Proton Transfer Networks and the Mechanism of Long Range Proton Transfer in Proteins

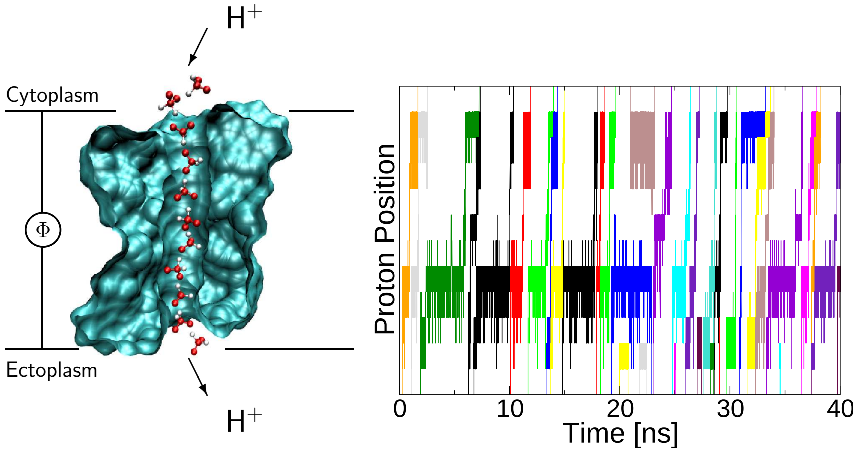
Dissertation zur Erlangung der Doktorwürde

der Fakultät für Biologie, Chemie und Geowissenschaften

der Unversität Bayreuth

Mirco S. Till

Februar 2009



Die vorliegende Arbeit wurde im Zeitraum Januar 2006 bis Januar 2009 an der Universität Bayreuth unter der Leitung von Prof. Dr. G. Matthias Ullmann erstellt.

1. Referee: Prof. Dr. G. Matthias Ullmann

Contents

Nomenclature

$[A]$	Concentration of Species A
$\Delta G^{b}_{\nu\mu}$	Activation Energy for the Reaction from μ to ν
$\Delta G_{\nu\mu}$	Reaction Free Energy for the Reaction from μ to ν
η	Reaction Rate
μ, ν	Microstates
A	Preexponential Factor
G_{Φ}	Influence of the Membrane Potential
G_{intr}	Intrinsic Energy
H	Enthalpy
k	Rate Constant
$P_{\nu}(t)$	Probability that the System is in State ν at Time t
S	Entropy
T	Temperature in Kelvin
$W(x_i, x_j)$	Interaction Energy between microstates i and j
ATP	Adenosine Triphosphate
gA	Gramicidin A
HBN	Hydrogen Bonded Network
LRPT	Long Range Proton Transfer
SDMC	Sequential Dynamical Monte Carlo
TST	Transition State Theory

Danksagung

Mein besonderer Dank gilt...

Prof. Dr. Matthias Ullmann, der mir die Möglichkeit gegeben hat, diese Arbeit in seiner Arbeitsgruppe durchzuführen und durch seine fachliche Unterstützung in vielen hervorragenden Diskussionen einen großen Beitrag zum Gelingen dieser Arbeit geleistet hat.

Dr. Torsten Becker, der maßgeblich an der Entwicklung der Methoden beteiligt war und nicht müde wurde, diese Entwicklungen zu diskutieren und voranzutreiben.

Dr. Timm Essigke, der nicht nur dafür gesorgt hat, dass das Netzwerk unserer Arbeitsgruppe stets allen Ansprüchen gerecht wurde sondern mir vor allem mit schier unendlicher Geduld in allen Fragen der Softwareentwicklung weiter geholfen hat. Außerdem für das Bereitstellen seines Programms QMPB. Danke!

Dr. Eva-Maria Krammer für die interessanten und produktiven Diskussionen über jedes Netzwerk, das ich ihr vorgelegt habe.

Der Arbeitsgruppe Strukturbiologie/Bioinformatik für das angenehme Arbeitsumfeld und die vielen guten Gespräche, die in diesem Umfeld stattgefunden haben.

Der liebsten Freundin der Welt, die während der gesamten Zeit für mich da war und mich jeden Tag aufs neue motiviert hat.

Meiner Mutter, die mir zu jeder Tag und Nacht Zeit nicht nur mit ihrem Wissen sondern vor allem mit ihrer Liebe zur Seite gestanden hat.

Dem Team des Enchilada Bayreuth, vor allem Armin, Alex und Harry, die es geschafft haben, Arbeit und Spaß an einem Ort zu verbinden.

All denen, die hier ungenannt bleiben, aber zu dieser Arbeit beigetragen haben, sei es durch Anregungen, Kritik oder Diskussionen rund um diese Arbeit oder aber dadurch, dass sie mich an manchen Tagen vom arbeiten abgehalten haben. Danke!

1 Summary

The main energy providing reaction systems in living cells, for example the photosynthesis or the respiratory chain, are based on long range proton transfer (LRPT) reactions. Even since these LRPT reactions have been heavily investigated in the last decades, the mechanism of these reactions is still not completely understood. The reaction kinetics of the LRPT are under heavy discussion and it is not clear, whether the reorientation of the hydrogen bond network (HBN) or the electrostatic barrier for the charge transfer is rate limiting.

The main purpose of this work is to investigate the dynamics of chemical reactions inside of proteins, focused on long range proton transfer reactions. Electron transfer reactions, rotations of water molecules or conformational changes of the protein are also considered. The developed sequential dynamical Monte Carlo (SDMC) method is applicable to almost all kinds of chemical reactions.

For all proton transfer reactions, the HBN of a protein plays a major role. Protons are transferred along such hydrogen bonds. Therefore, knowledge about the hydrogen bond network of a protein is crucial for the simulation of LRPT systems. The HBN can be calculated from the protein structure and the rotational state of the amino acid side chains. The reaction rate can be calculated from the electrostatic energies of the participating proton donor and acceptor groups. These two criteria are combined for the decision if a proton transfer between two molecules is possible and how fast this transfer would happen.

While the calculation of electrostatic energies of protonatable amino acid side chains or relevant cofactors in proteins (among them also water molecules) is already solved - implemented in various programs - the remaining tasks - calculating the hydrogen bond network followed by calculating the reaction rates - were solved during this work. Before the hydrogen bond network and the electrostatic energies could be calculated, the lack of water positions in many available crystallographically resolved protein structures made it necessary to develop an algorithm to detect internal cavities in proteins and fill these cavities with water molecules. The derived water positions could be included in the electrostatic calculations as well as in the calculation of the HBN.

The simulation of the LRPT in Gramicidin A (gA) compared to experimental data of the proton transfer in this polypeptide showed the possibilities of the simulation of the LRPT by the SDMC algorithm. The promising results encouraged us to investigate the mechanism of the LRPT, especially, if the reorientation of the HBN or the electrostatic energy barrier of the charge transfer is rate limiting for the LRPT. The results indicate, that both effects influence the LRPT and none of them is exclusively responsible for the LRPT rate.

Further analysis of the hydrogen bond network topology showed that graph algorithms can be used to analyze these networks. Hydrogen bond networks can be clustered into regions which are close connected to each other. On the other hand, residues connecting two or more of these densely connected regions might play an important role for proton transfer pathways since a loss of such residues cuts a proton transfer pathway. A comparison of an analysis of the HBN topology of the photosynthetic reaction center with mutation studies of the same system showed, that residues identified as important for proton transfer by the mutation studies are identified as connection points between clusters by the network analysis.

The developed algorithms together with the introduction of a new method for the simulation of the LRPT process (SDMC) improved the picture of the proton transfer processes in proteins. Starting from the protein structure, the developed algorithms cover all steps from the detection of protein cavities, the placement of water molecules in these cavities, the calculation and analysis of the hydrogen bond network, the simulation of the LRPT and the investigation of the reaction kinetics. The analysis of the HBN by graph theoretical methods gives further insight into the HBN topology and identifies residues important for proton transfer pathways and therefore important for the protein activity.

2 Zusammenfassung

Protonentransferreaktionen bilden in allen lebendigen Zellen die Grundlage für die wichtigsten energieliefernden Systeme wie zum Beispiel die Photosynthese. Obwohl diese Protonentransferreaktionen in den letzten Jahrzehnten mit großem Eifer untersucht wurden, ist der zugrunde liegende Mechanismus dieser Reaktionen noch nicht vollständig bekannt. Die Reaktionskinetiken der Protonentransferreaktionen innerhalb eines Proteins werden weiterhin diskutiert, da der limitierende Faktor der Reaktionen noch nicht klar ist. Es wird diskutiert, ob die Umordnung des Wasserstoffbrückennetzwerks oder die Energiebarriere des Ladungstransfers ratenbestimmend ist.

Ziel dieser Arbeit ist es, die Kinetiken von chemischen Reaktionen innerhalb von Proteinen zu erforschen, wobei das Hauptaugenmerk auf Protonentransferreaktionen liegt. Elektronentransferreaktionen, Rotationen von Wassermolekülen sowie Konformationsänderungen werden ebenfalls berücksichtigt. Die entwickelte Methode (Sequential Dynammical Monthe Carlo, SDMC) kann auf nahezu alle Arten von chemischen Reaktionen angewendet werden.

Das Wasserstoffbrückennetzwerk (WBN) eines Proteins spielt für alle Protonentransferreaktionen eine wichtige Rolle, da alle Protonentransferreaktionen entlang einer Wasserstoffbrücke erfolgen. Daher ist das Untersuchen des WBNs eines Proteins die Grundlage für die Simulation der Protonentransferkinetiken. Das WBN kann auf Grundlage der Proteinstruktur berechnet werden, wenn man alle Rotamere der einzelnen Aminosäuren einbezieht. Die Ratenkonstante einer Protonentransferreaktion kann aus dem Energieunterschied der beteiligten Donoren und Acceptoren berechnet werden. Diese beiden Kriterien zusammen bestimmen, ob ein Protonentransfer zwischen zwei Molekülen möglich ist und wie schnell dieser ablaufen wird.

Während die Berechnung der elektrostatischen Energien von protonierbaren Aminosäuren und wichtigen Kofaktoren (darunter auch Wasser) bereits durch viele verfügbare Programme gelöst ist, wurden die Algorithmen zur Berechnung des Wasserstoffbrückennetzwerks sowie die Berechnung der Reaktionskinetiken während dieser Arbeit entwickelt. Das Fehlen von Wasserpositionen in Röntgenstrukturen von Proteinen erforderte außerdem das Entwickeln eines Algorithmus zum Auffinden von Hohlräumen in Proteinen. Diese Hohlräume können anschließend mit Wassermolekülen gefüllt werden. Die erhaltenen Wasserpositionen werden in die Proteinstruktur integriert und bei den elektrostatischen Berechnungen berücksichtigt.

Die Simulation der Protonentransferkinetiken in Gramicidin A (gA) wurde mit experimentellen Daten verglichen und zeigte die Möglichkeiten des SDMC Algorithmus. Diese vielversprechenden Ergebnisse ermutigten uns auch den Mechanismus des Protonentransfers durch dieses Polypeptid zu untersuchen. Dabei wurde vor allem die Frage angegangen, ob die Umorientierung des Wasserstoffbrückennetzwerks oder die Energiebarriere des Ladungstransfers ratenbestimmend für den Protonentransfer ist. Die Ergebnisse deuten darauf hin, dass beide Effekte den Protonentransfer durch gA beeinflussen, bzw. keiner von beiden alleinig ratenbestimmend ist.

Bei der Betrachtung der Wasserstoffbrückennetzwerke zeigte sich, dass Algorithmen aus der Graphentheorie angewandt werden können, um diese Netzwerke zu analysieren. WBNs können in Bereiche (Cluster) unterteilt werden, die untereinander dichter verbunden sind. Auf der anderen Seite könnten Reste, die zwei oder mehr dieser Bereiche miteinander verbinden eine wichtige Rolle für Protonentransferpfade spielen, da ein Verlust dieser Reste das Unterbrechen eines solchen Pfads bedeuten würde. Ein Vergleich der Ergebnisse aus Mutationsstudien des bakteriellen Reaktionszentrums mit unserer Analyse der Netzwerktopologie zeigte, dass die Aminosäurereste, die bei den Mutationsstudien als wichtige Punkte für den Protonentransfer gefunden wurden in unseren Analysen als Verbindungspunkte zwischen Clustern auftraten.

Die entwickelten Algorithmen zur Netzwerkanalyse und die neu entwickelte Methode zur Simulation von Protonentransferkinetiken geben wichtige Einblicke in den gesamten Prozess des Protonentransfers in Proteinen, angefangen beim Auffinden von Hohlräumen in Proteinstrukturen über das Platzieren von Wassermolekülen in diesen Hohlräumen, die Berechnung und Analyse des WBN, die Simulation des Protonentransfers in Proteinen und die Betrachtung der Reaktionskinetiken dieser Prozesse. Außerdem gibt die Analyse des WBN Aufschluss über die Topologie solcher Netzwerke und kann Aminosäurereste identifizieren, die wichtig für den Protonentransfer und somit für die Funktion des Proteins sein können.

3 Introduction

Life is based on chemical reactions. At the very beginning of all living processes, RNA molecules were formed from sugar, a base and a phosphate group.[4,35] The chemical reactions forming the first RNA molecules may have started the evolution of live. The RNA molecules became building plans for proteins, proteins and RNA were grouped together in compartments known as cells today. All of these processes were based on chemical reactions and they still are based on chemical reactions. Every living cell produces proteins catalyzing chemical reactions which keep the cell alive. Amongst these reactions, proton transfer reactions may be the most important reactions.[24] The establishment of a proton gradient across the cell membrane is the key element of the energy housekeeping for every cell.[6,33] The proton gradient is established by proteins which are part of reaction mechanisms, using energy stored in energy rich molecules like sugar or energy sources like photons, to pump protons through the cell membrane out of the cell. The proton gradient is afterwards used to form adenosine triphosphate (ATP) , the general energy currency of the cell. ATP is necessary for almost all energy consuming reactions in the cell like biosynthesis, mobility or cell division. Two reaction cycles widely used for the establishment of the proton gradient are the respiratory chain and the photosynthesis.

The respiratory chain transforms electrochemical energy stored in NAD(P)H by oxidizing the NAD(P)H to NAD(P) into a proton gradient. During the oxidation, protons are pumped from the cytoplasm through the proteins of the respiratory chain to the ectoplasm. Following the chemiosmotic theory,[33] this gradient is afterwards used by the ATP synthetase[1] to store the energy of the proton gradient in the energy rich ATP molecule. During the ATP synthesis protons are transferred through the ATP synthetase along the proton gradient, providing the necessary energy for the ATP synthesis.[40]

The energy supply of all plants is based on a similar process. All photosynthetic active plants have light harvesting pigments,[14,21,31] collecting photons and transferring the energy of these photons to a photosynthetic reaction center. The photosynthetic reaction centers are located at the cell membrane using the energy provided by the light harvesting complexes to pump protons out of the cell.[22,42] The resulting proton gradient is again used to build ATP performed by an ATP synthetase.

3.1 Chemical reaction kinetics

3.1.1 The nature of chemical reactions

Looking at the processes inside a living cell, we can see, that all of these processes are based on chemical reactions. The formation of new covalent bonds catalyzed by enzymes, the transfer of protons along hydrogen bonds, the formation and breaking of hydrogen bonds, translocation or conformational changes of molecules, diffusion of molecules or the dissociation (breaking of covalent bonds). A chemical reaction is defined as the interconversion of one or several reactants into one or several products. Classically, during a chemical reaction the movement of electrons leads to breaking and forming of chemical bonds. Chemical reactions can therefore be grouped by their reaction character:

- The combination of two reactants to a single product (can be termed synthesis).
- The decomposition of a reactant into two or more products (can be termed analysis).
- The transfer of a part of one reactant to the other reactant, for example the transfer of a proton between two water molecules (can be termed substitution).

Acid-Base reactions as well as redox reactions can be seen as special types of substitutions. During Acid-Base reactions in water, an acid dissociates into the deprotonated acid and a proton (most likely forming an H_3O^+ ion), whereas a base accepts a proton from a water molecule leaving OH^-.[38] During redox reactions, the electron configuration of the reactants changes. Reaction kinetics can be described as a measure of how the concentration (or pressure) of the reaction partners change within time. Reaction kinetics are dependent on the concentration of the reactants, the available contact area, the pressure, an activation energy and the temperature. The concentration, pressure and contact area can be combined to the probability that all reactants necessary for the reaction meet at the same place. The activation energy and the temperature determine how fast the reaction takes place when the reactants are in contact with each other. The presence of a catalyst could also influence the reaction kinetics by lowering the activation energy.

Beside the activation energy, which determines if a reaction takes places at the moment all reactants meet, the energy levels of the reactants and products play a major role. Endothermic reactions, where the energy levels of the products are higher than the energy levels of the reactants consume energy during the reaction. Exothermic reactions, where the energy levels of the reactants are higher than the energy levels of the products free energy, most likely by releasing heat to the environment.

3.1.2 Reaction kinetics

Reaction kinetics describe the change in concentration or pressure of the reactants and products of a reaction. In 1864, Peter Waage and Cato Guldberg[44]developed the rate laws to describe experimental data of reaction kinetics in a mathematical way. In the following $[A]$ is the concentration of the species A at a certain point in time. We will look at the reaction $A + B \to C$ as an example for the explanation of rate laws.

Most reaction rates are dependent on the concentration of the reactants. The reaction rate η can therefore be expressed by

$$\eta = k[A][B] \tag{1}$$

where k is called the rate constant. The rate constant is independent of the concentrations but depends on the temperature. Eq. 1 is called the rate law of a reaction.The rate law is determined by experiment and can not be inferred from the chemical equation of the reaction. Once we have determined the rate law, we can predict the state of the reaction mixture at any point in time, based on the initial concentrations.

