_{I_t M_{t+1}} + (r - 1) \log a_{I_t} \quad (18)$$

where $a_{M_t I_t}$ is the transition probability from the hidden match state M to the insertion state I at the position j in the pHMM, $a_{I_t M_{t+1}}$ is the transition probability from the insertion state I_t to the next match state M_{t+1} , and a_{I_t} is transition probability that the insertion state is followed by another insertion state⁵. The costs for an insertion describe the score for the corresponding gap in the alignment. Thus, also the gap penalties $\gamma(g)$ are position-dependent unlike in progressive MSAs, where the gap penalties are position-independent⁹.

2.5 Multiple Sequence Alignment with a profile Hidden Markov Model.

Once a pHMM is build, it can be used to produce a MSA from a set of unaligned sequences. Therefore, first each sequence is aligned to the model using the Viterbi algorithm, and second the individual alignments are merged to the resulting MSA. The Viterbi algorithm finds the most probable path Q^* through the model, computing the best hidden state path for a given observation sequence O_x . The best hidden state path Q^* resembles the correct alignment of the used sequence.

The probability $P(q_0, \dots, q_i, o_1, \dots, o_l | \lambda)$ defines how probable it is to be in the obtained hidden state q_i and observing o_l of the observation sequence O_x after having seen $(l - 1)$ observations (o_1, \dots, o_{l-1}) of the sequence O_x and following the hidden path Q ($Q = q_0, \dots, q_{i-1}$) in the pHMM. To find the most probable path Q^* from the Start state to the position n through the model λ , one has to maximize the probability $P(q_0, \dots, q_i, o_1, \dots, o_{l-1} | \lambda)$. Therefore so called Viterbi trellis $v_i(o_l)$ are introduced:

$$v_i(o_l) = \max_{q_0, \dots, q_{i-1}} P(q_0, \dots, q_{i-1}, o_1, \dots, o_l, q_i | \lambda) \quad (19)$$

Reformulating eq.[19] leads to:

$$v_i(o_l) = \max_{q_0, \dots, q_{i-2}} P(q_0, \dots, q_{i-2}, o_1, \dots, o_{l-1}, q_{i-1} | \lambda) P(o_l, q_i | \lambda) \quad (20)$$

$$= \max_{q_0, \dots, q_{i-1}} v_{(i-1)}(o_{l-1}) \dot{a}_{(i-1)i} \dot{b}_i k \quad (21)$$

Moves

Observations

	Start	V	G	Y	...
<i>Start</i> / <i>I</i> ₀	$V(0) = 0$	$V_0^I(1) = \log a_{SI_0}$	$V_0^I(2) = \log a_{I_0I_0}$	$V_0^I(3) = \log a_{I_0I_0}$...
<i>M</i> ₁	forbidden	$V_1^M(1) = V(0) + \log a_{SM_1}$
<i>I</i> ₁	forbidden	$V_1^I(0) = \log \frac{b_{1V}}{R_V} + \max \begin{cases} V_1^M(0) + \log a_{M_1I_1} \\ V_1^I(0) + \log a_{I_1I_1} \\ V_1^D(0) + \log a_{D_1I_1} \end{cases}$
<i>D</i> ₁	$V_1^D(0) = \log a_{SD_1}$	$V_1^D(1) = V_1^I(0) + \log a_{I_0D_1}$
<i>M</i> ₂
<i>I</i> ₂
<i>D</i> ₂
⋮	⋮	⋮	⋮	⋮	⋮

Hidden States

Figure A7: Table of Viterbi Trellis. The principle filling the Viterbi trellis V in a table is depicted for an example sequence VGY . The hidden state $Start$ is shown as S in the table. For more details see text.

meaning that the Viterbi trelli of going from the observed state $i - 1$ to observed state i and observing o_l is just the maximum of the transition probability ($a_{(i-1)i}$) for moving from the observed state $i - 1$ to observed state i , of observing the symbol k at the position i (b_{ik}) and the Viterbi trelli of the position $i - 1$. In other words, knowing already the most probable path from the first to the $(i - 1)$ th position in the state path Q , the next most probable step in the path is easy to calculate^{1,5}. Since in a pHMM three different types of hidden states exist, these three states have also to be taken into account. Moreover, one normally uses the log-odd values of the Viterbi trellis $V_i(o_l)$. Assuming that one obtains a Match state at position i in the hidden state path Q and observing the symbol k at the observation o_l , one can define the Viterbi trelli V_i^M as:

$$V_i^M(o_l) = \log \frac{b_{ik}}{R_k} + \max \begin{cases} V_{i-1}^M(o_{l-1}) + \log a_{M_{i-1}M_i} \\ V_{i-1}^I(o_{l-1}) + \log a_{I_{i-1}M_i} \\ V_{i-1}^D(o_{l-1}) + \log a_{D_{i-1}M_i} \end{cases} \quad (22)$$

