**JMB**

Available online at www.sciencedirect.com

**ScienceDirect**

**ELSEVIER**

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

# Eva-Maria Krammer[1]†, Mirco S. Till[1]†, Pierre Sebban[2] and G. Matthias Ullmann[1]*

[1]*Structural Biology/ Bioinformatics, University of Bayreuth, Universitätsstrasse 30, BGI, Bayreuth 95447, Germany*

[2]*Laboratoire de Chimie Physique, UMR 8000, Université Paris-Sud XI/CNRS, Faculté des Sciences d'Orsay, Bâtiment 350, 91405 Orsay Cedex, France*

*Edited by D. Case*

In the bacterial reaction center (bRC) of *Rhodobacter sphaeroides*, the key residues of proton transfer to the secondary quinone ($Q_B$) are known. Also, several possible proton entry points and proton-transfer pathways have been proposed. However, the mechanism of the proton transfer to $Q_B$ remains unclear. The proton transfer to $Q_B$ in the bRC of *Blastochloris viridis* is less explored. To analyze whether the bRCs of different species use the same key residues for proton transfer to $Q_B$, we determined the conservation of these residues. We performed a multiple-sequence alignment based on profile hidden Markov models. Residues involved in proton transfer but not located at the protein surface are conserved or are only exchanged to functionally similar amino acids, whereas potential proton entry points are not conserved to the same extent. The analysis of the hydrogen-bond network of the bRC from *R. sphaeroides* and that from *B. viridis* showed that a large network connects $Q_B$ with the cytoplasmic region in both bRCs. For both species, all non-surface key residues are part of the network. However, not all proton entry points proposed for the bRC of *R. sphaeroides* are included in the network in the bRC of *B. viridis*. From our analysis, we could identify possible proton entry points. These proton entry points differ between the two bRCs. Together, the results of the conservation analysis and the hydrogen-bond network analysis make it likely that the proton transfer to $Q_B$ is not mediated by distinct pathways but by a large hydrogen-bond network.

*Keywords:* photosynthetic reaction center; proton transfer; hydrogen-bond network; graph-theoretical analysis; sequence alignment using profile hidden Markov model

## Introduction

A central protein of photosynthesis is the photosynthetic bacterial reaction center (bRC). The L and M subunits form together with the H subunit—and in some bacterial species also a C subunit—the bRC protein. The ultimate step of conversion of excitation energy into chemical energy takes place at the terminal electron acceptor, a quinone molecule bound at the secondary quinone ($Q_B$) binding site of the bRC. In the course of two light-induced electron-transfer reactions, $Q_B$ binds two protons that are taken up from the cytoplasm. The proton uptake is mediated by the protein. These reactions lead to an electrochemical gradient and to the full reduction of the quinone into a dihydroquinone. In the bRC of *Rhodobacter sphaeroides*, the ultimate proton donors to $Q_B$ are AspL213[1] and GluL212[2] for the first proton and the second proton, respectively. The way in which the protons are taken up and how they are transiently kept during the electron-transfer reactions are still a matter of debate.[3–8] Several groups have proposed different proton-transfer pathways with different entry points (see Fig. 1). Examples for such proton-transfer pathways are a single branched

*Corresponding author. E-mail address:
Matthias.Ullmann@uni-bayreuth.de.

† E.-M.K. and M.S.T. contributed equally to this work.

Abbreviations used: bRC, bacterial reaction center; MSA, multiple-sequence alignment; pHMM, profile hidden Markov model.
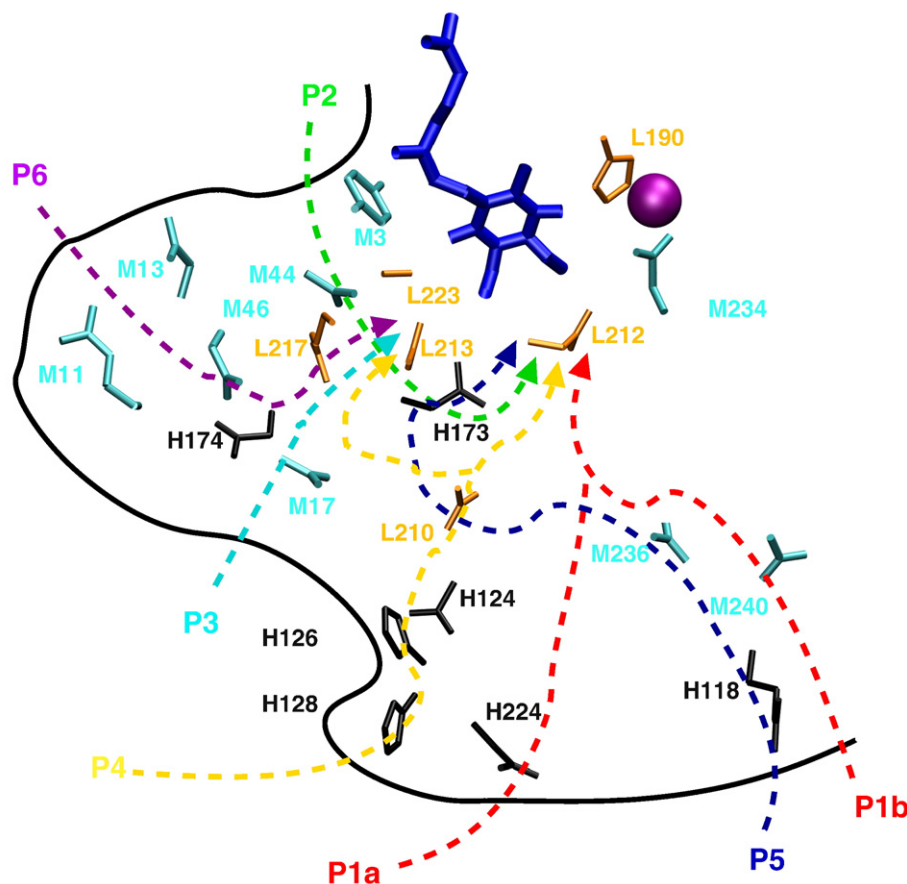
**Fig. 1.** Key residues for proton transfer to $Q_B$. All residues are colored according to their subunit (M = cyan, L = orange and H = black). Only side chains are shown. The proposed proton-transfer pathways P1 (red), P2 (green) and P3 (light blue),[4] P4 (yellow),[3] P5 (dark blue) and P6 (purple)[6] are shown. Additionally, the non-heme iron (purple) and $Q_B$ (blue) are depicted. The figure is based on the crystal structure with PDB code 2I8C and was prepared with VMD.

proton-transfer pathway with the entry point at the $Cd^{2+}$ binding site formed by AspH214, HisH126 and HisH128[3,5,9–13] and a combination of three branched proton-transfer pathways with the entry points TyrM3, AspM17, AspM240 and GluH224.[4] Recently, two extended proton-transfer pathways starting at ArgH118 and ArgM13 were proposed.[6] Inside the protein, several residues are involved in the proton transfer to $Q_B$. These residues are HisL190, AspL210, GluL212, AspL213, ArgL217, SerL223, AsnM44, GluM46, GluM234, GluM236, GluH173 and GlnH174.[3,6,9–12,14–18] There is an agreement in the literature[7,8,19–21] that in the bRC of *R. sphaeroides*, protons are taken up during the first electron transfer to $Q_B$ and are transiently stored in a delocalized hydrogen-bond network of protein residues and water molecules.[22] Not so much information exists about the proton-transfer system and key residues in the bRC of *Blastochloris viridis* since the introduction of mutations in this bacterium is not possible. A $Zn^{2+}/Cu^{2+}$ binding site has been proposed as a possible proton entry point in the bRC of *B. viridis*.[13] This binding site might be located near HisM16 and HisH178.[13] Continuum electrostatic calculations showed that GluL212, GluH177 and GluM234 (numbering refers to *B. viridis*; GluL212,

GluH173 and GluM236 in *R. sphaeroides*) are likely to be involved in proton transfer.[23–25] Moreover, another theoretical study determined a strongly interacting cluster of protonatable residues being coupled to $Q_B$.[26] In this study, possible proton-transfer pathways are also discussed.