The reaction order is a simplistic description of the reaction. Many reactions are found to have rate laws of the form

$$\eta = k[A]^a[B]^b \tag{2}$$

The reaction order of such a rate law is $a + b$. A rate law like in Eq. 1 is a second order rate law. Reactions with a zero order rate law are independent of the reactant concentrations. Reactions where one reactant is in large excess can be simplified from a second order rate law to a first order rate law, since the concentration of the excess reactant is assumed to be constant. These rate laws are called pseudo first order rate laws. Reaction orders higher than 2 are unlikely, since a reaction order of, for example, three would mean, that three reactants have to meet at the same time. The probability of such an event is rather small. Therefore most of the reactions for which a high order rate law was found can be separated into a sequence of reactions with second order rate laws.

In order to find the concentration of reactants as a function of time, we need to integrate the rate laws. The first order rate law for the consumption of a reactant A

$$\frac{d[A]}{dt} = -k[A] \tag{3}$$

has the solution

$$[A] = [A]_0 e^{-kt} \tag{4}$$

Therefore, first order rate constants can be determined by plotting $ln(\frac{[A]}{[A]_0})$ against t. The slope of this straight line is the rate constant k.

Temperature dependence and Arrhenius law. Most of the known rate constants in chemistry increase with increasing temperature. Molecules with higher temperature have more thermal energy. The increased collision frequency of the molecules is one fact for the increased rate constant, but the major contribution is derived from the fact, that all reactions require an activation energy to take place. Fig. 2 shows the energy landscape of an exothermic reaction. The product state μ has a lower energy level than the reactant state ν. Therefore the reaction will occur spontaneous. The reaction still requires an activation energy. At higher temperatures, more molecules have sufficient energy to react, i.e. their thermal energy is higher than the activation energy. The amount of molecules, which have a high enough thermal energy is given by the Boltzmann distribution.

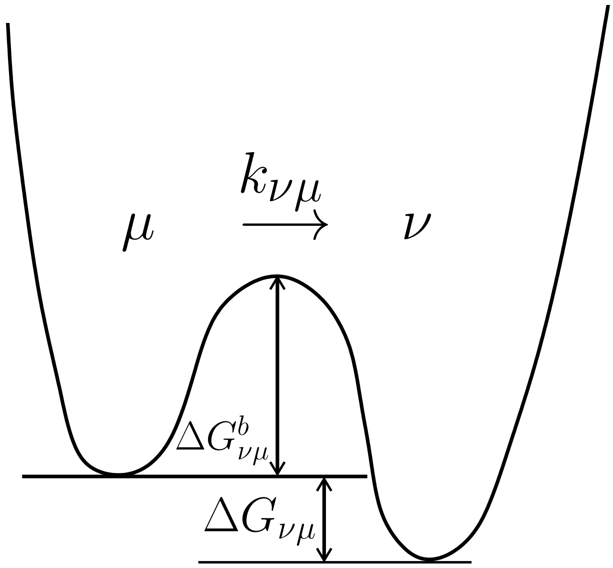


Figure 1: *Energy landscape of an exothermic reaction from the reactant state μ to the product state ν . $k_{\nu\mu}$ is the reaction rate constant, $\Delta G^b_{\nu\mu}$ is the activation energy or energy barrier of this reaction and $\Delta G_{\nu\mu}$ is the reaction free energy.*

It was found experimentally, that a plot of $ln(k)$ against T gives a straight line. The slope of this line can be used to determine the activation energy. The Arrhenius equation follows this empirical observations

$$lnk = lnA - \frac{E_a}{RT} \tag{5}$$

$$k = Ae^{\frac{-E_a}{RT}} \tag{6}$$

where A is the so called preexponential factor or frequency factor. E_a is the activation energy. The higher the activation energy, the stronger the temperature dependence of the rate constant. A zero activation energy indicates a temperature independent rate constant. The exponential character of the rate law can be explained as follows. In order to react, the reactants need a minimum amount of energy, the activation energy. At an absolute temperature, the fraction of molecules which have this energy as an kinetic energy are given by the Boltzmann distribution and are proportional to $e^{\frac{-E_a}{RT}}$. The preexponential factor is a measure of the rate at which the reaction would occur if there is no activation barrier. In other words, this is the maximal rate constant of an uncatalyzed reaction.

Since the Arrhenius law was based on empirical data, which was not satisfying, Eyring developed the Transition State Theory (TST)[13,25]. Comparing the Eyring equation

$$lnk = ln\frac{k_bT}{h} - \frac{\Delta H}{RT} + \frac{\Delta S}{R} \tag{7}$$

with the Arrhenius equation 6 shows a correlation between lnA and ΔS (the entropy of activation), and E_a and ΔH (the enthalpy of activation). In this work we assume the Arrhenius preexponential factor with 10^{13} according to the term $ln\frac{k_bT}{h}$ $(6 \cdot 10^{12}$ at room temperature) of the Eyring equation. The activation energy E_a is a given value for the energy barrier for exothermic reactions and is the same given barrier plus the Gibbs' free energy of the reaction for endothermic reactions (see Fig. 2). This approximation describes the proton transfer very well in our calculations.

3.1.3 The simulation of proton transfer reactions

The simulation of proton transfer reactions is the aim of many computational approaches. Two factors influence the possibilities of most of the know methods. On the one hand, breaking and forming of bonds is necessary to simulate chemical reactions. On the other hand, the reactions occur on very different time scales. The proton transfer between two molecules is very fast, in the picosecond timescale. The LRPT through a whole protein can take several milliseconds. The simulation of the long range proton transfer needs to simulate these time spans but with the accuracy of the fastest step, the proton transfer. Two blocks of well known approaches[8,9,17,30,32,39,43,45–47] are ruled out by these criteria. Molecular dynamics simulations, which might be capable of simulating the fast reactions on the picosecond time scale over several milliseconds are not able to simulate bond breaking or forming. Quantum mechanical methods on the other hand, are able to simulated the formation and breakage of bonds, but are way to slow to reach milliseconds in simulation time.

The method developed in this work solves the problems of the simulation of the LRPT as described above by solving the master equation for a proton transfer system using a Monte Carlo approach. The biological charge transfer is described as a transition between microstates of the system where one microstate is represented by a state vector. Each element of this vector represents the state of one site. For example, the state vector of a proton transfer system with three sites might look like $[010]$. The second site of this system is protonated, the other two sites are deprotonated. Thus, assuming p possible states for n sites, there could be p^n possible microstates, for example a proton transfer from site two to site three would be a microstate transition from $[010]$ to $[001]$. A charge transfer within this system is a transition between two of these microstates. Each of these transitions is defined by only one charge transfer. The transfer rate of this charge transfer is also the rate for the transition between these microstates. The transfer rates for proton transfer are calculated using the Arrhenius law. For each microstate, the set of possible transitions is limited by the possible proton transfer reactions, i.e. a proton transfer from site two to site three is only possible if site two is protonated and site three is deprotonated.

The master equation[5,15] describes the time evolution of such a microstate system:

$$\frac{d}{dt}P_\nu(t) = \sum_{\mu=1}^{M} k_{\nu\mu}P_\mu(t) - \sum_{\mu=1}^{M} k_{\mu\nu}P_\nu(t) \tag{8}$$

where $P_\nu(t)$ denotes the probability that the system is in state ν at time t, $k_{\nu\mu}$ denotes the probability per unit time that the system will change its state from μ to ν or in other words the rate at which the system changes from μ to ν. For small systems, solving Eq. 8 numerically is possible. By using Arrhenius law as described above to calculate the rate constants for the state transition and tabelize these rate constants for all possible transitions between microstates, one could calculate the time evolution of a proton transfer system with the resolution of the fast proton transfer reaction but over large simulation times.[5] Unfortunately, the number of microstates even of small biological systems is too large and solving the master equation of those systems directly is impossible.

3.2 Sequential Dynamical Monte Carlo

As mentioned above, solving the mater equation for a biological system of moderate size directly is computationally prohibited since the number of possible microstates is overwhelming. However most of these microstates are never populated, meaning that the probability P_ν

in Eq. 8 is near 0, since they are energetically unfavored compared to other microstates. Microstates with a high energy are never or only occasionally reached. Cancel these microstates out of the reaction mechanism would introduce a bias which consequences are hard to estimate. The solution to this problem presented in this work is a Sequential Monte Carlo Algorithm (SDMC), which is based on an algorithm developed by Gillespie.[18,26] Rate constants are only calculated, if they lead away from states, which are populated during the simulation. The algorithm starts at a given microstate and a given point in time. The algorithm decides - based on two criteria which are influenced by the rate constants of all reactions and therefore influenced by the difference in energy between the microstates - which microstate will be populated in the next time span. The criteria ensure, that energetically favorable microstates are populated more often than energetically unfavorable microstates, or in more detail, that the microstates are populated according to the Boltzmann distribution under equilibrium conditions. Letting the system evolve for a number of steps and averaging over the recorded trajectories gives a correct description of the time evolution of the system without the need of solving the master equation directly or calculating the whole partition function of the system.

Starting from a given microstate, two criteria are utilized to chose which reaction will take place in what time span. The first criteria[16,18] chooses which reaction m will take place:

$$\sum_{l=1}^{m-1} k_l \leq \rho_1 K < \sum_{l=1}^{m} k_l \tag{9}$$

$$K = \sum_{l=1}^{L} k_l \tag{10}$$

K is the sum of the rate constants k_l of all L possible events for the given microstate; ρ_1 is a random number between 0 and 1. For each step of the algorithm, all possible reactions are determined and the rate constants for each possible reaction is calculated. These rate constants depend on the electrostatic energy of the participating microstates. The criteria described in Eq. 9 ensures, that a reaction a which is twice as fast as a reaction b - $k_A = 2 \cdot k_B$ - is selected twice as often as reaction b.

The time span Δt which elapsed during the Monte Carlo step is chosen by

$$\Delta t = \frac{1}{K} \ln[\frac{1}{\rho_2}] \tag{11}$$

which is a standard way to draw a random number Δt from an exponential distribution given a uniformly distributed random number ρ_2 between 0 and 1. Applying these two criteria on the set of possible reactions in each step of the Monte Carlo simulation ensures a correct description of the time evolution of a given microstate system. For each step, only the reaction rate constants of the possible reactions need to be calculated based on the current microstate, reducing the number of rate calculations by orders of magnitude compared to the number of calculations necessary to solve the master equation directly. However, calculating the reaction rate constants is still the crucial part of the simulation. The general workflow of the SDMC algorithm can be seen in Fig. 3. Starting from the initial microstate, all possible reactions are determined. Reaction rates are calculated and the next step is chosen. After a determined number of steps, the simulation terminates. As described above, the Arrhenius law together with the transition state theory provides a good approximation for the reaction rate constants of proton transfer reactions. To calculate the electrostatic energy difference between the two participating microstates, the linearized Poisson-Boltzmann equation was solved using the Poisson-Boltzmann solver of the mead package implemented by the QMPB-program.

3.2.1 Electrostatic calculations

The electrostatic energies used for the rate calculations during the SMDC simulation are calculated using the microstate description as explained above. Three energies contribute to the electrostatic energy of a microstate. The so called intrinsic energy ($G_{intr}(x_i)$) , the influence of the membrane potential ($G_{\Phi}(x_i)$) and the interaction energy ($W(x_i, x_j)$) between each pair of sites for all instances. Therefore, the electrostatic energy of a microstate is expressed in the following sum:

$$G^{\circ}_{\nu} = \sum_{i=1}^{N} \Big(G_{intr}(x_i) + G_{\Phi}(x_i) \Big) + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} W(x_i, x_j) \tag{12}$$

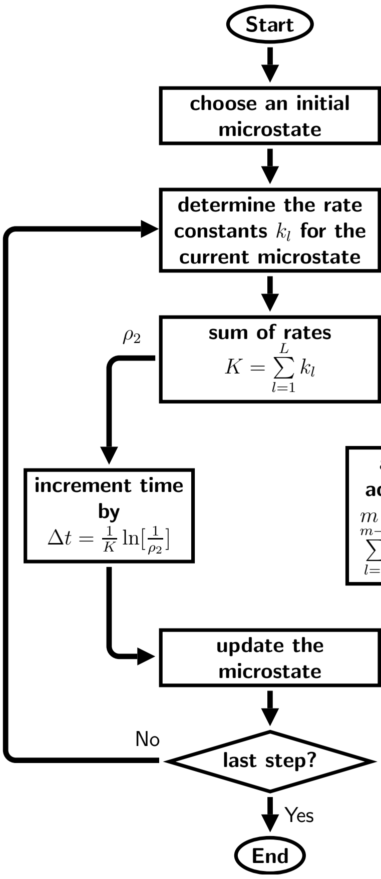

Start
choose an initial microstate
determine the rate constants k_l for the current microstate
sum of rates $K = \sum_{l=1}^{L} k_l$
ρ_2
ρ_1
increment time by $\Delta t = \frac{1}{K}\ln[\frac{1}{\rho_2}]$
advance system according to event m chosen such that $\sum_{l=1}^{m-1} k_l \leq \rho_1 K < \sum_{l=1}^{m} k_l$
update the microstate
last step?
No
Yes
End

Figure 2: *Flowchart of the sequential dynamical Monte Carlo algorithm. Starting from a specified microstate, the rate constants for all possible reactions which lead away from this microstate are calculated. The reaction which takes place and the time increment is determined. The microstate of the system is updated with the information from the chosen reaction rate and the time is incremented. If the termination criteria are not met, the next simulation step starts again with the calculation of the reaction rate constants.*

The energy contributions are calculated by solving the linearized Poisson-Boltzmann equation. The derived energy contributions for all instances of all microscopic sites as well as the derived interaction energies between all pairs of instances are tabelized. These tables are part of the input for the SDMC calculations. Reaction rates are calculated by using Arrhenius law (see Eq. 6). The activation barrier E_a for each reaction is calculated from the energy difference between the reactant and the product microstate and a constant energy barrier for the reaction. For exothermic reactions, E_a is equal to the energy barrier of the reaction, for endothermic reactions E_a is equal to the energy barrier of the reaction plus the energy difference between the reactant and product microstate (see Fig. 2).

3.2.2 Two possible mechanisms of LRPT

Simulating the long range proton transfer through the Gramicidin A membrane channel led us to a discussion about the general mechanism of the long range proton transfer from a more generalized point of view. The proton transfer rate in water is much faster than an estimated diffusion rate of protons in water. In 1809, Grotthuss[2,34] published his mechanism of long range proton transfer as a chain of subsequent hopping events between water molecules. If these water molecules are already oriented in a hydrogen bond network, the transfer of a proton from one end of a chain to the other end is fast, since the proton which is transferred between water molecule one and two is not necessarily the proton, which is transferred through the whole chain. After such a proton transfer along a water chain, the water chain needs to reorient to form new hydrogen bonds between the water molecules. Grotthuss suggested this reorientation as rate limiting for the long range proton transfer.

Braun-Sand et al.[8] published a mechanism for the long range proton transfer and identified the electrostatic energy barrier of the charge transfer as rate limiting.

By simulating the long range proton transfer through gA, we addressed the question of the long range proton transfer mechanism by investigating the influence of the rotation rate as well as the electrostatic energy barrier on the long range proton transfer.

3.2.3 The Hydrogen Bond Network of a Protein

A mandatory prerequest for a proton transfer is an established hydrogen bond. Proton transfer can be seen as a relatively small translocation of the hydrogen atom along the axis of an already existing hydrogen bond. A small energy barrier has to be crossed on the way from a location near the donor heavy atom (like oxygen or nitrogen) towards a location closer to the acceptor heavy atom. In bulk water, such a proton transfer reaction has a free reaction energy of 0.0 kcal/mol and an energy barrier which is rather small, less than 0.5 kcal/mol.

Since the SDMC algorithm is capable of simulating the long range proton transfer within an hydrogen bond network, the definition of such an hydrogen bond network within a protein is the first step, which needs to be done.

The Definition of a Hydrogen Bond The main element of each HBN, the hydrogen bond itself is a very diffuse definition. In general one can say, that a hydrogen bond is possible between an electronegative heavy atom and a hydrogen, bound to an electronegative atom if the distance of the heavy atoms is less than 4-5 Å. An example of such a combination is $OH--O$ where the $O--O$ distance is less than 4-5 Å. Additionally the angle spanned by the three atoms is used as a criteria for the possibility and the strength of a hydrogen bond. The angle range for a possible hydrogen bond varies with a maximum of 55°around 180°.

Analyzing a Hydrogen Bonded Network After identifying all hydrogen bonds within a protein, one can calculate one or more hydrogen bond networks. A network (or graph) in a mathematical sense is composed of an arbitrary number (more than one) of nodes and edges, which connect these nodes. A hydrogen bond network within a protein is therefore also a bidirectional graph in a mathematical sense. Bidirectional means, that the connections with the network have no direction. This is true for hydrogen bond networks if we focus on the possibility of proton transfer. Each pair of hydrogen bond partners can transfer a proton in both directions.