Equivalent the Viterbi trellis $V_i^I(o_l)$ are defined for obtaining at position i in the state path an Insertion state:

$$V_i^I(o_l) = \log \frac{b_{ik}}{R_k} + \max \begin{cases} V_{i-1}^M(o_{l-1}) + \log a_{M_{i-1}I_i} \\ V_{i-1}^I(o_{l-1}) + \log a_{I_{i-1}I_i} \\ V_{i-1}^D(o_{l-1}) + \log a_{D_{i-1}I_i} \end{cases} \quad (23)$$

and for a obtaining a a Deletion state at position i in the state path Q :

$$V_i^D(o_l) = \max \begin{cases} V_{i-1}^M(o_{l-1}) + \log a_{M_{i-1}D_i} \\ V_{i-1}^I(o_{l-1}) + \log a_{I_{i-1}D_i} \\ V_{i-1}^D(o_{l-1}) + \log a_{D_{i-1}D_i} \end{cases} \quad (24)$$

In the last equations, the transition probabilities such as for example $a_{M_{i-1}I_i}$ stand for the transition probabilities for the to the position i in the state path corresponding position t in the model λ . Using the equations for the Viterbi trellis V_i^M , V_i^I , and V_i^D a table can be filled having in the

a Example Sequences

VGYSIGTL
VGYSIGTL
TFQYASFNNR
IFQYASFNNR
IFQYASFNNR

b Viterbi Calculation

		Moves		Observations					
		Deletion	Match	Insertion	Start	V	G	Y	...
Hidden States	Start/ I_0		0.00		-0.70	-0.80	-1.10	...	
	M_1	-		0.00	-1.29	-0.41	...		
	I_1	-	-0.30		-0.56	-1.37	...		
	D_1	-0.78	-1.10		-1.20	-1.02	...		
	M_2	-	-0.92		-0.34	-0.93	...		
	I_2	-	-1.23		-0.58	-0.45	...		
	D_2	-1.23	-1.45		-1.32	-0.90	...		

c Most probable path for each sequence

Position	1	2	insert	3	4	5	6	7	8
	V	G	Y	S	I	-	G	T	L
	V	G	Y	S	I	-	G	T	L
	T	F	QYA	S	F	N	N	S	R
	I	F	QYA	S	F	N	N	S	R
	I	F	QYA	S	F	T	N	S	R

d Resulting Alignment

VGY..SI---GTL
VGY..SI---GTL
TFQYASF--NNSR
IFQYASF--NNSR
IFQYASF--TNSR

Figure A8: Computation of a Multiple Sequence Alignment using an existing profile Hidden Markov Model. (a) Sequences, which should be aligned to an existing pHMM. (b) Calculation of the Viterbi trellis. The resulting most probable path Q^* is marked by red arrows and frames. (c) The most probable path Q^* for each sequence through the pHMM. (d) The resulting alignment. Dots represent, that the insertion at this position is shorter than the maximal one. Dashes stand for a deletion.

columns the observation sequence O_x and in the rows the possible hidden states (see Figure A7).

Since for the Start state, some hidden states cannot be obtained (like for example M_0), the Viterbi equations can be simplified for these states. In Figure A7 such an table of Viterbi trellis is shown. During the filling each cell, a reference to the previous obtained hidden state with maximal probability is set.

The trellis V^{Stop} in the final cell (not depicted in Figure A7) has the best score. The most probable path Q^* is build up in reverse order using the references set in the filling procedure and starting in the end of the matrix (*traceback* procedure). At each step one moves from the current cell (i, l) to one of the cells $(i - 1, l - 1)$ to which the reference is pointing. Thus, one can add the symbol of q_{i-1} to the hidden state path Q^* . *Traceback* algorithms are also used in 'normal' sequence alignments such as in the Needleman-Wunsch algorithm¹², which is used to find the optimal global alignment of two sequences, or in the Smith-Waterman algorithm¹³, which is used to find the optimal local alignment of two subsequences⁵.