In the work presented here, we investigated the organization of proton transfer in the bRC by analyzing the hydrogen-bond network and determining the degree of conservation of key residues using multiple-sequence alignment (MSA). The MSAs are based on profile hidden Markov models (pHMMs) that include structural information of the bRC. The comparison of the hydrogen-bond networks of the bRC from *R. sphaeroides* and that from *B. viridis* gives new insight into the general organization of the proton transfer to $Q_B$. To the best of our knowledge, it is the first time that the hydrogen-bond network involved in proton transfer to $Q_B$ is analyzed using graph theory. Our analysis of the hydrogen-bond network indicates that the proton transfer to $Q_B$ is organized in a large network consisting of several connected clusters and not in distinct pathways. This observation finds an analogy in electron-transfer pathways that are organized in bundles of pathways.[27–31]

## Results and Discussion

The study presented here used MSAs and hydrogen-bond network analysis to examine the conservation and organization of the proton-transfer network from cytoplasm to $Q_B$ in the bRCs of different species. There is a large controversy in the field whether the proton transfer to $Q_B$ occurs along distinct proton-transfer pathways or in a highly delocalized proton-transfer network. Our results on the conservation and structural organization of the network open a new view of this problem.

### Conservation of functional key residues of proton transfer in the bRC

For the bRC of *R. sphaeroides*, several proton pathways with different proton entry points have been proposed (see Fig. 1).[3,4,6,9–12,14–16] But, until today, the exact mechanism of the proton transfer to $Q_B$ is not known. However, from crystallographic, mutational and spectroscopic studies with the bRCs on *R. sphaeroides* and *Rhodobacter capsulatus*, key residues of proton transfer (GlnH173, GluL212, HisL190, AspL210, AspL213, ArgL217, SerL223, AsnM44, GluM46, GluM234 and GluM236) and several possible proton entry points (TyrM3, ArgM13, AspM17, AspM240, ArgH118, AspH124, HisH126, HisH128 and GluH224) have been determined.[3,4,6,9–13] These residues are used as the starting point of our conservation analysis to determine whether these residues are of functional importance for proton transfer. If a key residue is only exchanged to functionally similar amino acids, we assumed that it has a general

**Table 2.** Character of the amino acid at position L210 in dependence on the amino acid pattern at positions L213 and M44 determined from an MSA of 50 bRC sequences

| | [L213, M44] | | L210 | |
|---|---|---|---|---|
| Pattern | Occurrence [%] | | Glu [%] | Asp [%] |
| [Asn, Asp] | 42 (21) | | 100 (21) | 0 (0) |
| [Asp, Asn] | 38 (19) | | 32 (6) | 68 (13) |
| [Asp, Met] | 2 (1) | | 0 (0) | 100 (1) |
| [Asp, Gln] | 18 (9) | | 100 (9) | 0 (0) |

The numbers in parentheses give the absolute number of occurrences of the patterns.

functional role in proton transfer in all analyzed bRCs. The results of this conservation analysis are shown in Table 1. Apart from AspM240, none of the putative proton entry points is totally conserved. Some of them (at positions M13, M17, H124, H126 and H224; numbering refers to *R. sphaeroides*) are mostly changed to other protonatable residues—i.e., they might keep their ability to transfer protons. However, HisH128 and ArgH118 are exchanged to non-polar amino acids in nearly 25% of the analyzed sequences. Thus, in these species, residues H128 and H118 cannot be involved in proton transfer to $Q_B$.

Many of the non-surface residues identified to participate in the proton transfer are highly conserved (at positions H173, L190, L212, L217, L223, M46 and M234). AspL210 is exchanged in 73.2% of the sequences to a glutamate, and GluM236 is exchanged in 15.6% of the sequences to an aspartate. Both glutamate and aspartate are able to participate in proton transfer; thus, L210 and M236 can have the

**Table 1.** Conservation of residues involved in proton transfer

| Subunit | Residue of *R. sphaeroides* | Conservation (%) | Exchanged to (%) | | | |
|---|---|---|---|---|---|---|
| | | | Negative | Positive | Polar | Other |
| L | HisL190 | 100.0 | | | | |
| | AspL210 | 26.8 | E (73.2) | | | |
| | GluL212 | 100.0 | | | | |
| | AspL213 | 60.0 | | | N (40.0) | |
| | ArgL217 | 100.0 | | | | |
| | SerL223 | 100.0 | | | | |
| M | *TyrM3* | 95.7 | | | | F/I (4.3) |
| | GlnM11 | 97.9 | | R (2.1) | | |
| | *ArgM13* | 42.6 | D/E (5.3) | H/K (19.1) | T/S/Q (24.5) | A/G/V (8.5) |
| | *AspM17* | 8.5 | E (50.0) | H (18.1) | Y (21.3) | M/P (2.1) |
| | AsnM44 | 44.0 | D (35.0) | | Q (19.0) | M (2.0) |
| | GlnM46 | 98.2 | E (0.9) | | S (0.9) | |
| | GluM234 | 100.0 | | | | |
| | GluM236 | 83.5 | D (15.6) | | Y (0.9) | |
| | *AspM240* | 100.0 | | | | |
| H | *ArgH118* | 36.4 | D/E (15.1) | H/K (6.1) | Q/N/T (18.2) | P/A (24.2) |
| | *AspH124* | 42.4 | | | N/T (51.5) | G (6.1) |
| | *HisH126* | 39.4 | D/E (51.5) | | | A/G (9.1) |
| | *HisH128* | 45.5 | E (6.1) | K (3.0) | N/Q/T (12.1) | V/A/L/I (33.3) |
| | GluH173 | 97.0 | | | S (3.0) | |
| | GlnH174 | 33.3 | | H (6.1) | N/S/Y (12.1) | V/A/P/M/L/I (48.5) |
| | *GluH224* | 9.1 | | R (3.0) | Q/S/Y (78.8) | V/F (9.1) |

The amino acid exchanges to a negative (D, E), a positive (R, H, K), a polar (T, W, S, N, Q, Y, C) or some other group are listed. Residues that have previously been proposed to function as proton entry points are shown in italics. The numbering refers to *R. sphaeroides*.

same functional role in all species. At position M44, either a polar amino acid or a protonatable amino acid is found in the sequences (see Table 1). At position L213, either an aspartate or an asparagine is found. Our analysis shows that most putative proton entry points are not conserved, and even a high level of sequence variability is observed for some of them. We therefore think that proton entry points might differ from species to species and are not evolutionarily conserved. However, the non-surface key residues show a high degree of conservation, and if an exchange is observed, it is only an exchange to a functionally similar amino acid.

## Correlation of the amino acid character at positions L213 and M44

An interesting phenomenon that could be termed *correlated mutation* has been described for the amino acids at positions M44 and L213 in the bRC.[32,33] In the bRC of *R. sphaeroides*, the combination AsnM44/AspL213 is found, whereas the combination AspM44/AsnL213 is the wild-type pattern of the bRC of *B. viridis*. The double mutant AspL213→Asn/AsnM44→Asp of the bRC of *R. sphaeroides* grows photosynthetically, while the single mutant AspL213→Asn is not able to do so.[32,33] It seems very likely that the combination of a polar amino acid and a protonatable amino acid at positions M44 and L213 is required for proton transfer to $Q_B$.

We assessed the proposed correlation by analyzing an MSA of 50 sequences of the L subunit and the corresponding M subunit. This analysis shows that for residues [L213, M44], the pattern [polar, protonatable] or [protonatable, polar] is always found (see Table 2). In addition to the wild-type patterns of *R. sphaeroides* [Asp, Asn] and *Rhodopseudomonas viridis* [Asn, Asp], the patterns [Asp, Met] and [Asp, Gln], respectively, are present. The pattern [Asp, Met] was found only in the bRC of *Rubrivivax gelatinosus*. There are several sequences available for the M subunit of the bRC of this species in the databases. In all these sequences, a methionine is found at position M44 (numbering refers to *R. sphaeroides*), and wrong sequencing at this position is thus unlikely. By further examination of the alignment, we found an interesting phenomenon that was, to our knowledge, not described before. The character of the amino acid at position L210 is correlated with the pattern of the residues [L213, M44] (see Table 2). In all examined sequences with the pattern [AsnL213, AspM44], L210 is a glutamate. For sequences with the pattern [AspL213, AsnM44], L210 is either a glutamate (32%) or an aspartate (68%). In sequences with the pattern [AspL213, GlnM44], L210 is always a glutamate. At this point, we have no clear explanation for this correlation. Both aspartate and glutamate at position L210 can fulfill the function of L210 in proton transfer; however, they differ in size.