For the identification of hydrogen bond networks, we applied a breath first search. The algorithm finds connected graphs within a given set of nodes and edges. Connected means,

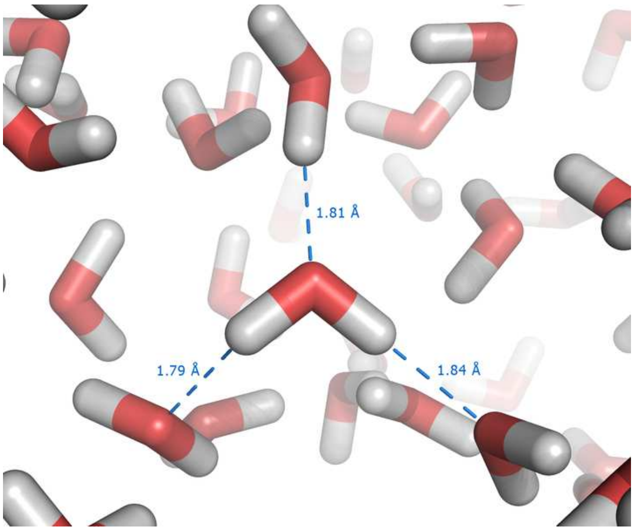


Figure 3: *An example for a small Hydrogen bond network of four water molecules. The Oxygen-Oxygen distance is less than 4 Å and the angle spanned by the two oxygen atoms and the Hydrogen atom varies less than 55°around 180°*

that each node within a graph is reachable from a second node by walking along the edges. The hydrogen bond network of a protein can therefore be parted into several unconnected subgraphs.

Analyzing the structure of such networks is the aim of clustering[19] connected graphs. Clustering tries to identify nodes with many connections between each others. Regions which are densely connected are called clusters. Applied to an hydrogen bond network, one could identify amino acid side chains which are heavily connected. On the other hand, one can identify important connections within a hydrogen bond network by looking at connections between two clusters. The loss of a connection between two clusters might be harder to compensate than the loss of a connection within a cluster. Analyzing the clustering of proton transfer networks gives insight into the proton transfer pathways within a protein, identifies possible proton entry points and predicts important connections or residues of proton transfer pathways.

3.2.4 Detecting Cavities and Surface Clefts in Proteins

Water molecules are of central importance for all proton transfer processes in proteins,[7,46] since water is not only the solvent for all proteins, water molecules can also be located in the

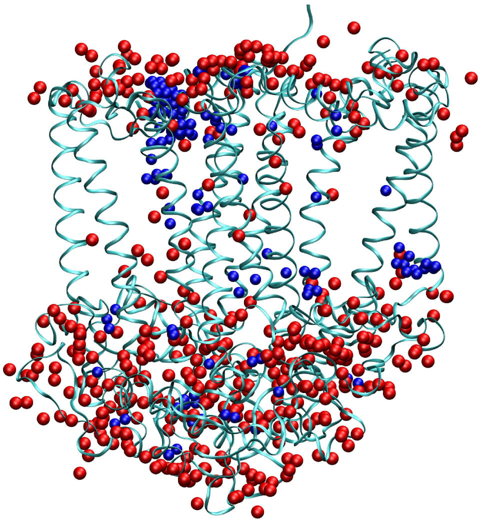


Figure 4: *Water molecules placed in cavities and surface clefts of the bacterial photosynthetic reaction center of rhodobacter sphaeroides placed by the McVol algorithm.*

protein interior. Since the mobility of these internal water molecules is relatively high compared to the robust protein backbone, these water molecules are not completely resolved by x-ray crystallography. Identifying protein cavities and placing water molecules in these cavities can have a strong influence on the simulation of the long range proton transfer in Proteins. The known algorithms[11,20,27–29,48] for this task suffer from several problems: If these algorithms are grid based, the resolution of the cavity detection is dependent on the grid resolution. The alpha shape theory,[29] independent of grid resolutions, is numerically not stable and is not always perfectly accurate for identifying surface clefts.

Detecting cavities or surface clefs in proteins is related to the problem of integrating the protein volume. Monte Carlo algorithms[37] have shown to solve these problems satisfyingly. Therefore we developed an algorithm to calculate the protein volume and detect protein cavities and surface clefts using a Monte Carlo method. This algorithm is independent of grid resolutions and not prone to numerical instabilities. The defined cavities were separated from the solvent by graph algorithms already invented for the hydrogen bond network analysis. Identified cavities and surface clefts were filled with water molecules in dependence of their size. The possibility of identifying cavities in proteins completed the task of simulating the proton transfer within proteins.

3.3 Proteins investigated in this work

During this work two systems were chosen for the application of the developed methods as well as to validate the new methods by comparison with experimentally determined data.

3.3.1 Gramicidin A

Gramicidin A (structure taken from pdb code 1jno[41]) is a well-studied system[3,10,12] consisting of to peptides in a helical secondary structure. The peptides are arranged in a head to head dimer, forming a channel through the cell membrane. The channel is filled by a water chain of about 11 water molecules. Protons can be transferred along this water chain. Beside protons, other cations can diffuse through the channel, however this diffusion is much slower than the proton transport. Gramicidin A perfectly fulfills the role of a test system. The system is very small and the proton transport is only mediated by the eleven water molecules located in the center of the channel. The peptides only provide additional hydrogen bond partners for the water molecules but do not take part in the proton transfer. Proton transfer rates were measured experimentally for the gA channel.

Gramicidin A was used as a test system to validate the correct simulation of the LRPT through this channel by comparing the experimentally derived data with the simulations performed with the SDMC algorithm.

3.3.2 Bacterial Photosynthetic reaction center

The detection of hydrogen bond networks and the graph-theoretical analysis of these networks was developed, tested an applied to two structures of the bacterial photosynthetic reaction center: One structure from Rb. sphaeroides (PDB code 2J8C[23]), the other one from Blastochloris viridis (PDB code 1EYS[36]). Both proteins span a large hydrogen bond network connecting the cytoplasmic bulk water with the ubiquinone cofactor. The proton entry points, i.e. the amino acid side chains which take up protons from the cytoplasmic bulk phase are not completely identified and the proton transfer pathways from the cytoplasmic bulk phase to ubiquinone are under heavy discussion. It is even not clear if the protons are transferred via distinct proton transfer pathways at all or if the protein works as a proton sponge, i.e. that the protons are transferred via groups of residues instead of certain special residues.

3.4 Aim of this Theses

The aim of this theses was to get insight into the reaction mechanisms of proton transfer reactions and the simulation of these reactions inside proteins. For the simulation of the long range proton transfer, a new method was developed, called SDMC. This method is able to simulate the proton transfer processes over time spans not accessible by other methods. This new method was applied to the proton transfer system of the Gramicidin A channel gaining new insights in the LRPT mechanism of the peptide as well as more knowledge about the rate limiting element of this LRPT. The analysis of hydrogen bond networks with graph-theoretical methods was, to the best of my knowledge, never before applied on proteins. A better understanding of the network topology, identification of key residues and knowledge whether the proton transfer in the photosynthetic reaction center is organized via distinct pathways or via a proton sponge were the results of this analysis.

References

[1] J. P. Abrahams, A. G. Leslie, R. Lutter, and J. E. Walker. Structure at 2.8 a resolution of f1-atpase from bovine heart mitochondria. *Nature*, 370(6491):621–628, Aug 1994.

[2] Noam Agmon. The grotthuss mechanism. *Chem. Phys. Lett.*, 244:456–462, 1995.

[3] O. S. Andersen, R. E. Koeppe, and B. Roux. Gramicidin channels. *IEEE Trans. Nanobioscience*, 4:10–20, 2005.

[4] S. Barazesh. How rna got started. *Science News*, 175(12):5, 2009.

[5] T Becker, R. T. Ullmann, and G. M. Ullmann. 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, 2007.

[6] J.M. Berg, J. L. Tymoczka, and L. Stryer. *Biochemie*. Spektrum, 2003.

[7] A. N. Bondar, M. Elstner, S. Suhai, J. C. Smith, and S. Fischer. Mechanism of Primary Proton Transfer in Bacteriorhodopsin. 12:1281–1288, 2004.

[8] S. Braun-Sand, A. Burykin, Z. T. Chu, and A. Warshel. Realistic simulations of proton transport along the gramicidin channel: demonstrating the importance of solvation effects. *J. Phys. Chem. B*, 109:583–592, 2005.

[9] S. Braun-Sand, M. Strajbl, and A. Warshel. Studies of proton translocations in biological systems: simulating proton transport in carbonic anhydrase by evb-based models. *Biophys. J.*, 87:2221–2239, 2004.

[10] B. M. Burkhart, N. Li, D. A. Langs, W. A. Pangborn, and W. L. Duax. The conducting form of gramicidin a is a right-handed double-stranded double helix. *Proc. Natl. Acad. Sci. U.S.A.*, 95:12950–12955, 1998.

[11] H. Edelsbrunner and E. P. Mucke. Simulation of Simplicity - A Techique to Cope with Degenerate Cases in Geometric Algorithms. *ACM Transactions on Graphics*, 9:66–104, 1990.

[12] G. Eisenman, B. Enos, J. Hagglund, and J. Sandblom. Gramicidin as an example of a single-filing ionic channel. *Ann. N.Y. Acad. Sci.*, 339:8–20, 1980.

[13] Henry Eyring. The activated complex in chemical reactions. *The Journal of Chemical Physics*, 3(2):107–115, 1935.

[14] Jack Fajer. Chlorophyll chemistry before and after crystals of photosynthetic reaction centers. *Photosynth Res*, 80(1-3):165–172, 2004.

[15] A. M. Ferreira and D. Bashford. Model for proton transport coupled to protein conformational change: Application to proton pumping in the bacteriorhodopsin photocycle. *J. Am. Chem. Soc.*, 128:16778–16790, 2006.

[16] K. A. Fichthorn and W. H. Weinberg. Theoretical foundations of dynamical monte carlo simulations. *J. Chem. Phys.*, 95:1090–1096, 1991.

[17] R. Friedman, E. Nachliel, and M. Gutman. Application of classical molecular dynamics for evaluation of proton transfer mechanism on a protein. *Biochim. Biophys. Acta*, 1710:67–77, 2005.

[18] D T Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[19] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.*, 99:7821–7826, 2002.

[20] M. Hendlich, F. Rippmann, and G. Barnickel. LIGSITE: Automatic and Efficient Detection of Potential Small Molecule-binding Sites in Proteins. *J. Mol. Graph.*, 15:359, 1997.

[21] W. Hillier and G. T. Babcock. Photosynthetic reaction centers. *Plant Physiol*, 125(1):33–37, Jan 2001.

[22] Pierre Joliot, Anne Joliot, and André Verméglio. 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(3):204–214, Feb 2005.

[23] J. Koepke, E.-M. Krammer, A. R. Klingen, P. Sebban, G. M. Ullmann, and G. Fritzsch. 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, 2007.

[24] L. I. Krishtalik. The mechanism of the proton transfer: an outline. *Biochim. Biophys. Acta*, 1458:6–27, 2000.

[25] K. J. J. Laidler. Development of transition-state theory. *J. Phys. Chem.*, 87(15):1657, 1983.

[26] D P Landau and K Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, 2000.

[27] R. A. Laskowski. SURFNET - A Program For Visualizing Molecular Surfaces, Cavities and Intermolecular Interactions. *J. Mol. Graph.*, 13:323, 1995.

[28] D. G. Levitt and L. J. Banaszak. POCKET - A Computer-graphics Method for Identifying and Displaying Protein cavities and their Surrounding Amino acids. *J. Mol. Graph.*, 10:229–234, 1992.

[29] J. Liang, H. Edelsbrunner, and C. Woodward. Anatomy of Protein Pockets and Cavities: Measurement of Binding Site Geometry and Implications for Ligand Design. *Prot. Sci.*, 7:1884–1897, 1998.

[30] M. A. Lill and V. Helms. Molecular dynamics simulation of proton transport with quantum mechanically derived proton hopping rates (q-hop md). *J. Chem. Phys.*, 115:7993–8005, 2001.

[31] P. J. Lockhart, A. W. Larkum, M. Steel, P. J. Waddell, and D. Penny. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. *Proc Natl Acad Sci U S A*, 93(5):1930–1934, Mar 1996.

[32] D. Marx. Proton transfer 200 years after von grotthuss: insights from ab initio simulations. *Com. Phys. Comm.*, 7:1848–1870, 2006.

[33] P. Mitchell. Chemiosmotic coupling in energy transduction: A logical development of biochemical knowledge. *Bioenergetics*, 3:5–24, 1972.

[34] J. F. Nagle and H. J. Morowitz. Molecular mechanisms for proton transport in membranes. *Proc. Natl. Acad. Sci. U.S.A.*, 75:298–302, 1978.

[35] J. Netting. Rna world gets support as prelife scenario. *Science News*, 159(14):212, 2001.

[36] T Nogi, I Fathir, M Kobayashi, T Nozawa, and K Miki. Crystal structures of photosynthetic reaction center and high-potential iron-sulfur protein from thermochromatium tepidum: Thermostability and electron transfer. *Proc. Natl. Acad. Sci. U.S.A.*, 97(25):13561–13566, DEC 5 2000.

[37] M. H. M. Olsson and A. Warshel. Monte carlo simulations of proton pumps: On the working principles of the biological valve that controls proton pumping in cytochrome c oxidase. *Proc. Natl. Acad. Sci. U.S.A.*, 103:6500–6505, 2006.

[38] Ralph G. Pearson. Acids and bases. *Science*, 151(3707):172–177, Jan 1966.

[39] U. W. Schmitt and G. A. Voth. The computer simulation of proton transport in water. *J. Chem. Phys.*, 111:9361–9381, 1999.

[40] D. Stock, A. G. Leslie, and J. E. Walker. Molecular architecture of the rotary motor in atp synthase. *Science*, 286(5445):1700–1705, Nov 1999.

[41] L. E. Townsley, W. A. Tucker, S. Sham, and J. F. Hinton. Structures of gramicidins a, b, and c incorporated into sodium dodecyl sulfate micelles. *Biochemistry*, 40:11676–11686, 2001.

[42] Andre Vermeglio and Pierre Joliot. Supramolecular organisation of the photosynthetic chain in anoxygenic bacteria. *Biochim Biophys Acta*, 1555:60–64, Sep 2002.

[43] G. A. Voth. Computer simulation of proton solvation and transport in aqueous and biomolecular systems. *Acc. Chem. Res.*, 39:143–150, 2006.

[44] P Waage and C. M. Guldberg. Forhandlinger: Videnskabs-selskabet i. *Christiana*, 1864.

[45] A. Warshel. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem. Rev.*, 93:2523–2544, 1993.

[46] A. Warshel. Molecular Dynamics Simulations of Biological Reactions. *Acc. Chem. Res.*, 35:385–395, 2002.

[47] A. Warshel and R. M. Weiss. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.*, 102:6218–6226, 1980.

[48] L. Xie and P. E. Bourne. A Robust and Efficient Algorithm for the Shape Description of Protein Structures and its Application in Predicting Ligand Binding Sites. *Bioinformatics*, 8:–, 2007.

4 Manuscripts

The central issue of this work was to gain further insights into the reaction kinetics of the long range proton transfer reactions inside of proteins. While the core reaction, a single proton transfer between two molecules which already form a hydrogen bond was already well studied by quantum chemical approaches, the mechanism of the proton transfer through a whole protein is still under discussion. Two elements, the reorientation of the hydrogen bond network or the energy barrier for the charge transfer are supposed to be rate limiting for the long range proton transfer. Solving the master equation for a proton transfer system described in a microstate formalism could solve some of the open questions. However, solving the master equation analytically is only possible for very small systems. The solution for this problem was the development of a Sequential Dynamical Monte Carlo algorithm (Manuscript A). The algorithm is based on an algorithm written by Gillespie which is known to solve the master equation statistically. Since the proton transfer reactions studied in this work are sequential, the Gillespie-algorithm was developed further to be able to simulate the sequential hopping events of a long range proton transfer system. This algorithm was applied to simulate the proton transfer system of the Gramicidin A channel gaining insight into the mechanism of the proton transfer in this system and addressing the question which of the two mentioned elements is rate limiting.

The SDMC algorithm requires knowledge of the proton transfer (or hydrogen bond) network of the system. The calculation of these networks is split into two problems. Water molecules not resolved in x-ray structures need to be placed in protein cavities inside proteins as well as in surface clefts (Manuscript B). To detect these cavities and clefts a Monte Carlo based algorithm for calculating the protein volume was developed, implemented and tested on several proteins. During the development, protein structures were compared with respect to their atom densities showing that proteins have a very similar atom to volume ratio independent of their size. The second problem of the hydrogen bond network calculations was the detection of hydrogen bonds. An algorithm based only on atom-atom distances was developed giving fast and accurate description of the hydrogen bond network of a whole protein.

The analysis of the hydrogen bond networks of the photosynthetic reaction centers from two bacterial species (Manuscript C) implied the application of graph theoretical methods on the hydrogen bond networks. To the best of my knowledge, this was the first time that graph theoretical methods were applied on hydrogen bond networks. The cluster analysis of the networks gained insight into the structural organization of these networks. Amino acid residues important for the long range proton transfer could be identified in agreement with experiments as well as proton entry points were found extending the list of already known points.

The work described in the manuscripts A to C completely covers the simulation of proton transfer by the new developed SDMC algorithm starting from the placement of water molecules in cavities, analysis of the proton transfer network up to the simulation of the whole proton transfer through the Gramicidin A channel by the SDMC algorithm.

4.1 Synopsis of the Manuscripts

Manuscript A:

Simulating the Proton Transfer in Gramicidin A by a Sequential Dynamical Monte Carlo Method

Mirco S. Till, Timm Essigke, Torsten Becker,* and G. Matthias Ullman

Received: February 19, 2008; Revised Manuscript Received: June 3, 2008

J. Phys. Chem. B 2008, 112, 13401 - 13410

DOI: 10.1021/jp801477b

The focus of this work was the development, implementation and validation of the SDMC algorithm. The SDMC algorithm is based on the Gillespie algorithm and was further developed to simulate the sequential hopping events of long range proton transfer systems. The implementation of the SDMC algorithm was tested and validated by simulating the proton transfer through the gA channel. The algorithm was able to simulate the proton transfer through the channel in good agreement with experimental data. After validating the new method with these simulations, we investigated the proton transfer mechanism in the gA channel addressing the question whether the reorientation of the hydrogen bond network or the energy barrier for the charge transfer is rate limiting. we could show, that as long as none of the two parameters is artificially set to extreme values, both of them influence the long range proton transfer on a similar level.