In Figure A8 an example of a MSA done with a pHMM is depicted. For each sequence (see Figure A8a) to be aligned, the table of trellis is calculated. After filling up the Viterbi table with the Viterbi trellis and references, the most probable path Q^* is found by back-tracking from the last trellis V^{Stop} in the hidden state path to the first maximal trellis $i = 0$ (see Figure A8b) as explained before. The Viterbi algorithm is repeated for each sequence, resulting in a most probable path of hidden states Q^* for each of the sequences (see Figure A8c). Afterward, the maximum number of inserted residues for each Insertion state is calculated and based on this information, the MSA can be build (see Figure A8d).

2.6 Limitations and Advantages of the profile Hidden Markov Model Approach.

The here described pHMM is used to build MSA of protein families. However, the model has some limitations arising from simplifications made in the model. First, the pHMM approach is linear and thus not able to model higher order correlations between the aminoacids such as hydrogen bonds, electrostatic interaction which are given by the three dimensional arrangement of the aminoacids in the protein. Second, it is assumed, that the probability of an aminoacid sequence can be found by multiplying the probability of the each aminoacid position in the sequence. This assumption is only valid if the probability of each aminoacid is independent of the other aminoacids, which is not always the case. Hydrophobic for example aminoacids often cluster together inside the protein because of their hydrophobicity and thus their positions would not be independent of each other.

However, the previously described limitations are not only limitations valid for the pHMM approach, but also for some of progressive MSA approaches. The described high order correlation and clustering of aminoacids can however be included in the MSA by using structural information of the protein for the MSA calculation like it is for example done in the *stacchato*¹⁸ program of the BioInfo3D program package¹⁹. A pHMM build on such a MSA including structural information is therefore more reliable especially for protein families with low sequence identity²⁰. Such an approach is used in the work presented here. But in contrast to normal progressive MSA

approaches (which uses position-independent scoring matrices), the pHMM uses a PSSM and position dependent gap penalties and can thus store the characteristic conservation pattern of a protein family. Consequently it is possible to align evolutionary related sequences with low sequence identity, for which the normal progressive alignment methods often fail^{16,17,21}.

3. Bibliography

- [1] Jurafsky D and Martin JH (2000): *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice Hall
- [2] Baum LE (1972): An equality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities* 3, 1–8
- [3] Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, and Wheeler DL (2008): GenBank. *Nucl. Acid Res.* 36, D25–D30
- [4] Eddy SR (1996): Hidden Markov Models. *Curr. Opin. Struct. Biol.* 6, 361–365
- [5] Dubin R, Eddy S, Krogh A, and Mitchison G (1998): *Biological sequence analysis*. Cambridge University Press
- [6] Krogh A, Brown M, Mian IS, Sjölander K, and Haussler D (1994): Hidden Markov Models in Computational Biology. *J. Mol. Biol.* 235, 1501–1531
- [7] Stolcke A and Omohundro S (1993): Hidden Markov Model induction by Bayesian model merging. *Advances in Neural Information Processing Systems* 5
- [8] Chu W, Ghahramani Z, Podtelezhnikov A, and Wild DL (2006): Bayesian segmental models with multiple sequence profiles for protein secondary structure and contact map prediction. *IEEE/ACM transactions on computational biology and Bioinformatics* 3, 98–113
- [9] Krogh A, Brown M, Mian IS, Sjölander K, and Haussler D (1993): Hidden Markov models in computational biology: applications to protein modeling. *Technical report UCSC-CRL-93-32 University of Santa Cruz*
- [10] Brown M, Hughey R, Krogh A, Mian IS, Sjölander K, and Haussler D (1993): Using Dirichlet mixture priors to derive hidden Markov models for protein families. *Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology* 44–51. AAAI Press.
- [11] Sjölander K, Karplus K, Brown M, Hughey R, Krogh A, Mian IS, and Haussler D (1996): Dirichlet mixtures: a method for improved detection of weak but significant protein sequence homology. *Comput. Appl. Biosci.* 12, 327–345
- [12] Needleman SB and Wunsch CD (1970): A general method applicable to search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48, 443–453