## Description of the hydrogen-bond network

To further investigate the organization of the proton transfer to $Q_B$, we analyzed the hydrogen-bond network that includes $Q_B$ for the bRC proteins of two species, *R. sphaeroides* and *B. viridis*. In the bRCs of both species, we found several unconnected hydrogen-bond networks. Among these networks, a large hydrogen-bond network connects $Q_B$ to the cytoplasm. This network will be called $Q_B$ network in the following. In the bRC from *R. sphaeroides*, it consists of 50 protein residues and 79 water molecules; in the bRC from *B. viridis*, 55 protein residues and 82 water molecules. Another large hydrogen-bond network is found around $Q_A$. However, $Q_A$ is not part of this network or any other hydrogen-bond network. Thus, even if it would be energetically possible, the reduced $Q_A$ cannot be protonated, since a proton cannot be transferred from the cytoplasm to $Q_A$.

We clustered the $Q_B$ network in order to analyze its structural organization. To identify the optimal division of this network, we determined the modularity in dependence of the number of clusters, which was varied between 2 and 50. The number of clusters at which the modularity is maximal represents the optimal clustering of the network. A modularity above 0.7 indicates that a network is highly structured—i.e., it can be well divided into several clusters. As can be seen in Fig. 2, the optimal clustering with a modularity of 0.77 for *R. sphaeroides* and that of 0.75 for *B. viridis* is obtained with 11 clusters for the $Q_B$ network. The locations of the different clusters in the bRC structures are depicted in Fig. 3. Figure 4 shows schematically the clusters and their connections. Some but not all residues that have been discussed before to be part of proton-transfer pathways are connections between clusters. From visual examination of the clusters in Fig. 3, it can be seen that the network and clusters are similar for both species and differ only in details. Several residues close to the cytoplasmic surface of the protein could function as proton entry points. These
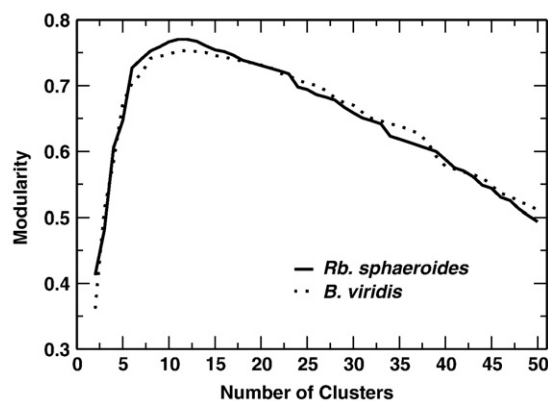


**Fig. 2.** Modularity of the clustering in dependence on the number of clusters for the $Q_B$ network of the bRCs from *B. viridis* (dotted line) and from *R. sphaeroides* (continuous line).

residues are listed in Table 3. Many of these possible proton entry points are not conserved, as shown in Table 4. However, some of these residues show a high degree of functional conservation. Interestingly, in both species, the cluster containing $Q_B$ includes no proton entry point. Thus, proton-transfer connections in the protein interior and to clusters with proton entry points are needed for the protonation of $Q_B$. The connections of the $Q_B$ cluster play a critical role for the proton transfer from cytoplasm to $Q_B$. The existence of at least one of these connections is essential, because otherwise the proton cannot reach $Q_B$.

Based on our analysis, a large hydrogen-bond network connecting $Q_B$ to the cytoplasm exists in both species. This network can be divided into several clusters. It thus seems likely that the proton transfer occurs not along certain residues but along certain clusters.

## Key residues included in the hydrogen-bond network

As shown in Table 5, the known non-surface residues involved in proton transfer are all part of the hydrogen-bond network. For the MSA, we found that the character of the amino acid at positions L210, L213, M44 and H174 in the bRC of *R. sphaeroides* differs from that in the bRC of *B. viridis* (see Table 5).

Compared with the non-surface residues involved in proton transfer, the situation for the proton entry points proposed in earlier studies is different.[3,4,6,9–13] First, not all of them are part of the calculated hydrogen-bond network in both investigated bRCs. Second, based on our calculations, not all of them are directly connected to the cytoplasm. In the bRC of *R. sphaeroides*, the proposed proton entry points TyrM3, ArgM13, AspH124 and HisH126 are part of the $Q_B$ network and are connected to the cytoplasm (see Table 3). In the bRC of *B. viridis*, only TyrM3 and ArgM13 are part of the $Q_B$ network (see Table 5). Both residues could act as proton entry points. All other proposed proton entry points (M17, H118, H124, H126, H128 and H224; numbering refers to *R. sphaeroides*) are not part of the $Q_B$ network of the bRC of *B. viridis*. In the bRC of *R. sphaeroides*, the $Cd^{2+}$ binding site formed by H124, H126 and H128 was proposed to function as a proton entry point.[3,9,13] Also, in our calculations, AspH124 and HisH126 are

## (a) *Rb. sphaeroides*
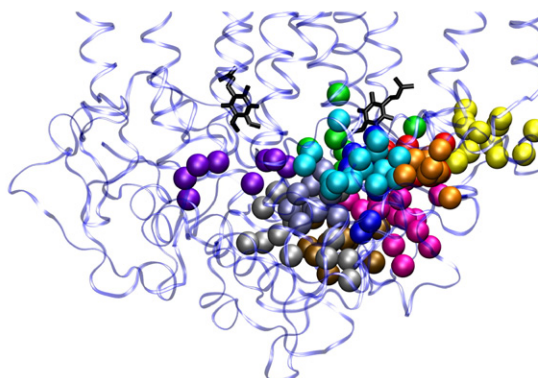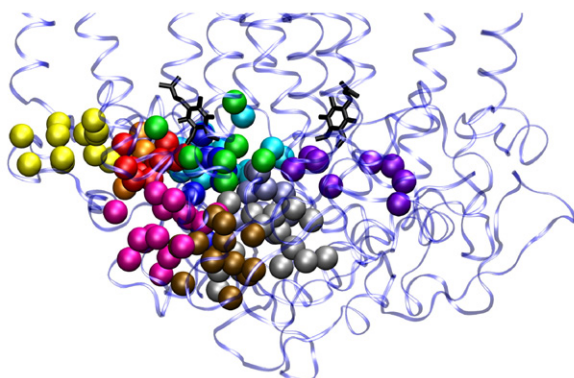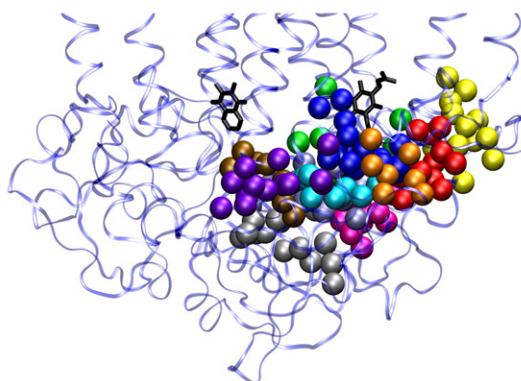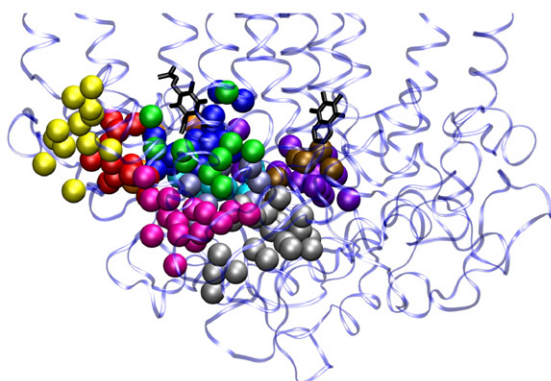