Together with G. Matthias Ullmann and Torsten Becker I developed the theory for the sequential dynamical Monte Carlo approach. For the electrochemical calculations I used a program written by Timm Essigke. Developing the SDMC algorithms, testing the software and applying this software to the Gramicidin A system was done by me.

Manuscript B:

McVol - A program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm

Mirco S. Till & G. Matthias Ullmann

Received: 31 March 2009 / Accepted: 23 May 2009

J Mol Model. 2009 Jul 22. [Epub ahead of print]

DOI 10.1007/s00894-009-0541-y

The detection of integral protein cavities as well as surface clefts on proteins was a crucial step during the calculation of the hydrogen bond network of proteins as well as the simulation of the long range proton transfer. Since all available methods were prone to errors, I developed together with G. Matthias Ullmann a Monte Carlo algorithm which is able to calculate the volume of a protein and detect cavities and clefts without numerical instabilities. The algorithm is fast and accurate, which was tested by identifying cavities in the hen egg lysozyme which where also detected by experiment. The gained data sets enabled us to analyse the atom density and volume to void ratio within proteins which both showed to be independent of the protein size.

My contribution to this work was the development of the algorithms for the graph searches (separating the cavities from the solvent), the water placement and the definition of the surface clefts and pockets. Furthermore I ported the algorithms developed by G. Matthias Ullmann (Monte Carlo volume calculation and neighbor lists) to C++ for a better abstraction of the sources. All calculations done for this paper were also my contribution.

Manuscript C:

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

Received 7 January 2009, accepted 8 March 2009

J. Mol. Biol. (2009) 388, 631 - 643

DOI:10.1016/j.jmb.2009.03.020

The availability of a fast algorithm for the calculation of hydrogen bond networks and the fact, that a hydrogen bond network can be expressed as a graph in mathematical sense implied to apply graph search and clustering algorithms to these networks. Together wit Eva-Maria Krammer, I compared the hydrogen bond networks of the photosynthetic reaction centers from two bacterial species. We clustered the networks using two different clustering methods. Using the betweenness clustering algorithm brought the best results. By analyzing the clustering of these networks we were able to identify amino acid residues important for the proton transfer from the cytoplasm to the Qb which were already identified by mutation experiments. We were also able to add some amino acid residues to the list of possible proton entry points. This was the first time that graph theoretical methods were applied to hydrogen bond networks.

While the sequence alignments were contributed by Eva Maria Krammer, I developed the algorithms to calculate hydrogen bond networks, search for connected graphs in these networks and cluster them by the two described methods. We combined our results and discussed them with G. Matthias Ullmann and Pierre Sebban. The results of the calculations and the conclusions from the discussions are shown in this publication.

4.2 Manuscript A

Simulating the Proton Transfer in Gramicidin A by a Sequential Dynamical Monte Carlo

Mirco S. Till, Timm Essigke, Torsten Becker,* and G. Matthias Ullman

Received: February 19, 2008;

Revised Manuscript Received: June 3, 2008

J. Phys. Chem. B 2008, 112, 13401 - 13410

DOI: 10.1021/jp801477b

Simulating the Proton Transfer in Gramicidin A by a Sequential Dynamical Monte Carlo Method

Mirco S. Till, Timm Essigke, Torsten Becker,* and G. Matthias Ullmann*

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

Received: February 19, 2008; Revised Manuscript Received: June 3, 2008

The large interest in long-range proton transfer in biomolecules is triggered by its importance for many biochemical processes such as biological energy transduction and drug detoxification. Since long-range proton transfer occurs on a microsecond time scale, simulating this process on a molecular level is still a challenging task and not possible with standard simulation methods. In general, the dynamics of a reactive system can be described by a master equation. A natural way to describe long-range charge transfer in biomolecules is to decompose the process into elementary steps which are transitions between microstates. Each microstate has a defined protonation pattern. Although such a master equation can in principle be solved analytically, it is often too demanding to solve this equation because of the large number of microstates. In this paper, we describe a new method which solves the master equation by a sequential dynamical Monte Carlo algorithm. Starting from one microstate, the evolution of the system is simulated as a stochastic process. The energetic parameters required for these simulations are determined by continuum electrostatic calculations. We apply this method to simulate the proton transfer through gramicidin A, a transmembrane proton channel, in dependence on the applied membrane potential and the pH value of the solution. As elementary steps in our reaction, we consider proton uptake and release, proton transfer along a hydrogen bond, and rotations of water molecules that constitute a proton wire through the channel. A simulation of 8 μs length took about 5 min on an Intel Pentium 4 CPU with 3.2 GHz. We obtained good agreement with experimental data for the proton flux through gramicidin A over a wide range of pH values and membrane potentials. We find that proton desolvation as well as water rotations are equally important for the proton transfer through gramicidin A at physiological membrane potentials. Our method allows to simulate long-range charge transfer in biological systems at time scales, which are not accessible by other methods.

Introduction

Long range proton transfer (LRPT) plays a major role in many biochemical processes.[1] Among them, biological energy transducing reactions such as cellular respiration, photosynthesis, and denitrification are of central importance for life. Although LRPT has been investigated extensively both experimentally and theoretically, the mechanism of these reactions is still not fully understood. One often discussed scenario is the so-called Grotthuss mechanism.[2,3] This mechanism assumes that the proton transfer reaction occurs in an already existing hydrogen bonded network. A subsequent rotation of the hydrogen bond partners restores the original network. In the Grotthuss mechanism, it is assumed that the rearrangement of the hydrogen bonded network is rate limiting for the LRPT. The actual transfer through the hydrogen bonded network is considered to be fast. Another proposed mechanism considers the energy barrier for transferring the proton through the hydrogen bonded network as rate limiting.[4] The rearrangement of the hydrogen bond pattern occurs during the LRPT and is thus not rate limiting.

To simulate LRPT in solution and in biological molecules, several approaches were developed. Many theoretical studies at different levels of approximation led to a detailed view of proton transfer reactions.[4−13] However, simulating the dynamics of LRPT processes in proteins still remains challenging. Two problems govern the simulation of LRPT processes. First, breaking of covalent bonds, which is typically addressed by quantum chemical methods, is necessary for proton transfer. Second, proton transfer processes across a cellular membrane occur on the microsecond time scale, which can not be simulated with current QM/MM methods.

The aim of the present work is to develop a general method for simulating LRPT in biomolecules. The approach that we are following is based on the master equation.[14,15] The elementary steps of the overall reaction are proton transfer and structural changes of the hydrogen bonded network. Since the number of possible states is rather large, we use a dynamical Monte Carlo (DMC) approach to solve the master equation.[16,17] In contrast to standard Metropolis Monte Carlo, DMC allows to simulate the kinetics of a reaction system.

We applied our DMC approach, to study the LRPT through gramicidin A (gA). This well-studied system consist of a head-to-head dimer of two helical peptides spanning the membrane.[18−20] The channel, which is formed in the center of the peptide, is filled by a file of water molecules.[4,21,22] Gramicidin A functions as an antibiotic exerting its activity by increasing the cation permeability of the target plasma membrane. Besides water and monovalent cations, also protons can pass the channel. While water molecules and cations diffuse through the channel, protons are transferred along a file of water molecules. This proton transfer across the membrane was measured experimentally in dependence on the pH value and the membrane potential.[23−26]

In this article, we describe a new DMC algorithm to simulate charge transfer in biomolecules. We discuss the theoretical

* Corresponding authors. E-mail: Matthias.Ullmann@uni-bayreuth.de (G.M.U.); Torsten.Becker@uni-bayreuth.de (T.B.). Fax: +49-921-55-3071.

10.1021/jp801477b CCC: $40.75 © 2008 American Chemical Society
Published on Web 09/30/2008

background and the implementation of the method. The method is applied to study the LRPT in gA for which we compare our results to experimental data. Due to the efficient Monte Carlo sampling, large molecular systems can be simulated over time ranges of biological interest. This approach will allow to investigate the underlying mechanism of biological charge transfer systems such as for example the photosynthetic reaction center, cytochrome *c* oxidase, and cytochrome bc_1.

Theory

Microstate Description. Biological charge transfer can be described as transitions between microstates of a system.[14,15,27−29] A microstate of a proton transfer system can be represented as an N-dimensional vector $\vec{x} = (x_1,..., x_i,..., x_N)$, where N is the number of protonatable sites of the system; x_i specifies the instance of site i, i.e., a combined representation of its protonation and rotameric form. Thus, assuming p possible instances x_i, there are in total $M = p^N$ possible microstates for the system. To keep the notation concise, microstates will be numbered by the Greek letters ν and μ, while we will use the roman letters i and j as site indices.

The standard energy for a given microstate $\vec{x}_\nu$ (i.e., the electrochemical potential of all ligands is zero) can be calculated by[30,31]

$$G_\nu^\circ = \sum_{i=1}^{N} (G_{\text{intr}}(x_i) + G_\Phi(x_i)) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} W(x_i, x_j) \qquad (1)$$

$G_{\text{intr}}(x_i)$ is the so-called intrinsic energy of the instance x_i, $G_\Phi(x_i)$ denotes the instance-specific energy contribution due to the membrane potential, and $W(x_i, x_j)$ takes into account the interactions between pairs of instances of different sites. If the electrochemical potential of the ligands is different from zero, the energy of the microstate differs from the standard energy. If we consider for simplicity that only protons can bind, the energy of the microstate ν at a given electrochemical potential $\bar{\mu}$ is given by

$$G_\nu = G_\nu^\circ - n_\nu\bar{\mu} \qquad (2)$$

where n_ν is the number of protons bound in microstate ν.

Equilibrium properties of a physical system are completely determined by the energies of its states. The equilibrium probability of a single state is given by

$$P_\nu^{\text{eq}} = \frac{e^{-\beta G_\nu}}{Z} \qquad (3)$$

with $\beta = 1/RT$ where R is the gas constant and T is the absolute temperature. Z is the partition function of the system.

$$Z = \sum_{\nu=1}^{M} e^{-\beta G_\nu} \qquad (4)$$

The sum runs over all M possible microstates. Macroscopic properties of the system can be obtained by summing up the individual contributions of all states. For example, the average number of bound protons is given by

$$\langle n \rangle = \sum_{\nu=1}^{M} n_\nu P_\nu^{\text{eq}} \qquad (5)$$

where n_ν denotes the number of bound protons in the microstate ν. For small systems, this sum can be evaluated explicitly. For larger systems, Monte Carlo techniques can be invoked to determine these probabilities.

Time Evolution of the System. The time evolution of the above-defined system can be described by a master equation

$$\frac{d}{dt} P_\nu(t) = \sum_{\mu=1}^{M} k_{\nu\mu}P_\mu(t) - \sum_{\mu=1}^{M} k_{\mu\nu}P_\nu(t) \qquad (6)$$

where $P_\nu(t)$ denotes the probability that the system is in state ν at time t, $k_{\nu\mu}$ denotes the probability per unit time that the system will change its state from μ to ν. The summation runs over all possible states μ. In principle, the time evolution of such a system can be solved analytically.[15] In the microstate description applied in this work, the number of states might become very large, so that solving eq 6 directly is computationally prohibited. To overcome this problem, stochastic methods, which have been developed to deal with complex kinetic systems, can be applied.[16,32,33] In such methods, the system—for example a chemical reaction system—is described by a discrete amount of particles of each species present. Transition rates are calculated for all possible reactions depending on the current number of particles. Although these stochastic methods are efficient in solving eq 6, they still require the calculation and the storage of all possible microstates and rate constants for all possible transitions. Such an approach would overstretch nowadays computational resources for a microstate description even of a biological molecule of moderate size.

In this paper, we introduce a DMC method which allows to solve eq 6 using affordable computational resources. The underlying idea is that although there is an overwhelming number of possible microstates, most of these states will never be populated, since they are energetically too unfavorable. However, deciding in advance, which microstates are important for the reaction dynamics of a system, could introduce a bias with consequences which are hard to estimate. To avoid this bias, we follow the time evolution of a single initial microstate and let our algorithm decide, which microstates will be populated in the course of the simulation. The time evolution of a given microstate is simulated by the Gillespie algorithm.[16] In order to get statistically significant results, the simulations need to be repeated several times. We call this variant of the DMC method sequential DMC. For a small test system with five sites,[15] we test the correctness of the implementation of our sequential DMC algorithm by comparing the analytically obtained kinetics with those calculated by the sequential DMC method (data not shown).

Figure 1 shows a flowchart of our sequential DMC algorithm which is based on the Gillespie algorithm. Starting from an initial microstate, rate constants are calculated for all events possible. An event is a transition between microstates. In our simulation, only one elementary step (proton uptake, proton release, proton transfers through a hydrogen bond, or rotation of a water molecule) is allowed in one event. The number of possible events for a given microstate is typically small and maximally on the order of N^2p, where N is the number of sites and p the number of instances per site. Thus, the total number of all possible events in the system (which is maximally in the order of p^{2N}) is drastically reduced. Given the rate constants of the possible events starting from the given microstate, the algorithm chooses the next event m according to the following criterion [16,17]

$$\sum_{l=1}^{m-1} k_l \leq \rho_1 K < \sum_{l=1}^{m} k_l \tag{7}$$

$$K = \sum_{l=1}^{L} k_l \tag{8}$$

K is the sum of the rate constants k_l of all L possible events for the given microstate; ρ_1 is a random number between 0 and 1. The rate constant k_l is equivalent to one of the rate constants in eq 6 and is a measure of the probability that event l happens during the next time step. To adequately represent the kinetic behavior of the system, it has to be ensured that the events are chosen in accordance with their respective probability. Thus, if a rate constant k_r is twice as large as a rate constant k_s, event r should on average be chosen twice as often as event s. This behavior is facilitated by eq 7. In the given example, k_r contributes twice as much as k_s to the sum K and, thus, the probability that event r fulfills eq 7 is twice as large as that for event s.

The time Δt that elapsed during the Monte Carlo step is given by

$$\Delta t = \frac{1}{K} \ln\left[\frac{1}{\rho_2}\right] \tag{9}$$

which is a standard way to draw a random number Δt from an exponential distribution given a uniformly distributed random number ρ_2 between 0 and 1. Thus, eq 9 is equivalent to the statement that the probability of any event to happen within time Δt is given by $\exp(-K\Delta t)$. In summary, the criteria in eqs 7 and 9 ensure that.

(i) all events are chosen according to their respective probability and

(ii) the average time evolution of the system follows a master equation.

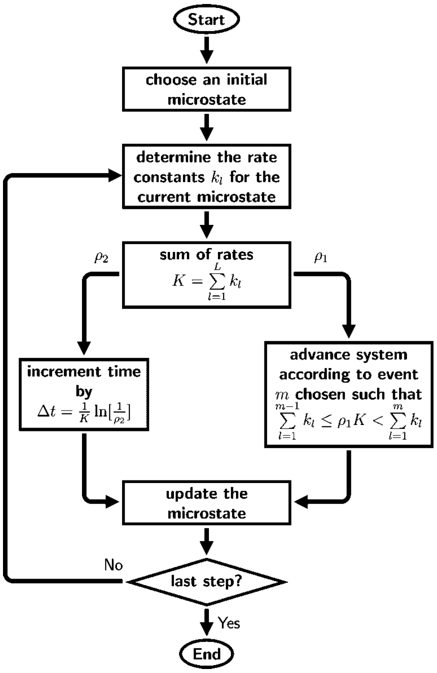

Start
choose an initial microstate
determine the rate constants k_l for the current microstate
sum of rates $K = \sum_{l=1}^{L} k_l$
ρ_2
ρ_1
increment time by $\Delta t = \frac{1}{K}\ln[\frac{1}{\rho_2}]$
advance system according to event m chosen such that $\sum_{l=1}^{m-1} k_l \leq \rho_1 K < \sum_{l=1}^{m} k_l$
update the microstate
last step?
No
Yes
End

Figure 1. Flowchart of the sequential DMC algorithm. Starting from a microstate, rate constants for all possible events are calculated. The time increment and the reaction to take place are chosen based on the calculated rate constants and two random numbers (ρ_1 and ρ_2) between 0 and 1.

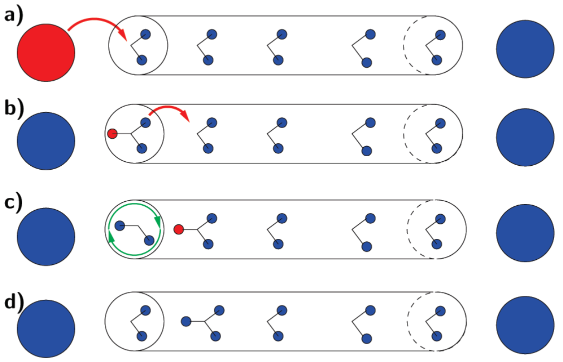


Figure 2. Different kinds of possible reactions included in the sequential DMC algorithm. (a) Representation of the uptake of a proton. (b) Proton transfer between neighboring water molecules. (c) Rotation of the first water molecule. (d) State of the channel after all three elementary reaction steps.