- [13] Smith TF and Watermann MS (1981): Identification of common molecular subsequences. *J. Mol. Biol.* 147, 813–818
- [14] Henikoff S and Henikoff J (1992): Amino acid substitution matrices from protein block. *Proc. Natl. Acad. Sci. USA* 89, 10915 – 10919
- [15] Dayhoff MO, Schwartz MR, and Orcutt BC (1978): A model of evolutionary change in proteins. *Atlas of Protein Sequence and Structure* 5, 345–352
- [16] Pearson WR and Sierk ML (2005): The limits of protein sequence comparison. *Curr. Opt. Stru. Biol.* 15, 224–260
- [17] Edgar RC and Sjölander K (2004): A comparison of scoring functions for protein sequence alignment. *Bioinformatics* 20, 1301–1308
- [18] Shatsky M, Nussinov R, and Wolfson HJ (2006): Optimization of multiple-sequence alignment based on multiple-structure alignment. *Proteins* 62, 209–217
- [19] Shatsky M, Dror O, Schneidman-Duhovny D, Nussinov R, and Wolfson HJ (2004): BioInfo3D: a suite of tools for structural bioinformatics. *Nucleic Acids Res.* 32, W503–W507
- [20] Bernardes JS, Dávila AMR, Costa VS, and Zaverucha G (2007): Improving model construction of profile HMMs for remote homology detection through structural alignment. *BMC Bioinformatics* 8, 435–447
- [21] Madera M and Gough J (2002): A comparison of profile hidden Markov model procedures for remote homology detection. *Nucl. Acid Res.* 30, 4321–4328

4 List of Abbreviations

A	Transition probability matrix
a_{nm}	Transition probability to get from state n to m in the statistical model
B	Emission probability matrix
B_n	Vector of emission probability of all K symbols at the state n
b_{nk}	Probability of observing symbol k and being in the state n ; also named emission probability
D_t	Deletion state at position t in the pHMM
eq.	equation
HMM	Hidden Markov Model
I_t	Insertion state at the position t in the pHMM
I	Length of the state path created by the statistical model
i	Any position in the state path ($i \in \{0, \dots, i, \dots, I\}$)
K	Number of symbols in the set of observation symbols s_o
k	One of the symbols out of the set of observation symbols
L	Length of the observation sequence O
l	Any position in the observation sequence ($l \in \{1, \dots, l, \dots, L\}$)
λ	A statistical model used in the Markov Chain, HMM or pHMM
λ^*	Optimal model of a pHMM
λ^p	Preliminary model of a pHMM
λ_S	Structural model
λ_A	Adjustable parameters of a model
M_t	Match state at the position t in the pHMM
MSA	Multiple sequence alignment

MAP	Maximum <i>a posteriori</i>
m	Any state of the statistical model ($m \in \{1...N\}$)
N	Number of states in the statistical model of a Markov Chain, a HMM or a pHMM
n	Any state of the statistical model ($n \in \{1...N\}$)
O	Sequence of observations
o_l	An observation at the position l in the observation sequence O
PSSM	Position-dependent Scoring Substitution Matrix
pHMM	profile Hidden Markov Model
p	Length of an ungapped block of matches (or insertions) in the MSA
π	Initial charge distribution of the probability matrices A and B
Q	State path through the statistical model ($Q = \{q_0, ..., q_i, ..., q_I\}$)
Q^*	Optimal path through the model
O_x	Any observation sequence of the set of sequences X
q_i	obtained state symbol at the position i in the state path Q
q_0	Start state
q_I	Stop state
R_k	the probability of observing the aminoacid k in a random model
S	Set of observation sequences O ($= \{O_1, ..., O_x, ..., O_X\}$); representing an MSA in case of the pHMM
s_o	Set of observation symbols, which can be observed at each observation o_l
s_q	Set of state symbols, each of them defines a states; In addition q_i can be any of these state symbols
$s(n, k)$	Logg-odd score of the emmision probability of the aminoacid k at the position n in the model

T	Length of a pHMM
t	Any position in the pHMM
$v_n(o_t)$	Viterbi trelli
V	Log-odds of the Viterbi trelli
X	Amount of observation sequences of the set of observation sequences S
x	Any position in the set of observation sequence S ($s \in \{1, \dots, x, \dots, X\}$)
$\gamma(g)$	Gap penalty function

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Ferner erkläre ich, dass ich weder diese noch eine gleichartige Doktorprüfung an einer anderen Hochschule endgültig nicht bestanden habe.

Bayreuth, den 5. November 2008

Eva-Maria Krammer