## (b) *B. viridis*



**Fig. 3.** Clusters of the $Q_B$ network. The colors of the participating groups refer to the clusters (1 = green, 2 = magenta, 3 = red, 4 = yellow, 5 = blue, 6 = cyan, 7 = orange, 8 = violet, 9 = ice blue, 10 = gray and 11 = ocher). The clusters are shown for the bRCs of (a) *R. sphaeroides* and (b) *B. viridis*. For each protein residue or water molecule participating in a cluster, a sphere is shown at the center of mass of the corresponding group. In the left panel, $Q_B$ is situated on the left; in the right panel, $Q_B$ is situated on the right. The figures are based on the crystal structures 2J8C[6] and 2I5N[34] and were prepared with VMD.[35]
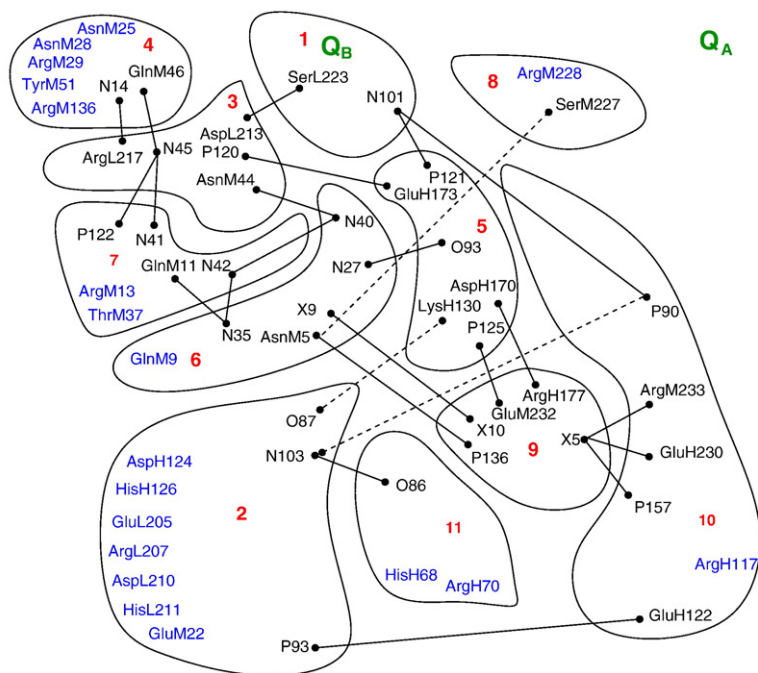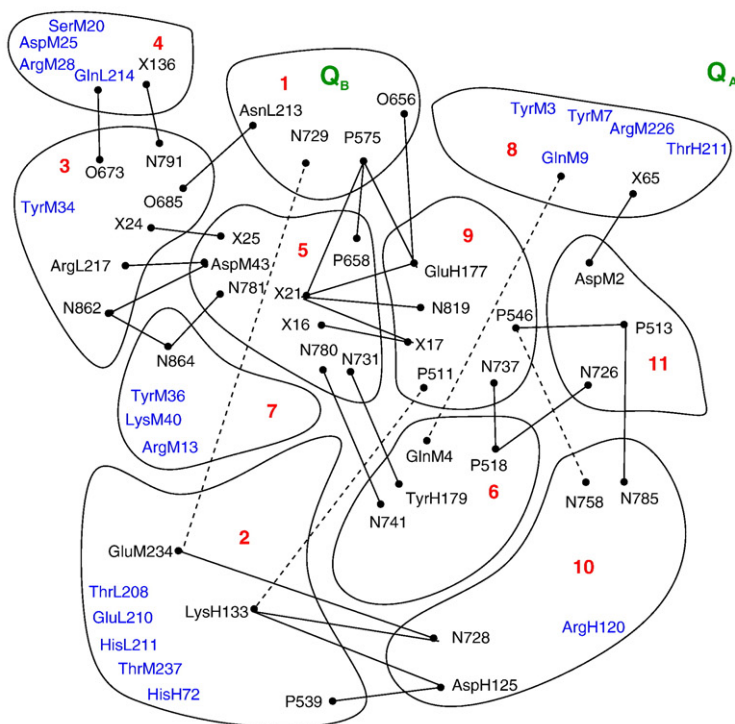
## (a) *Rb. sphaeroides*



## (b) *B. viridis*



**Fig. 4.** Clusters of the $Q_B$ network for the bRCS of (a) *R. sphaeroides* and (b) *B. viridis*. Cluster numbers are shown in red. The connections between the clusters and possible proton entry points (blue) are shown. Connections are shown as continuous lines or as dashed lines if a connection crosses other clusters in this representation. The water molecules with chains M, L and H in the PDB file are named N, O and P, respectively, or X if they were added in this study.

possible proton entry points. A metal binding site is also found in the bRC of *B. viridis*,[13] but it is not located at the same position as in the bRC of *R. sphaeroides*. It was proposed that this binding site may be formed by HisM16 and HisH176. Based on our calculations, HisM16 is not part of the $Q_B$ network. HisH178 is part of the network, albeit not in direct contact with the cytoplasm. Interestingly, in both networks, AspL210 is close to the cytoplasm and could function as a proton entry point.

Based on the calculated hydrogen-bond networks, it is likely that the proton entry points differ in different species but that the non-surface key residues involved in proton transfer are in similar positions in the graph representing the $Q_B$ network of all bRCs.

## Conclusions

The proton transfer to the $Q_B$ of the bRC was examined by a combined analysis of amino acid conservation and the hydrogen-bond network. In all used bRC sequences, the known non-surface key residues of proton transfer are conserved or exchanged to functionally equivalent amino acids. In contrast, most of the previously proposed proton entry points are not conserved, and some of them even show a high level of sequence variability. Thus, it is very likely that the proton transfer to $Q_B$ is mediated by the same functional key residues in all bacterial species but that the proton entry points differ from species to species. The hydrogen-bond networks of the examined bRC proteins from *R. sphaeroides* and *B. viridis* do not show distinct hydrogen-bond pathways from the cytoplasm to $Q_B$. In contrast, a large hydrogen-bond network spanning from the cytoplasm to $Q_B$ was found in both bRC proteins. These networks include all experimentally determined key residues involved in proton transfer. Possible proton entry points were determined in both bRCs. The proton entry points in these two networks are not identical. The analysis of hydrogen-bond network supports further the idea that the proton transfer to $Q_B$ is organized as a proton sponge—i.e., having several proton entry points and transferring the protons in a delocalized network from the entry points via certain key residues to $Q_B$. However, this sponge seems to be structured in several clusters. It thus seems likely that the proton transfer occurs not along certain residues but along certain clusters. The biological significance of such clusters could be that they are more robust against mutations than defined proton-transfer pathways. Nevertheless, the clusters provide an approximately defined route for the proton.

## Materials and Methods

### Multiple-sequence alignment

MSAs are made using pHMMs.[36–41] Respectively, 100, 114 and 33 sequences for subunits L, M and H were used for the MSA. These sequences were taken from a BLAST[42] search on the National Center for Biotechnology Information Web page[43] using the sequences of subunits L, M and H of the bRC from *R. sphaeroides* as query sequences. All sequences found in the database were considered. Redundant sequences were removed from the data set. The construction of the pHMMs used for the MSA of the L and M subunits has been described in an earlier publication.[44] For the construction of the pHMM of the H subunit, we followed a similar strategy. We generated an MSA of the bRCs from *R. sphaeroides* [Protein Data Bank (PDB) code 2J8C][6], *Thermochromatium tepidum* (PDB code 1EYS)[45] and *B. viridis* (PDB code 1PCR)[46] with the program Staccato,[47] which uses structure and sequence information. Default settings were used for this alignment. In order to validate the correctness of the sequence alignments, we identified regions (marker regions) with conserved structure and sequence. Structurally conserved regions of the H subunit were identified by visual inspection of the known bRC structures. We found four maker regions: a β-sheet from H58 to H75, a β-sheet from H148 to H180, an α-helix from H226 to H249 and a loop from H37 to H42 (numbering refers to *R. sphaeroides*). To the MSA with structural information, 12 additional H subunit sequences were aligned. We constructed the pHMM for the H subunit from this alignment. To validate the pHMM, we aligned 33 sequences and calculated the degree of conservation (sequence logos)[48] for the marker regions. The marker regions, the resulting structural alignment and the sequence logos are depicted in Fig. 5. The agreement of the conservation pattern of the marker regions between the MSA with structural information and the MSA obtained with the pHMM made us confident that the pHMM of the H subunit is correct. For the H subunit, the complete lists of the sequences used for building, validating and analyzing the pHMM, the obtained alignment and the pHMM file are given in Supporting Information. For the L and M subunits, these data have been provided in a previous publication.[44]

### Structure preparation

The network calculations are based on high-resolution crystal structures of the bRCs of *R. sphaeroides* (PDB code 2J8C)[6] and *B. viridis* (PDB code 2I5N).[34] For the bRC of