Application of the described algorithm provides a trajectory, i.e., a succession of microstates with accompanying time information. Comparison to experimental data can be made by averaging over several trajectories:

$$\langle B \rangle = \frac{1}{N_{\mathrm{Tr}}} \sum_{l=1}^{N_{\mathrm{Tr}}} B_l \tag{10}$$

where $\langle B \rangle$ is any given measurable quantity and B_l is its value for a given trajectory l, N_{Tr} is the number of trajectories. The flux F of protons through the channel, for example, is calculated as follows:

$$\langle F \rangle = \frac{1}{N_{\mathrm{Tr}}} \sum_{l=1}^{N_{\mathrm{Tr}}} F_l = \frac{1}{N_{\mathrm{Tr}}} \sum_{l=1}^{N_{\mathrm{Tr}}} \frac{f_l}{t_{\mathrm{Tr}}} \tag{11}$$

where t_{Tr} is the time elapsed in one trajectory and f_l is the number of the protons that are transferred from ectoplasm to cytoplasm in trajectory l.

Description of the Model System. The dimeric proton channel gA was chosen as a model system to test the DMC approach. The proton transfer through this channel occurs along a file of water molecules. In our simulation, the water molecules can rotate and protonate. Proton transfer can only occur between neighboring water molecules. Proton uptake and release takes place only at the water molecules at the two ends of the channel. The water molecules can assume different orientations: four for the protonated water molecule (H_3O^+) and six for the neutral water molecule (H_2O, see the section Water Representation in Computational Details). A rotation is the transition between different orientations of a water molecule; the protonation is not allowed to change during a rotation. Since our system contains eleven water molecules that can exist in ten different instances, the total number of different microstates is 10^{11}. In the simulation, only one elementary step (proton uptake, proton release, proton transfer, or rotation of a water molecule) is allowed in one Monte Carlo step. The model system and the possible reactions are schematically depicted in Figure 2.

Calculation of the Rate Constants. The rate constant $k_{\nu\mu}$ of the transition from state μ to state ν is calculated using an Arrhenius approach

$$k_{\nu\mu} = A_{\nu\mu} e^{-\beta G^{\ddagger}_{\nu\mu}} \tag{12}$$

The preexponential factor $A_{\nu\mu}$ was set to 10^{13} s^{-1}, which approximates the preexponential factor kT/h derived from

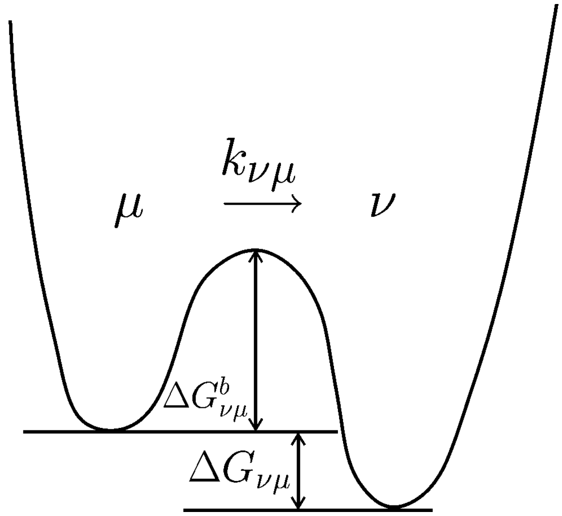


Figure 3. Energy profile of a reaction (proton uptake, proton release, proton transfer, or rotation of a water molecule) within our system. μ and ν are the microstates, $k_{\nu\mu}$ is the reaction rate constant for the reaction from μ to ν. $\Delta G^{\mathrm{b}}_{\nu\mu}$ is the energy barrier, and $\Delta G_{\nu\mu}$ is the difference between the microstate energies of ν and μ.

transition state theory, where k is the Boltzmann constant, T is the temperature and h is Planck's constant. This preexponential factor $A_{\nu\mu}$ represents the maximal rate corresponding to an activationless transition. The activation energy $G^{\ddagger}_{\nu\mu}$ is given by

$$G^{\ddagger}_{\nu\mu} = \begin{cases} \Delta G_{\nu\mu} + \Delta G^{\mathrm{b}}_{\nu\mu} & : \quad \Delta G_{\nu\mu} > 0 \\ \Delta G^{\mathrm{b}}_{\nu\mu} & : \quad \Delta G_{\nu\mu} \leq 0 \end{cases} \tag{13}$$

$\Delta G_{\nu\mu}$ is the energy difference between the microstates μ and ν. $\Delta G^{\mathrm{b}}_{\nu\mu}$ is the energy barrier between the microstates μ and ν. The meaning of the symbols is illustrated in Figure 3. The way of obtaining energy barriers for the elementary reactions of our system is described in the following.

Proton Transfer Along a Hydrogen Bond. Proton transfer can only occur between a hydronium ion and a water molecule that form a hydrogen bond. Which pairs of molecules form a hydrogen bond can be determined based on geometric criteria: the O−O distance between these water molecules is less than 4 Â and the hydrogen atom of the donor molecule points toward the lone pair of the acceptor molecule. An angle criteria for an hydrogen bond is derived from the regular tetrahedron structure of the water molecules. Only hydrogen bonds with an hydrogen bond angle that deviates from 180° by less than 55° are considered. The energy difference between the reactant state and the product state is calculated from eq 1. The energy barrier for a proton transfer along a hydrogen bond in water is rather small.[10,34,35] Therefore, we set the energy barrier $G^{\mathrm{b}}_{\nu\mu}$ for the proton transfer reaction to a fixed value of 0.5 kcal/mol in agreement with quantum chemical calculations.[10,34,35] With an average proton transfer rate constant of 3 ps^{-1} (taken from a simulation without membrane potential), we can estimate a transfer time of about 330 fs from our calculations which is in the same order of magnitude as proton transfer times determined from simulations of proton transfer in water.[36,37] The two calculations should result in comparable proton transfer rates, since the environment within the gA channel is similar to that in bulk water phase. In both cases, a water molecule forms several hydrogen bonds. In the gA channel, hydrogen bonds are formed with waters and the peptide backbone.

Proton Uptake and Release. The rate of proton uptake and release depends on the proton electrochemical potential $\bar{\mu}$ of the surrounding medium.

$$\bar{\mu} = -RT\ln(10)\mathrm{pH} + zF\phi \tag{14}$$

where R is the gas constant, T the absolute temperature, z is the charge of a proton, F is Faraday's constant, and ϕ is the membrane potential. The energy difference $\Delta G_{\nu\mu}$ between the product state ν and the reactant state μ is given by

$$\Delta G_{\nu\mu} = \Delta G^{\circ}_{\nu\mu} - \Delta G^{\circ}_{\mathrm{H_2O}} - \lambda\bar{\mu} \tag{15}$$

where λ is −1 for proton release reactions and +1 for proton uptake reactions. $\Delta G^{\circ}_{\mathrm{H_2O}}$ is the energy for protonating a water molecule in the bulk at standard conditions, which takes into account that the proton is taken up from or released to the bulk water. This value can be calculated from the pK_a value for the protonation of a water molecule and is 2.3 kcal/mol.

The energy barrier $\Delta G^{\mathrm{b}}_{\nu\mu}$ for taking up a proton from the bulk water into the gA channel has two contributions. First, the energy barrier for transferring a proton in bulk water, which is at least 1.9 kcal/mol[3]. Second, the transfer of a proton from the bulk to the surface of the membrane, which was estimated to be about 2.7 kcal/mol.[38,39] These two contributions lead to a value of at least 4.6 kcal/mol for the energy barrier of the proton uptake and release, which is the value used in this study.

Rate Constants for Rotations of Molecules. The barrier of the rotation of a water or a hydronium molecule is assumed to depend on the number of hydrogen bonds that need to be broken to allow this rotation, no matter if these hydrogen bonds are formed again after the rotation. Hydrogen bonds are defined as explained above. The energy for breaking the hydrogen bonds determines the energy barrier $G^{\mathrm{b}}_{\nu\mu}$. To calculate the energy for breaking a hydrogen bond, we apply an empirical formula (eq 16).[40] The energy barrier $G^{\mathrm{b}}_{\nu\mu}$ is given by summing over the contribution of all H hydrogen bonds that need to be broken,

$$G^{\mathrm{b}}_{\nu\mu} = \sum_{l=1}^{H} a\mathrm{e}^{-cr_l} \tag{16}$$

where r_l is the O•••H distance; a and c are empirical constants which have the values 6042 kcal/mol and 3.6 Å^{-1}, respectively. Equation 16 leads to hydrogen bond energies between 4.5 and 0.5 kcal/mol for H•••O distances between 2 and 5 Å, respectively. The energy difference between the reactant state and the product state are again calculated from eq 1. In order to avoid barrierless rotation events, the minimum barrier is set to 1.0 kcal/mol. Woutersen et al.[41] measured the rotation rate of water molecules in bulk water by IR-spectroscopy. These authors found two rotation times for water molecules, 0.7 ps for weakly, and 13 ps for strongly hydrogen bonded water molecules. Since in our system water molecules have less hydrogen bonds than in liquid water, a rotation time of 1.1 ps, which we obtained from our simulations, is in good agreement with the experimental data.

Computational Details

Structure Preparation. Coordinates of gA are taken from the PDB (code 1jno).[42] A cube of dummy atoms (20 × 20 × 20 Å^3) with zero charge is placed around the structure to represent the lipid bilayer. Since the structure is determined by NMR, no positions for water molecules are available in the structure. To generate water positions, the system is placed in a water box. All water molecules overlapping with the system are deleted. A short steepest descent energy minimization (1000 steps)

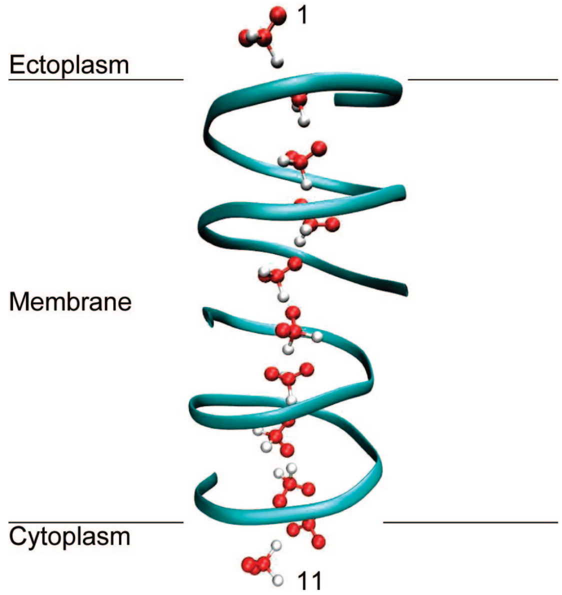


Figure 4. Gramicidin A system used in the simulation. The system contains eleven water molecules buried inside the gramicidin A membrane channel. The water model is depicted with the oxygen atom at the center and two lone pairs (red) and two hydrogen atoms (white).

followed by an adopted basis Newton−Raphson minimization (10 000 steps) is done using CHARMM.[43] Peptide heavy atoms and membrane atoms are kept fixed for both minimizations. In agreement with previous simulations,[4,21,22] we found nine water molecules in the channel. Two additional water molecules, one on each side of the channel, are selected to connect the water file within gA to the bulk solvent. These water molecules are in contact with the water molecules in the channel. The total number of water molecules thus amounts to eleven. Finally, the surrounding water box is removed and the eleven water molecules are replaced by our five-center water model (see next section). The resulting structure (see Figure 4) is used in all electrostatic calculations.

Water Representation. The incorporation of rotation events in our simulations requires an efficient way of calculating the contributions of the different rotameric forms of a water molecule to the microstate energy. For this purpose, we designed a symmetric water model based on a regular tetrahedron with five interaction centers, one at the center of the tetrahedron and the remaining four at each corner of the tetrahedron. The distance between the central and the four peripheral interaction centers is 0.95. The central interaction center represents the oxygen atom and the peripheral interaction centers represent either lone pairs or hydrogen atoms. The peripheral centers are permutated to sample all possible rotameric forms. No coordinates need to be changed, only atom labels and charges are assigned to already existing interaction centers. This water representation makes the calculation of state energies (eq 1) very efficient. Multipole-derived charges[44] for the possible protonation forms (H_2O and H_3O^+) are calculated using ADF.[45] For the H_2O molecule, the oxygen atom, the hydrogen atoms, and the lone pairs have a charge of −0.22, 0.21, and −0.10, respectively. For the H_3O^+ molecule, the respective atoms have a charge of 0.13, 0.32, and −0.09. Zundel ions were not considered explicitly, but geometries that correspond to Zundel ions where included in the simulation.

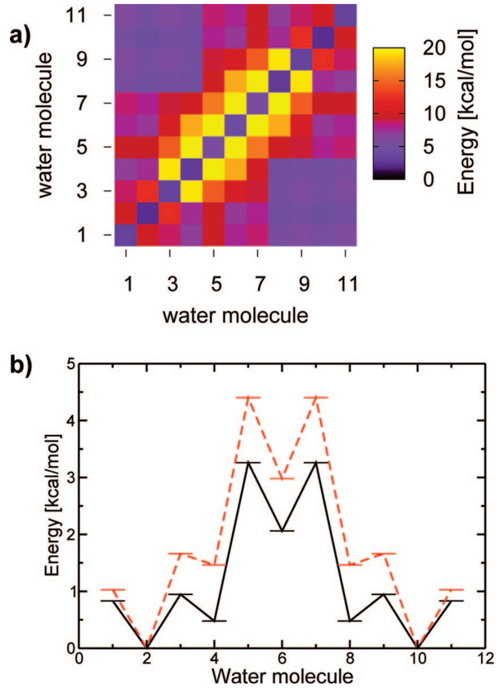


Figure 5. (a) Two dimensional potential of mean force for binding protons to the gA channel without membrane potential. The diagonal represents states with one proton bound. All other squares represent states with two protons bound, i.e., the entry (1,5) represents the state in which one proton is bound to water molecule 1 and the other proton to water molecule 5. The plot is symmetric, because entry (1,5) and entry (5,1) represent the same physical situation. (b) Energy profiles for the gA channel with exactly one proton inside the channel. The solid line depicts the potential of mean force. The dashed line is the minimum energy profile. For better comparison, both profiles are shifted with their minimum value to 0 kcal/mol.

Electrostatic Calculations. The energetic parameters in eq 1 ($G_{\mathrm{intr}}(x_i)$, $G_\Phi(x_i)$, $W(x_i, x_j)$) are calculated from the solution of the Poisson−Boltzmann equation.[30,31] The intrinsic energies $G_{\mathrm{intr}}(x_i)$ and the interaction energies $W(x_i, x_j)$ are obtained by using the MEAD package.[46] The dielectric constant for the protein and the membrane is set to 4 and the dielectric constant of the solvent is set to 80. The ionic strength is set to 0.1 M. The electrostatic potential is calculated by focusing using two grids of 81^3 grid points and a grid spacing of 1.0 and 0.25 Å. The first grid is centered on gA, and the second grid, on the water molecule of interest. Partial charges for the water molecules are taken from the ADF calculations as described before, partial charges for the peptide are taken from the CHARMM force field.[47] Energy contributions due to the membrane potential $G_\Phi(x_i)$[31] are calculated by the PBEQ module[48,49] of CHARMM[43] using the same settings as for the MEAD calculations. In order to account for the symmetry of gA, we symmetrized the energetic parameters in eq 1, i.e., we assigned the same energy parameters ($G_{\mathrm{intr}}(x_i)$, $W(x_i, x_j)$) to symmetry related water molecules.

DMC Calculations. The time evolution of the system is simulated by calculating possible transitions between the microstates. A microstate is described by a vector with eleven elements, each element represents one water molecule. Water molecules 1 and 11 are connected to the ectoplasm and cytoplasm, respectively. All other water molecules are connected only to their neighboring water molecules.

TABLE 1: Comparison of the Experimentally and Computationally (DMC) Determined Proton Flux through the gA Channel for Different pH Values and Membrane Potentials[a]

	Proton Flux [pA]							
	pH 0		pH 0.3		pH 2		pH 2.3	
membrane potential [mV]	exp[b]	DMC	exp[c]	DMC	exp[d]	DMC	exp[e]	DMC
0	2	0 ± 1	0	1 ± 1	0.4	0.0 ± 0.2	−0.05	0.0 ± 0.1
−50	22	3 ± 1	8	4 ± 1	0.7	0.1 ± 0.2	0.42	0.2 ± 0.2
−100	42	7 ± 1	21	8 ± 2	0.9	0.6 ± 0.5	0.71	0.3 ± 0.2
−150	63	14 ± 2	34	15 ± 2	1.1	1.3 ± 0.5	*f*	0.6 ± 0.4
−200	79	23 ± 2	*f*	25 ± 3	1.1	1.9 ± 0.9	*f*	1.1 ± 0.6
−300	105	50 ± 3	*f*	47 ± 3	*f*	4.1 ± 1.1	*f*	2.3 ± 0.7
−400	120	76 ± 5	*f*	69 ± 5	*f*	8.3 ± 1.4	*f*	5.7 ± 1.6
−500	*f*	110 ± 7	*f*	97 ± 4	*f*	13.3 ± 1.3	*f*	7.8 ± 1.4

[a] Experimental data were obtained from published diagrams using the program g3data. Error values given for the DMC calculations are calculated as standard deviations of twenty independently simluated trajectories. [b] Reference 23. [c] Reference 24. [d] Reference 25. [e] Reference 26. [f] Not determined.

For each DMC simulation, 20 trajectories are generated. Since the initial state is set arbitrarily, the system is simulated for 10 000 Monte Carlo steps so that the system can adopt a steady state. The final state of this short simulation is than used as starting configuration of a production run of 5×10^6 Monte Carlo steps. Properties are calculated as average over these 20 trajectories.