**Table 3.** Possible proton entry points

| Subunit | Protein residues | | | |
| | *R. sphaeroides* | | *B. viridis* | |
| | Residue | Cluster | Residue | Cluster |
|---|---|---|---|---|
| L | GluL205 | 2 | LysL205 | 0 |
| | ArgL207 | 2 | LysL207 | 0 |
| | ThrL208 | (2) | ThrL208 | 2 |
| | AspL210* | 2 | GluL210 | 2 |
| | HisL211* | 2 | HisL211 | 2 |
| | ThrL214 | (3) | GlnL214* | 4 |
| M | TyrM3 | 6 | TyrM3 | 8 |
| | PheM7 | 0 | TyrM7 | 8 |
| | GlnM9 | 6 | GlnM9 | 8 |
| | ArgM13 | 7 | ArgM13 | 7 |
| | GluM22 | 2 | SerM20 | 4 |
| | AsnM25 | 4 | AspM25 | 4 |
| | AsnM28 | 4 | ArgM28 | 4 |
| | ArgM29 | 4 | ValM30 | 0 |
| | PheM35 | 0 | TyrM34 | 3 |
| | ThrM37 | 7 | TyrM36 | 7 |
| | TrpM41 | 0 | LysM40 | 7 |
| | TyrM51 | 4 | TyrM50 | 0 |
| | ArgM136 | 4 | ArgM134 | (4) |
| | ArgM228* | 8 | ArgM226 | 8 |
| | AlaM239 | 0 | ThrM237 | 2 |
| H | HisH68 | 11 | HisH72 | 2 |
| | LysH70 | 11 | - | 0 |
| | ArgH117 | 10 | ArgH120 | 10 |
| | AspH124 | 2 | ThrH127 | 0 |
| | HisH126 | 2 | AspH129 | 0 |
| | AsnH206 | 0 | ThrH211 | 8 |

Residues that are not in direct contact with the cytoplasm but are connected through a water molecule are marked by an asterisk. For comparison, the corresponding residues in the bRCs of *R. sphaeroides* and *B. viridis* are shown. The numbering refers to the corresponding species. The table indicates to which cluster a residue belongs. If the residue is not part of the $Q_B$ hydrogen-bond network, we assigned the cluster number 0. If the residue is not in contact with the cytoplasm, we listed the cluster number in parentheses.

**Table 4.** Conservation of the possible proton entry points proposed based on our network analysis

| Subunit | Residue of *R. sphaeroides* | Conservation (%) | Exchanged to (%) | | | |
|---|---|---|---|---|---|---|
| | | | Negative | Positive | Polar | Other |
| L | GluL205 | 28.9 | D (10.3) | K/R (8.2) | N/S/T (14.4) | A/I/L/P/V (38.1) |
| | ArgL207 | 6.2 | | K (87.6) | C (1.0) | G/M (5.2) |
| | ThrL208 | 57.7 | | H (10.3) | S/Y (22.7) | A/F (9.3) |
| | AspL210* | 26.8 | E (73.2) | | | |
| | HisL211* | 76.3 | | | N/T/Y (22.7) | A (1.0) |
| | ThrL214 | 86.3 | | | Q/S (4.2) | A/I/M (9.5) |
| M | *TyrM3* | 95.6 | | | | F/I (4.4) |
| | PheM7[a] | 90.5 | | | Y (5.3) | L (4.2) |
| | GlnM9 | 71.6 | | R (15.8) | T/S (7.3) | A/G/L/P (5.3) |
| | *ArgM13* | 42.6 | D/E (5.3) | H/K (19.1) | T/S/Q (24.5) | A/G/V (8.5) |
| | GluM22 | 6.1 | D (8.2) | H (7.1) | N/S/T/Y (15.3) | A/G/I/L/M/P/V (63.3) |
| | AsnM25 | 17.5 | D/E (21.6) | H (1.0) | Q/S/T (24.8) | A/G/I/L/M/V (35.1) |
| | AsnM28 | 5.1 | D/E (28.3) | K/R (63.6) | | A/G (3.0) |
| | ArgM29 | 36.4 | E (6.1) | K (1.0) | Q/S/T/Y (17.1) | I/L/M/V/F (39.4) |
| | PheM35[a] | 53.5 | D (1.0) | H (15.2) | N/Q/S/Y (26.3) | L (4.0) |
| | ThrM37 | 15.3 | | H/K/R (16.4) | N/Q/S/W/Y (65.3) | P (2.0) |
| | TrpM41[a] | 20.0 | | K/R (57.0) | Q/Y (6.0) | I/L/V (17.0) |
| | TyrM51 | 84.8 | | H (4.5) | N/W (3.6) | L/P/F (7.1) |
| | ArgM136 | 72.8 | | | | I/C/L (27.2) |
| | ArgM228 | 98.2 | | H (0.9) | | L (0.9) |
| | AlaM239* | 2.8 | | | T/Y (57.6) | I/L/M/V/F (39.6) |
| H | HisH68 | 45.5 | D (24.2) | | N/S/T (9.0) | G (21.2) |
| | LysH70 | 9.1 | | H/R (57.6) | N/Q (12.1) | A/G (18.2) |
| | ArgH117 | 97.0 | | K (3.0) | | |
| | *HisH126* | 39.4 | D/E (51.5) | | | A/G (9.1) |
| | *HisH128* | 45.5 | E (6.1) | K (3.0) | N/Q/T (12.1) | V/A/L/I (33.3) |
| | AsnH206 | 15.2 | D/E (30.3) | K/R (30.3) | Q/T (18.2) | A/G (6.0) |

The conservation analysis is based on our MSAs. Residues that have previously been proposed to function as proton entry points are shown in italics. Residues that are not in direct contact with the cytoplasm but are connected through a water molecule are marked by an asterisk. The amino acid exchanges to a negative (D, E), a positive (R, H, K), a polar (T, W, S, N, Q, Y, C) or some other residue are listed. For residues ThrM37 and LysH70, the complete percentage does not lead to 100% since gaps (one for ThrM37 and three for LysH70) were found at these positions in the MSA. The numbering refers to *R. sphaeroides*.
[a] This residue is a proton entry point in the bRC of *B. viridis*.

**Table 5.** Previously determined key residues of proton transfer and their participation in the $Q_B$ network in the bRCs of *R. sphaeroides* and of *B. viridis*

| | Protein residues | | | |
|---|---|---|---|---|
| | *R. sphaeroides* | | *B. viridis* | |
| Location | Residue | Cluster | Residue | Cluster |
| Non-surface residues | HisL190 | 1 | HisL190 | 1 |
| | AspL210 | 2 | GluL210 | 2 |
| | GluL212 | 1 | GluL212 | 1 |
| | AspL213 | 3 | AsnL213 | 1 |
| | ArgL217 | 3 | ArgL217 | 3 |
| | AspL218 | 4 | AspL218 | 4 |
| | SerL223 | 1 | SerL223 | 1 |
| | AsnM44 | 3 | AspM43 | 5 |
| | GlnM46 | 4 | GlnM45 | 3 |
| | GluM236 | 11 | GluM234 | 2 |
| | GluH173 | 5 | GluH177 | 9 |
| | GlnH174 | 0 | HisH178 | 3 |
| Proposed proton entry points | TyrM3 | 6 | TyrM3 | 7 |
| | ArgM13 | 7 | ArgM13 | 9 |
| | AspM17 | 0 | HisM16 | 0 |
| | AspM240 | 0 | AspM138 | 0 |
| | ArgH118 | 0 | AlaH121 | 0 |
| | AspH124 | 2 | ThrH127 | 0 |
| | HisH126 | 2 | AspH129 | 0 |
| | HisH128 | 0 | LysH131 | 0 |
| | GluH224 | 0 | GlnH229 | 0 |

The numbering refers to the corresponding species. The table indicates to which cluster a residue belongs. If the residue is not part of the $Q_B$ network, we assigned the cluster number 0.

*R. sphaeroides*, only the proximal position of $Q_B$ is used for the calculations since the distal position is thought to be unproductive.[49–52] For both structures, the lipids of the crystal structures were included in the calculations. Hydrogen atoms are placed with the HBUILD module[53] of CHARMM,[54] followed by energy optimization of the hydrogen positions, while the heavy-atom positions are kept fixed. In the used crystal structure of the bRC from *B. viridis*, no coordinate is given for the loop region from H46 to H53. Since this loop is located in the cytoplasmic part of the protein and could thus be important for proton transfer, the loop is modeled into the structure. Starting coordinates for this loop are taken from a lower-resolution crystal structure (PDB code 1PRC).[46] The atom coordinates of the loop residues are minimized, while the rest of the protein is kept fixed. To define the membrane-spanning part of the proteins, we superimposed the used structures with the crystal structure that was obtained by the lipidic cubic phase method (PDB code 1OGV).[55] For this structure, the region of the lipid bilayer can be easily deduced.[55] The hydrophobic region of the membrane spans from −6.55 to 27.45 Å on the *z*-axis. Since not necessarily all water positions are resolved in crystal structures, we searched for internal cavities using the program McVol. The algorithm evaluates whether points, which are randomly placed in a box containing the protein, are inside the protein or inside the solvent. Clusters of solvent points inside the protein, which have no connection to other solvent points, are identified as cavities. Additional water molecules were placed in these cavities if their volume was more than 18 Å³.
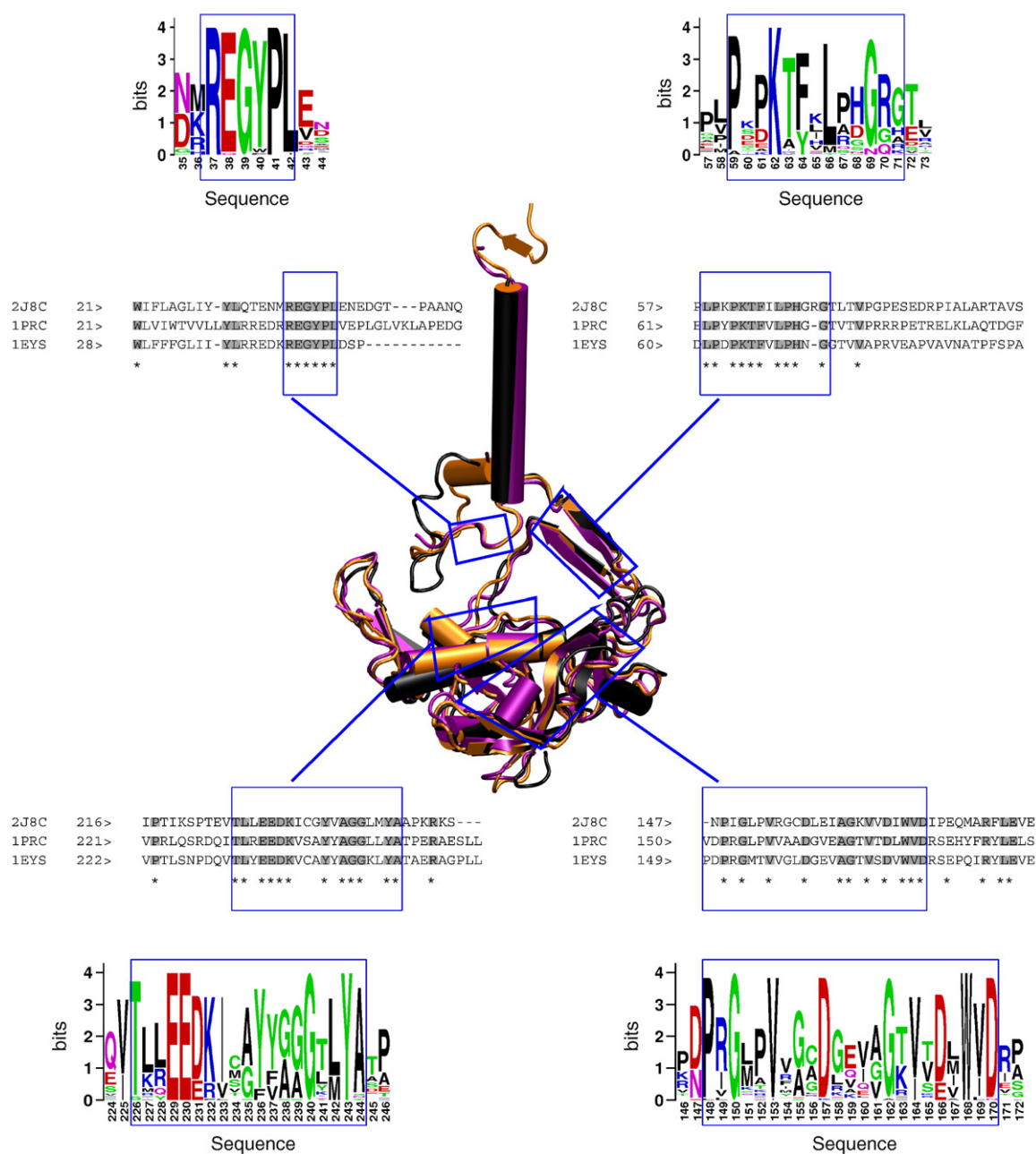
```
2J8C   21>  WIFLAGLIY-YLQTENMREGYPLENEDGT---PAANQ        2J8C   57>  HLPKPKTFILPHGRGTLTVPGPESEDRPIALARTAVS
1PRC   21>  WLVIWTVVLLYLRREDRREGYPIVEPLGLVKLAPEDG        1PRC   61>  HLPYPKTFVLPHG-GTVTVPRRRPETRELKLAQTDGF
1EYS   28>  WLFFFGLII-YLRREDKREGYPLDSP----------        1EYS   60>  DLPDPKTFVLPHN-GGTVVAPRVEAPVAVNATPFSPA
            *        **  ******                               ** ****  ***  *          *
```

```
2J8C  216>  IPTIKSPTEVTLLEEDKICGYVAGGLMYAAPKRKS---       2J8C  147>  -NPIGLPVRGCDLEIAGKVVDIWVDIPEQMARFLEVE
1PRC  221>  VPRLQSRDQITLRREDKVSAMYAGGLLYATPERAESLL       1PRC  150>  VDPRGLPVVAADGVEAGTVTDLWVDRSEHYFRYLELS
1EYS  222>  VPTLSNPDQVTLYEEDKVCAMYAGGKLYATAERAGPLL       1EYS  149>  PDPRGMTVVGLDGEVAGTVSDVWVDRSEPQIRYLEVE
            *         ** ****  * *** **  *                      * *   *    ** * *** *   * ***
```

**Fig. 5.** Superposition of the H subunits of *R. sphaeroides* (PDB code 2J8C; black), *T. tepidum* (PDB code 1EYS; purple) and *R. viridis* (1PRC; orange).[6,45,46] Regions with high conservation are marked in the superposition, in the structural alignment and in the sequence logo. In the sequence logo, the maximum conservation at a certain position is given by $\log_2 20 = 4.32$ bits, since 20 amino acids are, in principle, possible.[48] These regions were used for validation of the pHMM. The corresponding sequence logo of the resulting profile alignment is given next to the structural alignment. Sequence logos were done using the WebLogo program.[48]

Since water clusters on the protein surface give no information about the proton transfer inside the protein, all water molecules on the protein surface were removed. All water molecules with a distance of less than 3.0 Å from the solvent-accessible surface of the protein were removed. The solvent-accessible surface was calculated by the program McVol, setting the probe sphere radius to 1.4 Å. The calculation of the solvent-accessible surface and the removal of the surface water molecules were done iteratively until no more water molecules were found at the protein surface. However, water molecules located in protein pockets (clefts) are potentially important as

hydrogen-bond partners. If such a water molecule is near the protein surface, it was removed by our algorithm. Thus, we placed water molecules in the clefts with a similar algorithm as described above for the placement of water molecules in cavities.