Results and Discussion

In this article, we combine a microstate formalism that describes charge transfer reactions[14,15] and a sequential DMC algorithm to simulate the kinetics of long-range proton transfer processes. Energetic parameters of this reaction system are obtained from continuum electrostatic calculations. We present simulations of the proton transfer through gramicidin A (gA) in dependence on external pH and membrane potential. The proton flux obtained by these simulations agrees with experimental values.[23−26]

Energy Profile of the Proton Channel. In order to analyze the proton transfer process within the channel, it is instructive to calculate first the energy profile along the proton transfer path. For this purpose, we computed the potential of mean force (PMF) of the gramicidin A channel for having one or two protons inside the channel. Since in our system only microstates with one or two protons in the channel are significantly populated, all relevant states are considered in the two-dimensional energy profile shown in Figure 5a. Due to the moderate size of our system, the partition function of the system with a limited number of protons in the channel can be calculated and thus the PMF can be obtained from the following two equations (given here for one proton in the channel):

$$\langle u_i \rangle = \frac{\sum_{\nu=1}^{M} u_i \cdot \mathrm{e}^{-G_\nu/RT}}{Z} \qquad (17)$$

$$G_{\mathrm{pmf}} = RT \ln\left(\frac{\langle u_i \rangle}{1 - \langle u_i \rangle}\right) \qquad (18)$$

$\langle u_i \rangle$ is the probability that the proton is on site i, u_i, is 1 or 0 depending on whether site i is protonated or deprotonated, respectively. Z is the partition function, R is the ideal gas constant, and T is the absolute temperature. The one-dimensional PMF obtained from eq 18 is plotted in Figure 5b (solid line). For a system with one proton in the channel, the minimum energy profile is shown as dashed line. An energy barrier for the charge transfer from one side of the channel to the other is located at the central three water molecules. This energy barrier is about 3.4 kcal/mol if calculated from the PMF profile and 4.6 kcal/mol if calculated from the minimum energy profile. The difference between the minimum energy profile and the PMF profile are entropic contributions due to water molecule rotation, which are taken into account in the PMF profile but not in the minimum energy profile. These entropic contributions lower the energy barrier by about 1.2 kcal/mol. For a one barrier process, such a lowering corresponds to an increase of the overall rate constant by about 1 order of magnitude, which underlines the importance of water rotations in the gA channel. The energy barrier obtained from the PMF is in good agreement with an earlier empirical valence bond calculation.[4]

From the two-dimensional PMF in Figure 5a, one can derive the localization of the protons inside the channel. The lowest energy states are found on the diagonal. This diagonal represents the states with one proton bound. All other states have two protons bound. Low energy states with two protons bound are those in which the protons are on opposite sides of the barrier, i.e., one proton is on water molecule 1 to 4 and the other one is on water molecule 8 to 11.

Proton Flux Through Gramicidin A. In order to compare our sequential DMC calculations with experimental data, the system was simulated at pH values of 0.0, 0.3, 2.0, and 2.3. For each pH value, the proton flux was calculated for membrane potentials ranging from 0 to −500 mV. A trajectory of 8 μs took about 5 min on an Intel Pentium 4 with 3.2 GHz.

Table 1 shows a comparison of the calculated proton flux with experimental data of several groups.[23−26] We obtained an agreement between theory and experiment within 1 order of magnitude with a slight trend of underestimating the proton flux. The calculated proton flux deviates from the experimental value normally only by a factor of 2. We are not only able to reproduce the dependence of the experimental fluxes on the membrane potential at a given pH value, but our simulations also reproduce the increase of the proton flux when the pH is lowered from 2.3 to 0.0. Especially at pH = 0, the discrepancy between theory and experiments is larger. Under these conditions, the model is also expected to describe the real system less satisfactorily, since the pH contributes considerably to the ionic strength at this pH, which is not considered in our calculations. Since in our calculations no other parameters than the pH value and the membrane potential are changed, our model describes correctly the behavior of LRPT in gA over a wide range of pH values and membrane potentials.

TABLE 2: Probability for a Proton Being Transferred from the Ectoplasm to the Cytoplasm[a]

	Proton Transfer Probability [%]					
	number of protons not limited			number of protons limited to one		
water molecule	−100 mV	−300 mV	−500 mV	−100 mV	−300 mV	−500 mV
1	9	37	53	16	62	88
2	9	37	54	16	62	88
3	16	66	82	16	63	88
4	25	83	95	27	75	92
5	37	95	99	52	94	99
6	44	95	99	55	95	99
7	68	97	99	70	97	99
8	92	99	100	90	99	100
9	98	99	100	97	99	100
10	99	99	100	98	99	100
11	99	100	100	99	100	100

[a] All values are calculated at pH = 0. A transfer probability of 68% for water molecule 7 at −100 mV means that 68% of the protons which reached water molecule 7 were transferred across the whole channel afterwards. Columns 2−4 present transfer probabilities of protons if the simulation is not limited to a certain number of protons. Columns 5−7 show the transfer probabilities if the number of protons inside the channel is limited to one.

Proton Transfer Mechanism. The good agreement of the proton flux with experimental values allows to further investigate the mechanism of the LRPT in gA. In the first subsection, we analyze the overall behavior of protons in the gA channel by determining the transfer probability of individual protons. Our sequential DMC approach enables us to follow single protons in the gA channel. Such an analysis is shown in the second subsection. In the last subsection, we will address the question whether the reorientation of the hydrogen bonded network or the electrostatic barrier for the charge transfer is rate limiting for the LRPT process in gA.

Proton Transfer Probability. Table 2 presents the probability that a proton is transferred through the whole channel after reaching a given water molecule. For example, a probability of 68% for water molecule 7 at a membrane potential of −100 mV means that 68% of the protons which reached water molecule 7 after being taken up from the ectoplasmic site are released on the cytoplasmic site. The remaining 32% of protons are returned to the ectoplasmic site. At low membrane potentials (−100 mV) only 9% of the protons that are taken up are transferred across the membrane (see Table 2). The remaining 91% of the protons are released on the same side of the membrane where they entered the channel. But already at this membrane potential, it is obvious from the transfer probabilities that once the proton crosses the central energy barrier, it is most likely transferred across the whole channel. At a membrane potential of −100 mV, once the proton at water molecule 1 has reached water molecule 8, the probability of leaving the channel from water molecule 11 is above 90%. With increasing membrane potential, the proton transfer probabilities increase for all water molecules. Interestingly, for membrane potentials as negative as −300 mV, the first water molecule for which more than 50% of the protons are transferred, is located in front of the energy barrier. Under these conditions, the energy barrier is already greatly diminished, nevertheless a small barrier remains. However, proton transfer through gA is a nonequilibrium process. Thus, discussing transfer probabilities solely on the basis of an energy profile can be misleading.

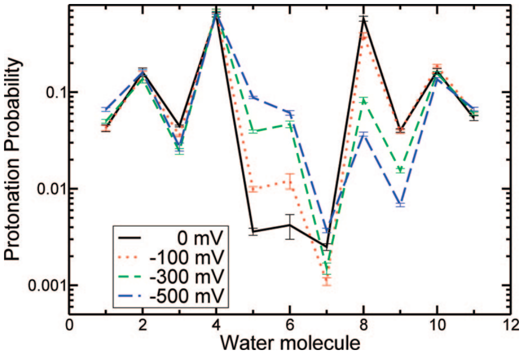


Figure 6. Protonation probability under steady state conditions at membrane potentials of 0, −100, −300, and −500 mV. The overall number of protons in the channel decreases with increasing membrane potential.

Figure 6 shows the protonation probability at pH = 0 of all water molecules at equilibrium conditions (without membrane potential) and at steady state conditions with different membrane potentials. The protonation probabilities of the three central water molecules are very much influenced by the membrane potential. Figure 6 shows that with increasing membrane potential, the protonation probabilities of the water molecules 5 and 6 increases strongly. In contrast, the protonation probabilities of water molecules 8 and 9 decrease with increasing membrane potential. This decrease also leads to an overall reduction of the average number of protons inside the channel from 1.7 (at a membrane potential of 0 mV) to 1.3 (at a membrane potential of −500 mV). The observed shifts of the protonation probabilities under the influence of the membrane potential can be interpreted as follows: Under the influence of the membrane potential, the proton reaches the barrier at the central three water molecules more frequently. Once the barrier is crossed, the proton tends to leave the channel more rapidly the stronger the membrane potential. Nevertheless, the energy profile that could be extracted from the protonation probabilities does not simply contain the membrane potential as an additive contribution, because the protons that crossed the barrier are eventually removed from the channel and thus the steady state protonation differs from an equilibrium protonation.

Analysis of Single Protons in the Channel. Our sequential DMC approach allows to analyze the simulation in analogy to single molecule experiments. We can for instance follow single protons inside the channel. Figure 7 shows such an analysis at different membrane potentials. At a membrane potential of −100 mV (Figure 7a), protons that entered the channel stay on the same site of the channel and are only rarely transferred across the central barrier. The protons enter from both sites and only reach water molecule 4 or 8, depending on the site from which they have entered the channel. This observation correlates with a very low transfer probability of only about 9% for −100 mV. At a membrane potential of −300 and −500 mV, more protons are transferred across the membrane. The actual crossing event is rather fast for all membrane potentials, reflecting the high protonation energies of the central three water molecules. The increase of the number of transferred protons is due to a decrease of the time span a proton stays in front of the barrier. Moreover, the protons also leave the channel faster. If the proton has crossed the barrier, it is generally released. The probability that the proton is crossing the barrier again in the opposite direction is negligibly small. Table 3 shows the average occupation times, i.e., the average time a single proton stays on a water molecule

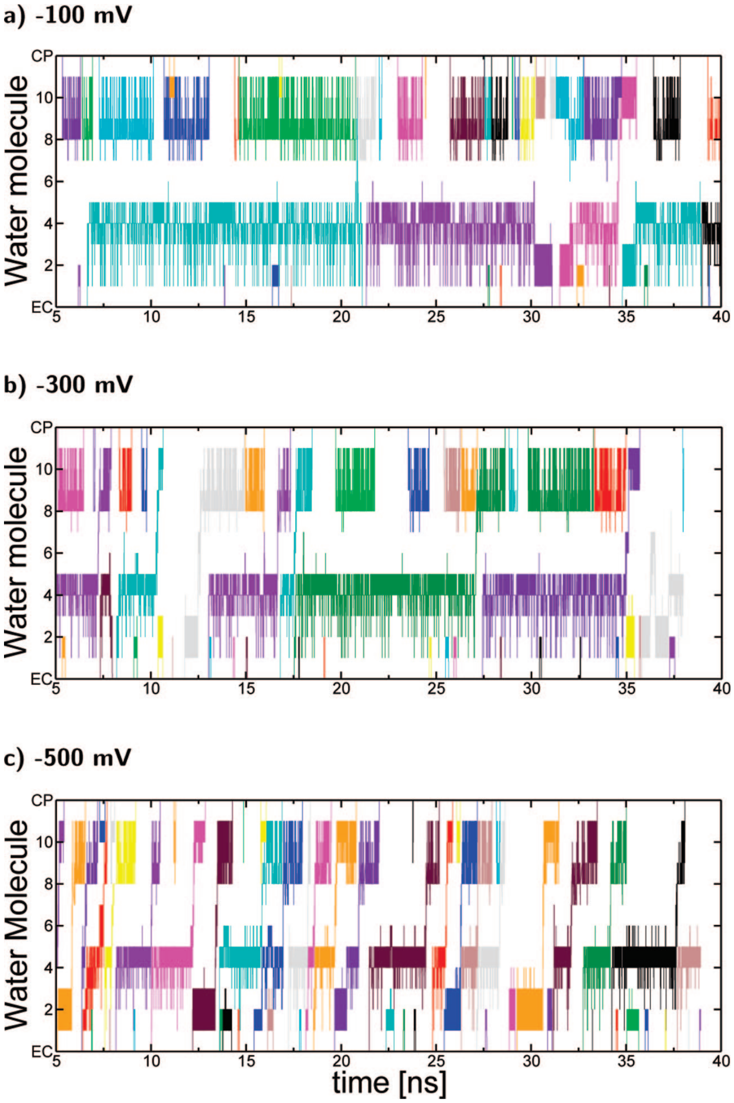


Figure 7. Sections of sequential DMC trajectories, which allow us to follow protons through the gA channel. The diagrams show the location of protons within the first 40 ns of our simulations for membrane potentials of (a) −100, (b) −300, and (c) −500 mV. The colors represent different protons. All simulations are performed at pH = 0.

while it is inside the channel. The proton stays most of the time at water molecules 4 and 8. The actual transfer over the barrier is fast.

As can be seen from Figure 7, there is typically more than one proton inside the channel, i.e., once a proton has crossed the central barrier, the next proton already enters the channel. One thus may ask whether the transfers of these protons are correlated with each other. Therefore, in a second set of simulations, the number of protons inside the channel is limited to one. In these simulations, uptake events are only allowed if no proton is inside the channel. If there is a correlation between two protons inside the channel, the limitation to one proton should result in different transfer probabilities and occupation times. Table 2 shows the transfer probabilities for each water molecule for simulations with an arbitrary number of protons inside the channel and simulations limited to one proton inside the channel. In Table 3, the occupation times of a proton are listed for each water molecule determined from these two sets of simulations. The transfer probabilities shown in Table 2 for the simulations with a limited number of protons increase for the first two water molecules. The changed transfer probabilities indicate that the first proton hinders the second proton from crossing the barrier. This hindrance is due to electrostatic interactions as can be seen from the two-dimensional potential of mean force (Figure 5a). Microstates with two protons close to each other (less than three water molecules distance) have a significantly higher energy than microstates with protons separated by more than three water molecules.

The same picture can be obtained by comparing the occupation times in the simulations with a limited number of protons to the occupation times in the simulations without this limitation. The proton stays much longer on water molecule 4 if there is more than one proton allowed in the channel, but the occupation time for water molecule 8 is similar or even decreased. The first proton, which is on the other side of the barrier, hinders

TABLE 3: Average Occupation Times for All Water Molecules[a]

	Average Occupation Times [ps]					
	number of protons limited to one			number of protons not limited		
water molecules	−100 mV	−300 mV	−500 mV	−100 mV	−300 mV	−500 mV
1	93 ± 7	51 ± 3	23 ± 6	90 ± 4	62 ± 2	51 ± 2
2	491 ± 37	335 ± 24	194 ± 52	316 ± 22	172 ± 14	124 ± 8
3	59 ± 5	62 ± 4	53 ± 12	127 ± 16	53 ± 3	32 ± 3
4	1313 ± 261	1348 ± 132	900 ± 206	3973 ± 545	1953 ± 164	905 ± 49
5	64 ± 12	87 ± 8	99 ± 21	80 ± 9	122 ± 8	125 ± 5
6	109 ± 20	136 ± 9	91 ± 11	103 ± 18	148 ± 9	87 ± 4
7	14 ± 3	5 ± 0	4 ± 1	3 ± 0	4 ± 0	5 ± 0
8	935 ± 225	318 ± 27	76 ± 12	1011 ± 61	143 ± 8	46 ± 2
9	27 ± 5	26 ± 2	10 ± 2	77 ± 6	24 ± 1	8 ± 0
10	325 ± 30	242 ± 15	178 ± 24	330 ± 15	240 ± 10	163 ± 5
11	105 ± 9	91 ± 5	81 ± 11	106 ± 4	91 ± 4	80 ± 3

[a] An occupation time is the total time a single proton stays on a water molecule while it is inside the channel. Columns 2−4 present transfer probabilities of protons if the simulation is not limited to a certain number of protons. Columns 5−7 show the transfer probabilities if the number of protons inside the channel is limited to one. Error values given for the DMC calculations are calculated as standard deviations of 20 independently simluated trajectories.

TABLE 4: Proton Flux through the gA Channel in Dependence on the Membrane Potential for Differently Lowered Electrostatic Barriers and Increased Rotation Rates[a]

	Proton Flux [pA]						
		lowering of the electrostatic energy barrier[b]			factor for rotation rate increase[c]		
membrane potential [mV]	reference flux	−1.0	−2.0	−3.0	5	10	100
0	0 ± 1	2 ± 2	7 ± 3	10 ± 2	2 ± 4	3 ± 5	4 ± 8
−50	3 ± 1	8 ± 2	35 ± 3	42 ± 3	9 ± 3	14 ± 8	30 ± 11
−100	7 ± 1	19 ± 2	62 ± 5	73 ± 3	23 ± 5	30 ± 9	71 ± 15
−150	14 ± 2	32 ± 2	84 ± 4	101 ± 4	41 ± 7	56 ± 13	128 ± 19
−200	23 ± 2	46 ± 4	106 ± 5	131 ± 4	70 ± 9	96 ± 14	184 ± 14
−300	50 ± 3	87 ± 4	146 ± 5	186 ± 4	136 ± 8	177 ± 20	332 ± 24
−400	76 ± 5	129 ± 5	189 ± 4	245 ± 4	208 ± 13	276 ± 21	488 ± 19
−500	110 ± 7	181 ± 5	236 ± 7	316 ± 5	298 ± 12	369 ± 25	625 ± 26

[a] All values are derived at pH = 0. Error values given for the DMC calculations are calculated as standard deviations of twenty independently simluated trajectories. [b] Intrinsic energies of the three central water molecules was lowered by 1, 2, and 3 kcal/mol. [c] The rotation rate is increased by multiplying $A_{\nu\mu}$ in eq 11 by 5, 10, and 100.

the second proton from crossing. The second proton has to stay in front of the barrier until the first proton has left the channel.

These findings indicate a strong correlation between the protons transferred through the gA channel at low pH when on average more than one proton is in the channel. At higher pH values, less protons are in the channel. At pH = 2.3 for instance, on average only 0.1 protons are found in the channel. Under these circumstances, the proton−proton interaction has nearly no influence on the proton flux.

Rate Limiting Step of the LRPT. The rate limiting step of the LRPT in gA is under ongoing discussion.[4,50,51] Two different aspects of the transfer process might be rate limiting. On one hand, protons have to overcome an electrostatic energy barrier to cross the channel;[4] on the other hand, the hydrogen bonded network has to rearrange to allow the next transfer. In order to address the question which aspect is rate limiting, we artificially reduce the electrostatic energy barrier of the LRPT process in gA as well as increase the rotation rates in our simulations. For comparison, it is instructive to describe the LRPT as an one barrier process. Assuming Arrhenius behavior, a decrease of the energy barrier by 1.0 kcal/mol increases the transfer rate by a factor of 5, a decrease by 1.35 kcal/mol increases the transfer rate by a factor of 10 and a decrease by 2.7 kcal/mol increases the rate by a factor of 100. In our simulations, the energy barrier is reduced by lowering the intrinsic energies ($G_{\text{intr}}(x_i)$ in eq 1) of the protonated forms of the central three water molecules by 1, 2, or 3 kcal/mol. The rotation rates are increased by multiplying the preexponential factor ($A_{\nu\mu}$ in eq 11) by 5, 10, or 100. If the electrostatic energy barrier is the rate limiting step of the LRPT, the reduction of this barrier should result in a higher proton flux through the gA channel. If the rearrangement of the hydrogen bonded network is rate limiting, the increased rotation rate should increase the proton flux. As can be seen from Table 4, lowering the electrostatic barrier has a significant effect on the observed proton flux. At a membrane potential of −100 mV, decreasing the barrier by 3 kcal/mol increases the flux about 10-fold. And even at membrane potentials as negative as −500 mV, we still observe an increase of the flux from 110 to 316 pA.

Increasing rotation rates also increases the flux. For membrane potentials between 0 and −200 mV, the influence of increasing the rotation rates on the observed flux is similar to the influence of lowering the electrostatic barrier. For membrane potentials more negative than −200 mV, increasing the rotation rates is even more effective than lowering the electrostatic barrier. At a membrane potential of −500 mV, an increase of the rotation rates by a factor of 100 increases the flux from 110 to 625 pA.

The increase of the proton flux by lowering the electrostatic energy barrier is expected to attenuate at higher membrane potentials, since stronger membrane potentials diminish the influence of the electrostatic energy barrier on the LRPT process. The increase of the proton flux by increasing the rotation rates is not influenced by stronger membrane potentials. At small membrane potentials between 0 and −200 mV both the electrostatic energy barrier and the rotations of the participating molecules similarly influence the LRPT in gA. An increased flux can be achieved both by lowering the electrostatic energy barrier as well as by increasing the rotation rates of the participating molecules.

Conclusions

In this work, we introduce a sequential dynamical Monte Carlo algorithm to simulate long-range proton transfer processes in biomolecules on timescales which are not accessible by other methods up to now. This algorithm allows us to simulate proton transfer processes in dependence on external parameters. We applied the new method to simulate the proton flux through gramicidin A as a function of pH and membrane potential. The calculated proton flux agrees well with experimental data, which gives us confidence to investigate the underlying proton transfer mechanism. In contrast to conventional dynamical Monte Carlo, the new method allows us to analyze our simulation in analogy to single molecule experiments. From this analysis, it can be seen that the proton can only cross the barrier, when the previously transferred proton has already left the channel. Thus at low pH, proton−proton interaction inside the channel is an important factor influencing the proton transfer through gramicidin A. By varying the electrostatic barrier for the proton transfer and the rotation rates of the water molecules, we analyzed the rate limiting process of the proton transfer through gramicidin A. We conclude that at physiological membrane potentials, i.e., between 0 and −250 mV, both aspects of the long-range proton transfer in gramicidin A, the electrostatic barrier and the reorientation of the hydrogen bonded network, are equally important.

By analyzing the proton transfer process, we could show that at low pH a proton which has entered the channel has to wait in front of the electrostatic energy barrier as long as a second proton is still in the channel. Once the proton has crossed the electrostatic energy barrier, it is transferred across the whole membrane with a probability of more than 90%.

The new sequential DMC algorithm can be applied to other proteins that are involved in charge transfer. The method can be straightforwardly extended to include electron transfer and coupled proton/electron transfer. It will allow to analyze the detailed mechanism of coupled charge transfer reactions in proteins on biologically relevant time scales.

Acknowledgment. This work was supported by the Deutsche Forschungsgemeinschaft (UL 174/7-1 and UL 174/4-3). We thank Donald Bashford for providing his program MEAD. R. Thomas Ullmann assisted us in designing the water model.

References and Notes

(1) Krishtalik, L. I. *Biochim. Biophys. Acta* **2000**, *1458*, 6–27.
(2) Nagle, J. F.; Morowitz, H. J. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 298–302.
(3) Agmon, N. *Chem. Phys. Lett.* **1995**, *244*, 456–462.
(4) Braun-Sand, S.; Burykin, A.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **2005**, *109*, 583–592.
(5) Warshel, A. *Acc. Chem. Res.* **2002**, *35*, 385–395.
(6) Warshel, A. *Chem. Rev.* **1993**, *93*, 2523–2544.
(7) Voth, G. A. *Acc. Chem. Res.* **2006**, *39*, 143–150.
(8) Braun-Sand, S.; Strajbl, M.; Warshel, A. *Biophys. J.* **2004**, *87*, 2221–2239.
(9) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226.
(10) Schmitt, U. W.; Voth, G. A. *J. Chem. Phys.* **1999**, *111*, 9361–9381.
(11) Friedman, R.; Nachliel, E.; Gutman, M. *Biochim. Biophys. Acta* **2005**, *1710*, 67–77.
(12) Marx, D. *Com. Phys. Commun.* **2006**, *7*, 1848–1870.
(13) Lill, M. A.; Helms, V. *J. Chem. Phys.* **2001**, *115*, 7993–8005.
(14) Ferreira, A.; Bashford, D. *J. Am. Chem. Soc.* **2006**, *128*, 16778–16790.
(15) Becker, T.; Ullmann, R. T.; Ullmann, G. M. *J. Phys. Chem. B* **2007**, *111*, 2957–2968.
(16) Gillespie, D. T. *J. Phys. Chem.* **1977**, *81*, 2340–2361.
(17) Fichthorn, K. A.; Weinberg, W. H. *J. Chem. Phys.* **1991**, *95*, 1090–1096.
(18) Burkhart, B. M.; Li, N.; Langs, D. A.; Pangborn, W. A.; Duax, W. L. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12950–12955.
(19) Eisenman, G.; Enos, B.; Hagglund, J.; Sandblom, J. *Ann. N.Y. Acad. Sci.* **1980**, *339*, 8–20.
(20) Andersen, O. S.; Koeppe, R. E.; B, R. *IEEE Trans. Nanobiosci.* **2005**, *4*, 10–20.
(21) Pomes, R.; Roux, B. *Biophys. J.* **2002**, *82*, 2304–2316.
(22) de Groot, B. L.; Tieleman, D. P.; Pohl, P.; Grubmüller, H. *Biophys. J.* **2002**, *82*, 2934–2942.
(23) Chernyshev, A.; Cukierman, S. *Biophys. J.* **2006**, *91*, 580–587.
(24) Cukierman, S. *Biophys. J.* **2000**, *78*, 1825–1834.
(25) Akeson, M.; Deamer, D. W. *Biophys. J.* **1991**, *60*, 101–109.
(26) Cukierman, S.; Quigley, E. P.; Crumrine, D. S. *Biophys. J.* **1997**, *73*, 2489–2502.
(27) Homeyer, N.; Essigke, T.; Ullmann, G.; Sticht, H. *Biochemistry* **2007**, *46*, 12314–12326.
(28) Klingen, A. R.; Bombarda, E.; Ullmann, G. M. *Photochem. Photobiol. Sci.* **2006**, *5*, 588–596.
(29) Ullmann, G. *J. Phys. Chem. B* **2003**, *107*, 1263–1271.
(30) Ullmann, G. M.; Knapp, E.-W. *Eur. Biophys. J.* **1999**, *28*, 533–551.
(31) Bombarda, E.; Becker, T.; Ullmann, G. M. *J. Am. Chem. Soc.* **2006**, *128*, 12129–12139.
(32) Landau, D. P.; Binder, K. *A Guide to Monte Carlo Simulations in Statistical Physics*; Cambridge University Press: New York, 2000.
(33) Prados, A.; Brey, J.; SanchezRey, B. *J. Stat. Phys.* **1997**, *89*, 709–734.
(34) Sadhukhan, S.; Munoz, D.; Adamo, C.; Scuseria, G. E. *Chem. Phys. Lett.* **1999**, *306*, 83–87.
(35) Pavese, M.; Chawla, S.; Lu, D. S.; Lobaugh, J.; Voth, G. A. *J. Chem. Phys.* **1997**, *107*, 7428–7432.
(36) Lill, M. A.; Helms, V. *J. Chem. Phys.* **2001**, *115*, 7985–7992.
(37) Barroso, M.; Arnaut, L. G.; Formosinho, S. J. *J. Phys. Chem. A* **2007**, *111*, 591–602.
(38) Cherepanov, D. A.; Junge, W.; Mulkidjanian, A. Y. *Biophys. J.* **2004**, *86*, 665–680.
(39) Mulkidjanian, A. Y.; Heberle, J.; Cherepanov, D. A. *Biochim. Biophys. Acta* **2006**, *1757*, 913–930.
(40) Espinosa, E.; Molins, E.; Lecomte, C. *Chem. Phys. Lett.* **1998**, *285*, 170–173.
(41) Woutersen, S.; Emmerichs, U.; Bakker, H. J. *Science* **1997**, *278* (5338), 658–660.
(42) Townsley, L. E.; Tucker, A, W.; Sham, S.; Hinton, F, J. *Biochemistry* **2001**, *40*, 11676–11686.
(43) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminatha, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(44) Swart, M.; Van Duijnen, P. T.; Snijders, J. G. *J. Chem. Phys.* **2001**, *22*, 79–88.
(45) Guerra, C. F.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Acc.* **1998**, *99*, 391–403.
(46) Bashford, D.; Gerwert, K. *J. Mol. Biol.* **1992**, *224*, 473–486.
(47) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(48) Im, W.; Beglov, D.; Roux, B. *Com. Phys. Commun.* **1998**, *111*, 59–75.
(49) Roux, B. *Biophys. J.* **1997**, *73*, 2980–2989.
(50) de Groot, B. L.; Frigato, T.; Helms, V.; Grubmüller, H. *J. Mol. Biol.* **2003**, *333*, 279–293.
(51) Jensen, M. O.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2003**, *85*, 2884–2899.

JP801477B

4.3 Manuscript B

McVol - A program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm

Mirco S. Till & G. Matthias Ullmann

Received: 31 March 2009 / Accepted: 23 May 2009

J Mol Model. 2009 Jul 22. [Epub ahead of print]

DOI 10.1007/s00894-009-0541-y

J Mol Model
DOI 10.1007/s00894-009-0541-y

ORIGINAL PAPER

McVol - A program for calculating protein volumes and identifying cavities by a Monte Carlo algorithm

Mirco S. Till · G. Matthias Ullmann

Received: 31 March 2009 / Accepted: 23 May 2009

Abstract In this paper, we describe a Monte Carlo method for determining the volume of a molecule. A molecule is considered to consist of hard, overlapping spheres. The surface of the molecule is defined by rolling a probe sphere over the surface of the spheres. To determine the volume of the molecule, random points are placed in a three-dimensional box, which encloses the whole molecule. The volume of the molecule in relation to the volume of the box is estimated by calculating the ratio of the random points placed inside the molecule and the total number of random points that were placed. For computational efficiency, we use a grid-cell based neighbor list to determine whether a random point is placed inside the molecule or not. This method in combination with a graph-theoretical algorithm is used to detect internal cavities and surface clefts of molecules. Since cavities and clefts are potential water binding sites, we place water molecules in the cavities. The potential water positions can be used in molecular dynamics calculations as well as in other molecular calculations. We apply this method to several proteins and demonstrate the usefulness of the program. The described methods are all implemented in the program McVol, which is available free of charge from our website at http://www.bisb.uni-bayreuth.de/software.html.

Keywords Cavities in proteins · Molecular volume · Monte Carlo · Water placement inside proteins

M. S. Till · G. M. Ullmann (✉)
Structural Biology/Bioinformatics, University of Bayreuth,
Universitätsstr. 30, BGI,
95447 Bayreuth, Germany
e-mail: Matthias.Ullmann@uni-bayreuth.de

Introduction

The identification of the surface of a protein has a long tradition in many fields of protein modeling and drug design [1–5]. The great interest in this subject is motivated by its importance for identifying ligand binding pockets and cavities in proteins. Moreover, protein crystal structures often show internal cavities that could be filled with water molecules. The identification of such water-filled cavities is important for the analysis of proton transfer networks in proteins, since these water molecules can play a role in hydrogen bond networks and therefore influence the long range proton transport within proteins [6–8]. Several methods have been developed to calculate the solvent accessible surface, molecular surface and molecular volume of a protein. Among them, algorithms based on the alpha shape theory are used in many approaches [2, 9, 10]. The alpha shape theory orders a subset of Delauny complexes with the aim of reducing the computational cost of an inclusion-exclusion formalism to calculate the protein surface and volume. An accurate computation of the molecular and solvent accessible surfaces and volumes is possible with this algorithm. However, the main drawbacks are numerical instabilities due to geometric degeneracy. The computation of the Delauny complexes are shown to be prone to such instabilities. A solution to this problem is found with the so-called “Simulation of Simplicity” [9] which is implemented for example in CASTp [2]. Other methods like LIGSITE [11], POCKET [12], or SURFNET [13] are grid based methods to define the protein surface and internal cavities or ligand binding sites. These methods are limited to the resolution of the grid they use. All these methods are basically methods for integrating the protein volume. Monte Carlo algorithms are known to be able to perform such integrations. A well-

Published online: 22 July 2009

known textbook example is the integration of a circle area for the determination of the number π [14]. Such an algorithm can also be used for determining the volume of proteins.

In this paper, we describe an efficient Monte Carlo algorithm for calculating protein volumes and for identifying internal cavities. Our new algorithm is neither dependent on grid resolutions nor is the algorithm prone to geometric degeneracy at any point of the integration. Based on the identified cavities, we suggest possible positions for water molecules and place these water molecules. We apply this program to several proteins of different sizes and compare our results with experimentally identified water positions. The program is available from our website at http://www.bisb.uni-bayreuth.de/software.html.

Methods

Theory of the volume integration

In our algorithm, we consider the protein to consist of spherical atoms. In order to define the molecular volume (MV), we calculate the solvent accessible surface (SAS), which is defined by rolling a probe sphere over the atoms of the protein [15]. The probe sphere represents a solvent molecule. Therefore the probe sphere radius is adjustable to match the desired solvent molecule radius. Figure 1 shows a schematic drawing of the scenario. The MV consists of two parts: the volume of the protein atoms and the volume of the voids, i.e., the volume between the atoms which is not solvent accessible. The MV can be determined by a Monte Carlo integration: A point is randomly placed in a box with known dimensions that contains the whole molecule and it is determined whether this random point is in the solvent or in the MV. From the ratio between points inside the MV and the total number of points, the MV can be calculated. If the box has a volume V_{box}, then the MV is given by

$$MV = \frac{n_{inside}}{n_{tot}} V_{box} \tag{1}$$

where n_{inside} is the number of points inside the MV and n_{tot} is the total number of points.

Whether a point is inside the MV or not is determined by the following steps:

1. If the point is closer to one atom than the van der Waals radius of this atom, the point is inside the van der Waals volume and therefore inside the MV, else
2. If the distance of the point to any atom center is smaller than the van der Waals radius of the atom plus the probe sphere radius and the distance to the closest point of the SAS is larger than the probe sphere radius, the point belongs to a void and therefore to the MV.
3. In any other case, the point belongs to the solvent.

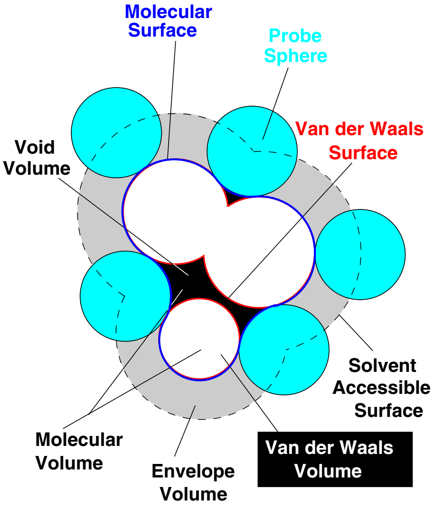


Fig. 1 Definition of volumes and surfaces of a molecule. The atoms of a molecule are represented as white spheres, the probe sphere as cyan spheres. The solvent accessible surface (dashed line) is defined by the center of the probe sphere when rolled over the atoms of the protein. The molecular surface (solid blue line) is defined by the surface points of the probe sphere closest to the protein atoms. The molecular volume consists of two parts, the Van der Waals volume of the atoms and the volume of the voids (shown in black) between these atoms. A void is defined as the space between atoms which is not solvent accessible.The molecular volume is represented by the area inside molecular surface (solid blue line). The solvent accessible surface encloses a volume that consists of three parts: the envelope region (gray), the Van der Waals volume (white), and the void volume (black)

For practical calculations, the SAS is represented by dots. The distance to the surface is than evaluated by calculating the distances to all surface points. In our implementation, we defined the surface points by the double cubic lattice method developed by Eisenhaber and coworkers [16]. This method can also be used to calculate the SAS by the following equation:

$$SAS = \sum_{i=1}^{N} 4\pi r_i^2 \frac{n_{surf f,i}}{n_{tot,i}} \tag{2}$$

where N is the number of atoms, r_i is the radius of atom i, $n_{surf,i}$ is the number of dots on the SAS of atoms i and $n_{tot,i}$ is the number of dots placed on atom i, no matter whether they are on the SAS or not.