**Building of the hydrogen-bond network**

We describe the hydrogen-bond network in the proteins as a graph. Graph theory has been used in previous studies to investigate electron-transfer pathways in

proteins.[26,28–30,56–58] In mathematics, a graph is a representation of a set of objects where some pairs of the objects are connected by links. The objects are called nodes, and the links are called connections. In our study, water molecules, protein residues with polar side chains (arginine, aspartate, lysine, glutamate, histidine, threonine, serine, tyrosine, tryptophan, the N-terminus and the C-terminus) and cofactors (quinone, cardiolipin, bacteriopheophytin and bacteriochlorophyll) are considered as nodes in the graph representation of the hydrogen-bond network. Possible hydrogen bonds are considered as connections between these nodes. Two distance criteria are used to identify a hydrogen bond between two possible hydrogen-bond partners. The distance between donor and acceptor heavy atoms should be less than 4.0 Å, and the distance between the acceptor heavy atom and the hydrogen should be less than the distance between the donor heavy atom and the acceptor heavy atom. Assuming that the distance between the donor heavy atom and the hydrogen varies between 0.9 and 1.0 Å and the distance between the donor heavy atom and the acceptor heavy atom varies between 2.0 and 4.0 Å, the angle between hydrogen, donor heavy atom and acceptor heavy atom is always less than 85°. Proton entry points are residues that are in contact with the cytoplasm—i.e., the proton donor or acceptor of this residue is less than 3.0 Å apart from the solvent-accessible surface of the protein. During our analysis, we realized that the distance between the carboxylate oxygen of GluL212 and $Q_B$ is about 4.5 Å; therefore, this hydrogen bond was not included in our network. We inspected the structure and electron density near GluL212 using the Coot[59] program. The electron density is not well defined at this position. We assumed that GluL212 is connected to the O2 oxygen of $Q_B$ either directly or by a water molecule in our calculations and introduced a hydrogen bond between these atoms.

### Network analysis and clustering

The hydrogen-bond network is clustered by the algorithm of Girvan and Newman (betweenness clustering algorithm)[60] (see Fig. 6). The algorithm is a divisive clustering algorithm and clusters the hydrogen-bond network based on its topological properties. The algorithm iteratively removes connections from the network, dividing the graph into more and more subgraphs. The decision which connection is deleted at each iteration step is based on an all-pairs-shortest-path search. The betweenness of a certain connection is defined as the number of shortest paths containing this connection. The connection with the highest betweenness is removed. Afterward, the number of unconnected subgraphs of the remaining network is evaluated. These three steps (i.e., calculating the betweenness, removing the connection with the highest betweenness and evaluating the remaining network) are done iteratively until the desired number of subgraphs is reached. Each of these subgraphs is then considered as a cluster. To evaluate the quality of clustering, we calculated the modularity[61] in dependence on the number of clusters. The modularity $Q$ of the clustering is given by the following equation:

$$Q = \sum_{i=1}^{K} \left( \frac{A_{ii}}{N} - \left( \sum_{j=1}^{K} \frac{A_{ij}}{N} \right)^2 \right) \tag{1}$$

where $K$ is the number of clusters, $N$ is the total number of connections in the network, $A_{ii}$ is the number of connec-
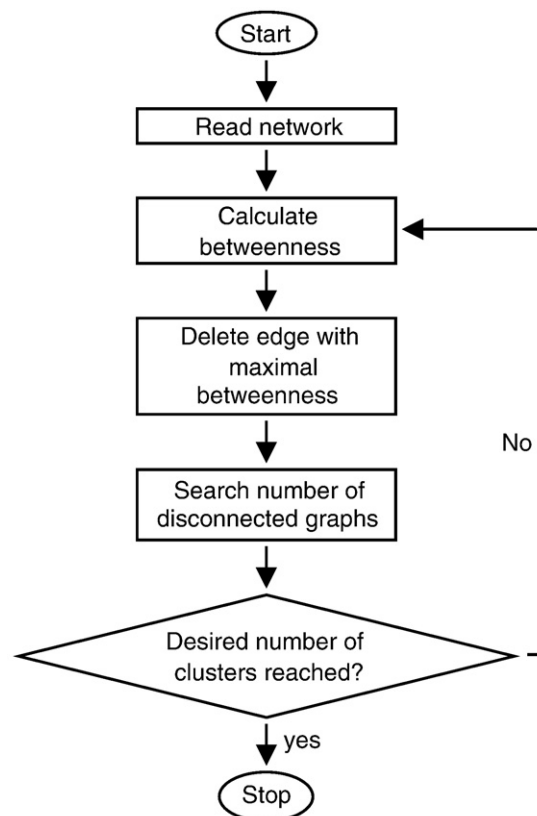


**Fig. 6.** Flowchart of the betweenness clustering algorithm. The algorithm iteratively removes the edge with the highest betweenness. The iteration runs until the network is divided into the desired number of subgraphs.

tions within cluster $i$ and $A_{ij}$ is the number of connections between cluster $i$ and cluster $j$. The second term in Eq. (1) requires some additional explanation. Let us consider two clusters, $i$ and $j$, one with $k_i$ connections and the other with $k_j$ connections. The average number of connections, $L_{ij}$, between these clusters is given by

$$L_{ij} = \frac{k_i k_j}{N} \tag{2}$$

if the connections are placed randomly. An equivalent equation can be used for the expected number of connections, $L_{ii}$, within a single cluster $i$. Since Eq. (3) is valid,

$$\frac{L_{ii}}{N} = \frac{k_i^2}{N^2} = \frac{\left( \sum_{j=1}^{K} A_{ij} \right)^2}{N^2} = \left( \sum_{j=1}^{K} \frac{A_{ij}}{N} \right)^2 \tag{3}$$

the last term in Eq. (1) represents the average number of connections of cluster $i$. The modularity can take values between one and zero and is related to the difference between the number of connections within each of the clusters and the average number of connections of each cluster. A randomly clustered network would give a modularity close to zero. Large values of the modularity indicate a high quality of the clustering. Newman and Girvan reported that modularities of 0.7 and higher indicate a strong clustering.[61]

## Acknowledgements

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.03.020

## References

1. Takahashi, E. & Wraight, C. A. (1990). A crucial role for Asp L213 in the proton transfer pathway to the secondary quinone of reaction centers from *Rhodobacter sphaeroides*. *Biochim. Biophys. Acta*, **1020**, 107–111.
2. Paddock, M. L., Rongley, S. H., Feher, G. & Okamura, M. Y. (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.
3. Ädelroth, P., Paddock, M. L., Tehrani, A., Beatty, J. T., Feher, G. & Okamura, M. Y. (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.
4. Abresch, E. C., Paddock, M. L., Stowell, M. H. B., McPhillips, T. M., Axelrod, H. L., Soltis, S. M. *et al.* (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.
5. Paddock, M. L., Feher, G. & Okamura, M. Y. (2003). Proton transfer pathways and mechanism in bacterial reaction centers. *FEBS Lett.* **555**, 45–50.
6. Koepke, J., Krammer, E. M., Klingen, A. R., Sebban, P., Ullmann, G. M. & Fritzsch, 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.
7. Miksovska, J., Schiffer, M., Hanson, D. K. & 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**, 13453–14348.
8. Cheap, H., Tandori, J., Derrien, V., Benoit, M., deOliveira, P., Koepke, J. *et al.* (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.
9. Ädelroth, P., Paddock, M. L., Sagle, L. B., Feher, G. & Okamura, M. Y. (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.
10. Xu, Q., Axelrod, H. L., Abresch, E. C., Paddock, M. L., Okamura, M. Y. & 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.
11. Paddock, M. L., Feher, G. & Okamura, M. Y. (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.
12. Paddock, M. L., Graige, M. S., Feher, G. & Okamura, M. Y. (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.
13. Utschig, L. M., Poluektov, O., Schlesselman, S. L., Thurnauer, M. C. & Tiede, D. M. (2001). $Cu^{2+}$ site in photosynthetic bacterial reaction centres from *Rhodobacter sphaeroides*, *Rhodobacter capsulatus* and *Rhodopseudomonas viridis*. *Biochemistry*, **40**, 6132–6141.
14. Miksovska, J., Kálmán, L., Schiffer, M., Maróti, P., Sebban, P. & Hanson, D. K. (1997). In bacterial reaction centers rapid delivery of the second proton to $Q_B$ can be achieved in the absence of L212Glu. *Biochemistry*, **36**, 12216–12226.
15. Sebban, P., Maróti, P., Schiffer, M. & Hanson, D. K. (1995). Electrostatic dominoes: long distance propagation of mutational effects in photosynthetic reaction centers of *Rhodobacter capsulatus*. *Biochemistry*, **34**, 8390–8397.
16. Hanson, D. K., Baciou, L., Tiede, D. M., Nace, S. L., Schiffer, M. & Sebban, P. (1992). In bacterial reaction centers protons can diffuse to the secondary quinone by alternative pathways. *Biochim. Biophys. Acta*, **1102**, 260–265.
17. Rabenstein, B., Ullmann, G. M. & Knapp, E. W. (2000). Electron transfer between the quinones in the photosynthetic reaction center and its coupling to conformational changes. *Biochemistry*, **39**, 10496–10587.
18. Taly, A., Sebban, P., Smith, J. C. & Ullmann, G. M. (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.
19. McPherson, P. H., Okamura, M. Y. & Feher, G. (1988). Light-induced proton uptake by photosynthetic reaction centers from *Rhodobacter sphaeroides* R-26. I. Protonation of the one-electron states $D^+Q_A^-$, $D^+Q_AQ_B^-$ and $DQ_AQ_B^-$. *Biochim. Biophys. Acta*, **934**, 348–368.
20. Maróti, P. & Wraight, C. A. (1988). Flash-induced $H^+$ binding by bacterial photosynthetic reaction centers: influences of the redox states of the acceptor quinones and primary donor. *Biochim. Biophys. Acta*, **934**, 329–347.
21. Maróti, P., Hanson, D. K., Schiffer, M. & Sebban, P. (1995). Long-range electrostatic interaction in the bacterial photosynthetic reaction centre. *Nat. Struct. Biol.* **2**, 1057–1059.
22. Tandori, J., Baciou, L., Alexov, E., Maroti, P., Schiffer, M., Hanson, D. K. & Sebban, P. (2001). Revealing the involvement of extended hydrogen bond networks in the cooperative function between distant sites in bacterial reaction centers. *J. Biol. Chem.* **276**, 45513–45515.
23. Rabenstein, B. & Ullmann, G. M. (1998). Calculation of protonation patterns in proteins with structural

relaxation and molecular ensembles—application to the photosynthetic reaction center. *Eur. Biophys. J.* **27**, 626–637.

24. Rabenstein, B., Ullmann, G. M. & Knapp, E. W. (1998). Energetics of electron-transfer and protonation reactions of the quinones in the photosynthetic reaction center of *Rhodopseudomonas viridis*. *Biochemistry*, **37**, 2488–2495.

25. Alexov, E. G. & Gunner, M. R. (1999). Calculated protein and proton motions coupled to electron transfer: electron transfer from $Q_A^-$ to $Q_B$ in bacterial photosynthetic reaction centers. *Biochemistry*, **38**, 8253–8270.

26. Lancaster, C. R. D., Michel, H., Honig, B. & Gunner, M. R. (1996). Calculated coupling of electron and proton transfer in the photosynthetic reaction center of *Rhodopseudomonas viridis*. *Biophys. J.* **70**, 2469–2492.

27. Regan, J. J., Risser, S. M., Beratan, D. N. & Onuchic, J. N. (1993). Protein electron transport: single *versus* multiple pathways. *J. Phys. Chem.* **97**, 13083–13088.

28. Farid, R. S., Moser, C. C. & Dutton, P. L. (1993). Electron transfer in proteins. *Curr. Opin. Struct. Biol.* **3**, 225–233.

29. Ullmann, G. M. & Kostić, N. M. (1995). Electron-tunneling paths in various electrostatic complexes between cytochrome *c* and plastocyanin. Anisotropy of the copper–ligand interactions and dependence of the iron–copper electronic coupling on the metalloprotein orientation. *J. Am. Chem. Soc.* **117**, 4766–4774.

30. Beratan, D. N. & Skourtis, S. S. (1998). Electron transfer mechanisms. *Curr. Opin. Chem. Biol.* **2**, 235–243.

31. Jones, M., Kurnikov, I. V. & Beratan, D. N. (2002). The nature of tunneling pathway and average packing density models for protein-mediated electron transfer. *J. Phys. Chem. A*, **106**, 2002–2006.

32. Paddock, M. L., Senft, M. E., Graige, M. S., Rongey, S. H., Turanchik, T., Feher, G. & Okamura, M. Y. (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.

33. Rongley, S. H., Paddock, M. L., Feher, G. & Okamura, M. Y. (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.

34. Li, L., Mustafi, D., Fu, Q., Tereshko, V., Chen, D. L., Tice, J. D. & Ismagilov, R. F. (2006). Nanoliter microfluidic hybrid method for simultaneous screening and optimization validated with crystallization of membrane proteins. *Proc. Natl Acad. Soc. USA*, **103**, 19243–19248.

35. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graphics*, **14**, 33–38.

36. Eddy, S. R. (1996). Hidden Markov models. *Curr. Opin. Struct. Biol.* **6**, 361–365.

37. Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.

38. Birney, E. (2001). Hidden Markov models in biological sequence analysis. *IBM J. Res. Dev.* **34**, 449–454.

39. Mount, D. W. (2001). *Bioinformatics. Sequence and Genome Analysis.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

40. Durbin, R., Eddy, S., Krogh, A. & Mitchison, G. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.* Cambridge University Press, Cambridge, UK.

41. Bernardes, J. S., Dävila, A. M. R., Costa, V. S. & Zaverucha, G. (2007). Improving model construction of profile HMMs for remote homology detection through structural alignment. *BMC Bioinformatics*, **8**, 435–447.

42. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

43. Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V. *et al.* (2006). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **35**, D5–D12.

44. Krammer, E. M., Sebban, P. & Ullmann, G. M. (2009). Profile hidden Markov models for analyzing similarities and dissimilarities in the bacterial reaction center and photosystem II. *Biochemistry*, **48**, 1230–1243.

45. Nogi, T., Fathir, I., Kobayashi, M., Nozawa, T. & 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.

46. Deisenhofer, J., Epp, O., Sinning, I. & Michel, H. (1995). Crystallographic refinement at 2.3 Å resolution and refined model of the photosynthetic reaction centre from *Rhodopseudomonas viridis*. *J. Mol. Biol.* **246**, 429–457.

47. Shatsky, M., Dror, O., Schneidman-Duhovny, D., Nussinov, R. & Wolfson, H. J. (2004). BioInfo3D: a suite of tools for structural bioinformatics. *Nucleic Acids Res.* **32**, W503–W507.

48. Crooks, G., Hon, G., Chandonia, J. & Brenner, S. (2004). WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190.

49. Zachariae, U. & Lancaster, C. R. D. (2001). Proton uptake associated with the reduction of the primary quinone $Q_A$ influences the binding site of the secondary quinone $Q_B$ in *Rhodopseudomonas viridis* photosynthetic reaction centers. *Biochim. Biophys. Acta*, **1505**, 280–290.

50. Breton, J., Boullais, C., Mioskowski, C., Sebban, P., Baciou, L. & Nabedryk, 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.

51. Pokkuluri, P. R., Laible, P. H., Crawford, A. E., Mayfield, J. F., Yousef, M. A., Ginsell, S. L. *et al.* (2004). Temperature and cryoprotectant influence secondary quinone binding position in bacterial reaction centers. *FEBS Lett.* **570**, 171–174.

52. Baxter, R. H. G., Seagle, B. L., Ponomarenko, N. & Norris, J. R. (2005). Cryogenic structure of the photosynthetic reaction center of *Blastochloris viridis* in the light and dark. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **61**, 605–612.

53. Brunger, A. T. & Karplus, M. (1988). Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins: Struct. Funct. Genet.* **4**, 148–156.

54. Brooks, B. R., Bruccoleri, B., Olafson, D., States, D., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization and dynamic calculations. *J. Comput. Chem.* **4**, 187–217.

55. Katona, G., Andréasson, U., Landau, E. M., Andréasson, L. E. & Neutre, R. (2003). Lipidic

cubic phase crystal structure of the photosynthetic reaction centre from *Rhodobacter sphaeroides* at 2.35 Å resolution. *J. Mol. Biol.* **331**, 681–692.

56. Onuchic, J. N., Beratan, D. N., Winkler, J. R. & Gray, H. B. (1992). Pathway analysis of protein electron-transfer reactions. *Annu. Rev. Biophys. Biomol. Struct.* **21**, 349–377.

57. Beratan, D. N., Betts, J. N. & Onuchic, J. N. (1991). Protein electron transfer rates set by the bridging secondary and tertiary structure. *Science*, **252**, 1285–1288.

58. Betts, J. N., Beratan, D. N. & Onuchic, J. N. (1992). Mapping electron tunneling pathways: an algorithm that finds the "minimum length"/maximum coupling pathway between electron donors and acceptors in proteins. *J. Am. Chem. Soc.* **114**, 4043–4046.

59. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **60**, 2126–2132.

60. Girvan, M. & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA*, **99**, 7821–7826.

61. Newman, M. E. & Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E: Stat. Nonlin. Soft Matter Phys.* **69**, 026113.