The pseudocode for determining whether a random point is inside the molecular volume or not is given in the following:

```
point.inside_solvent = true;
point.inside_prot = false;
point.inside_void = false;
for (all atoms(i))
    {if (distance(point,atom(i)) <= atom(i).radius)
        {point.inside_prot = true;
         point.inside_solvent = false;
         break;
        }
    }
if (point.inside_solvent == true)
   {for (all atoms (i))
        {if ((distance(point, atom(i)) <  (atom(i).radius + probe.radius))
             && (distance(point,surface) > probe.radius))
            {point.inside_void = true;
             point.inside_solvent = false;
            }
        }
   }
```

Implementation of the volume integration

A direct implementation of the algorithm described above would give correct results for the volume calculation. However, it would be quite slow, since many distances need to be evaluated. To reduce the number of distance calculations, we used two cell-based neighbor list [14] (see Fig. 2), one for the atoms and another one for the surface dots. Two steps are necessary to create the neighbor list with a given grid spacing. The first step is to place a grid on the protein, where the maximal and minimal Cartesian coordinates of the grid points are the maximal Cartesian coordinates of the protein atoms extended by the maximal radius of the atoms and the probe sphere radius. In our implementation, we allow that the grid cells can have negative indices [17]. Each grid point is initialized as an empty linked list. The second step is to fill the linked lists with the nearby atoms or surface dots. The assignment of atoms to grid cells is done by running over all coordinates, dividing them by the grid spacing and rounding these values to the nearest integer (using the standard C-function rint ()). The rounded coordinates give the indices of the grid cell to which the atom or surface dot is associated. A pointer to the atom or surface dot is appended to the linked list at this grid position. Calculating the distance of a random point to the closest atom or surface dot is then accomplished by the following steps: The coordinates of the random point are divided by the grid spacing and these values are rounded to the nearest integer (using the standard C-function rint()). This procedure gives the indices of the grid cell to which the point is assigned. Now only the distance to atoms or surface dots assigned to the neighboring

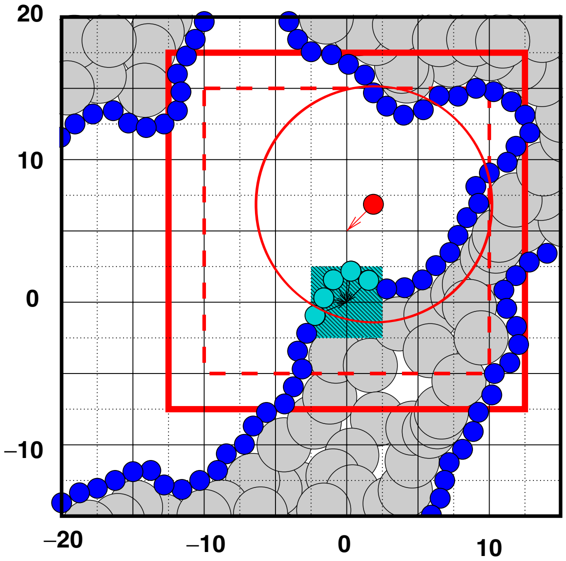


Fig. 2 Schematic drawing of the assignment of surface point to a neighbor list. The task is to find whether the distance of a random point (red) to a surface point (blue) is less than the probe sphere radius. Without a neighbor list, all surface points need to be evaluated until the first surface point within the probe sphere radius is found. In order to reduce the number of distance evaluations, a neighbor list is defined by mapping all surface points to a grid. For example, all cyan points are mapped to the grid point next to them (indicated by black arrows). All points within the cyan rectangle are mapped to this grid point. The random point is also assigned to a grid point. Now only the distance to the surface points in the neighboring grid cells (shown as the dashed red square) need to be evaluated. Only the surface points within the red circle have a distance to the random point that is smaller than the probe sphere radius. Using our neighbor list, only the distances to surface points that are within the solid read square are evaluated

grid cells needs to be calculated. How many neigboring grid cells need to be analyzed is determined as follows. All random points that are at least within a distance of the probe sphere radius plus the maximal atom radius need to be checked for determninig whether the random point is within the void or envelope region. In order to check whether the point is not in the envelope region, it needs to have a distance from any surface point that is larger than the probe sphere radius. These distances are divided by the respective grid spacing and rounded to the next highest integer *h* (using the standard function ceil()). Then, all distances to the atoms and surface points in the neighboring grid cells are evaluated. Suppose the random point was assigned to the grid cell with the index (i, j, k), the distances to all atoms or surface dots assigned to the grid cells $(i \pm h, j \pm h, k \pm h)$ are calculated. By this procedure, the number of distance calculations is reduced by orders of magnitude. It should be noted, that the grid resolution influences the speed of the program but not the accuracy of the volume calculations, since the points to calculate the volume are placed randomly in the box.

Identification of cavities

The procedure described above allows not only to calculate protein volume but also identify internal cavities. We have two ways to identify internal cavities in our calculation. First, it is possible to identify cavities based on the dot surface and second, based on the volume integration. We describe both possibilities in the following.

First, the surface is defined based on surface points marking the accessibility to the probe sphere. The surface of an internal cavity is described in the same way as the outside surface of the protein. We applied a graph search algorithm to separate surface points defining the outside surface of the protein from surface points defining internal cavities. The undirected graph is generated by connecting surface dots which are less than a certain distance (ca. 1 to 2 Å) apart using a cell-based neighbor list. The basic idea is to divide the graph in unconnected subgraphs. Typically, the largest subgraph describes the outer surface of the protein and smaller subgraphs describe internal cavities. The graph search is implemented as a breadth first search (BFS) [18]. To save memory, both, searching and building the graph is implemented in one routine, since it is not necessary to keep the connectivity matrix in the memory. The BSF methods starts by placing all surface dots in one graph. A vector representing all surface dots shows the graph division. This vector is initialized with 0 as graph number for all elements. Starting from the first element *i* in this vector, we assign the subgraph number 1 to this element and identify all neighboring surface dots. These neighboring surface dots are considered as connected in our graph and therefore the subgraph number 1 is assigned to these points. Additionally, these points are placed on a stack. If all connections of i are evaluated, a loop is started with an empty stack as termination condition. Within this loop, the last dot placed on the stack is taken from the stack and the subgraph number 1 is assigned to all neighboring dots, which do not already have a subgraph number. These dots are also placed on the stack. In each loop iteration, one dot is taken from the stack and all neighboring dots, which are not already in a subgraph are placed on the stack. Therefore, if the stack becomes empty, no more dots are in the whole graph which are connected to subgraph 1 but are not assigned to subgraph 1. If all dots of the surface are placed in subgraph 1, the whole graph is not dividable into subgraphs. If there are dots with 0 as subgraph number remaining in the vector, one of these dots is taken as the next starting point i for subgraph number 2. This procedure is repeated until all dots are assigned to a subgraph. If more than one subgraph is found by the BFS algorithm, subgraphs not connected to the outer protein surface can be defined as internal cavities. The surface of each subgraph can be calculated using Eq. 2.

Second, we can map the random points placed during the MC integration on a grid with a given resolution. Saving the number of points on a grid reduces dramatically the memory requirements compared to saving all random points individually. In each grid cell, we count the number of random points that were placed inside an atom, inside a void, and inside the solvent. A grid cell is marked as solvent as soon as one random point mapped to this grid cell was evaluated to be in the solvent. All grid cells not marked as solvent are considered to be inside the protein. Searching for cavities is accomplished by separating solvent grid cells completely surrounded by protein grid cells from solvent grid cells which are connected to the borders of the box. This separation is achieved by a BFS algorithm as explained above. An undirected graph is build from all grid cells. Within this graph a grid cell has a connection to a neighboring cell, if both grid cells are marked as solvent. After evaluating all grid cells at least one subgraph is found, defining the solvent surrounding the whole protein. If additional subgraphs of solvent grid cells are found these subgraphs are internal cavities. The volume of the internal cavities is integrated again by a Monte-Carlo algorithm. This time with a box placed only around the cavity. The resulting volume is more exact, since more random points are placed in a smaller volume. The volume is again evaluated by Eq. 1.

Detecting surface clefts

One problem connected to the calculation of the surface of a protein is the detection of large clefts on the surface reaching deep into the protein. A cleft is a solvent accessible pocket on the protein surface surrounded by a given ratio of protein. By default our algorithm would treat a cleft with a connection to the solvent as solvent accessible and therefore this cleft is treated as solvent and not as cavity. Several attempts to detected surface clefts were made [1, 2, 4, 5, 11–13, 19–24]. Our method for detecting internal cavities led us to an algorithm which is capable of detecting clefts on the protein surface. For testing if a solvent grid point belongs to a cleft, we place a box on each solvent grid point. The volume of this box is checked for points belonging to the protein or cleft. If more than a given percentage of grid points in the box are protein or cleft points, the solvent point is marked as cleft. Figure 3 schematically depicts the evaluation of a solvent point. This algorithm runs iteratively until no more cleft points are found. The points marked as clefts are divided into subgraphs using the BFS method describe above. The determined clefts are treaded like cavities in the program flow, except that the cleft volume is not reevaluated with a smaller box.

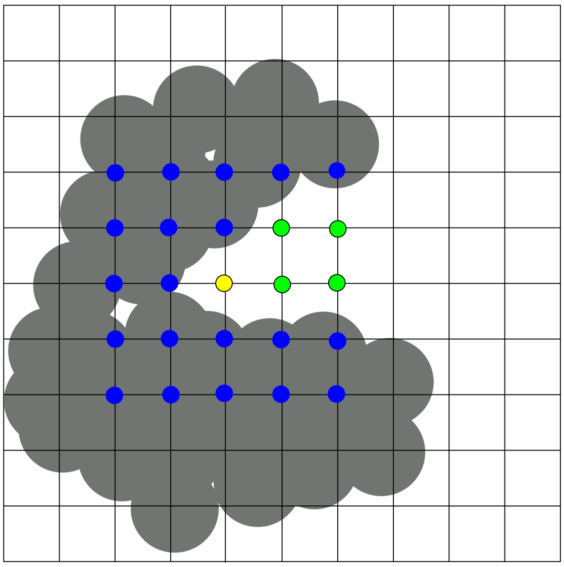


Fig. 3 Definition of clefts in proteins. The grey circles represent protein atoms. The yellow grid point (i,j,k) is a solvent grid point for which it is tested whether it is situated in a cleft or not. All grid cells in the two layers (i.e. i±2,j±2,k±2) are evaluated whether they are solvent grid points (green) or protein grid points (blue). The yellow grid point is considered to be situated in a cleft if a certain percentage of the surrounding grid points are protein grid points or cleft grid points

Placing water oxygen atoms

One reason for searching cavities in proteins is that they may contain water molecules. We place water molecules in all cavities with a volume larger than the volume of one water molecule. Based on the volume of each cavity, the number of water molecules each cavity can hold is determined by dividing the volume of the cavity by the volume of a water molecule. The result is rounded to the nearest integer. Initially, the atoms are place randomly inside the cavity by selecting a random solvent grid node that is far enough from the protein atoms. Starting from this configuration, a Monte Carlo method is applied to optimize the water positions on the grid.

We maximize the function D in Eq. 3

$$\begin{aligned} D = \sum_{i=1}^{K}\sum_{j=i+1}^{K} d(i,j) + \sum_{i=1}^{K} |x_i - x_{max}| \\ + \sum_{i=1}^{K} |y_i - y_{\mathrm{max}}| + \sum_{i=1}^{K} |z_i - z_{\mathrm{max}}| \\ + \sum_{i=1}^{K} |x_i - x_{\mathrm{m}i\mathrm{x}}| + \sum_{i=1}^{K} |y_i - y_{\mathrm{min}}| \\ + \sum_{i=1}^{K} |z_i - z_{\mathrm{min}}| \end{aligned} \tag{3}$$

where $d\,(i,j)$ is the distance between water molecule i and j and xyz_{min} and xyz_{max} are the minimal and maximal coordinates of the cavity, respectively. D is maximized by the Monte Carlo algorithm. Maximizing D ensures that the placed water molecules are as far apart from each other as possible and also as far apart as possible from the cavity borders. The algorithm moves one water molecule in a random direction at the grid and checks whether D has increased or not and if a water molecule at this position does not overlap with protein atoms. If the distance sum has increased, the new water position is accepted, otherwise, the move is discarded. The algorithm terminates after a given number of steps. By applying this algorithm, we ensure that the cavity is evenly filled with water molecules. Since no energy criteria are applied during the placement of water molecules, it is recommended to minimize the positions of the water molecules afterward.

Adding a membrane to membrane proteins

For electrostatic calculation on membrane proteins, it is often required to add dummy atoms around the protein representing the hydrophobic region of the membrane [25–27]. When such a membrane of dummy atoms is added, care must be taken, that internal cavities of the protein that are filled potentially by water molecules are not filled by dummy atoms. We implemented a procedure to add a dummy atom membrane in McVol to handle this problem.

Since the protein is placed in a box, all grid points of this box not assigned to a cavity or cleft are solvent grid points. On the basis of these grid points, McVol is capable of placing a membrane of dummy atoms around the protein. This membrane is built by defining an upper and lower border of the membrane. All solvent grid points within these borders (defined by the z-coordinates) are considered as membrane region. Grid points that are identified as cavities are not considered as membrane region in order to avoid that water filled cavities in the protein that are potentially important, for example for proton transfer, are filled with dummy atoms.

The overall flowchart of the program is given in Fig. 4.

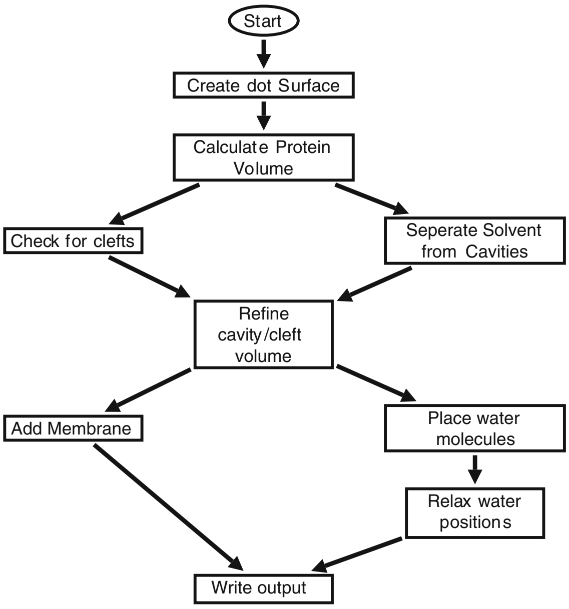


Fig. 4 Flowchart of the program McVol including the detection of potential water positions and adding a dummy atom membrane

Computational details

Structure preparation

All structures discussed in the following are derived from their pdb structures. Hydrogen atoms were added by the hbuild routine of CHARMM [28] and subsequently minimized. Atom radii were taken from Bondi [29] if not stated otherwise.

Computational details

All calculations were done with 50 Monte Carlo steps per Å^3 of the box volume and 2500 surface dots unless stated otherwise. The probe sphere radius was initially set to 1.3 Å in accordance to the water volume. The grid resolution for the initial grid was set to 1 Å, the cavity volume refinement was done with a grid resolution of 0.5 Å. Water molecules were only placed in cavities larger than 18 Å^3. The number of water molecules per cavity was determined by dividing the cavity volume by the volume of a water molecule and rounding the result.

Results

Convergence of the Monte Carlo algorithm

We tested the convergence of the Monte Carlo algorithm for calculating the volume of a molecule by varying the number of Monte Carlo steps per cubic Å of the box volume between 50 and 250. Moreover, we varied the number of points placed initially on each atom for the creation of the dot surface by the double cubic lattice method [16] between 500 and 10,000 per atom. We use 3-hydrobenzoate hydrolase (pdb code 2dkh) [30] as a test case. Each calculation was repeated 10 times in order to get an error estimation. The results are shown in Fig. 5. We observed no influence of the number of Monte Carlo steps

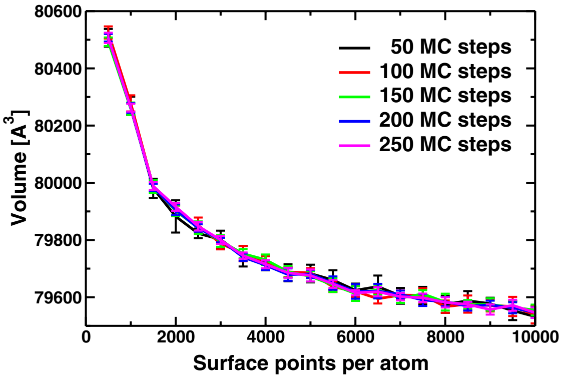


Fig. 5 Convergence of the protein volume determined by the program McVol in dependence on the number of Monte Carlo steps and surface points as atom. Molecular volume was calculated with 50 to 250 Monte Carlo steps per Å^3 box volume and 500 to 10000 surface points per atom. The protein 3-hydrobenzoate hydrolase (pdb code 2dkh) was used as an example

on the protein volume. All five calculations with the same number of surface points resulted in the same volume. The number of surface points influences the volume calculation, but only within a range of about 1%. Since the protein volume shows the strongest dependence for the increase in the number of surface points from 500 to 2000, we decided to take 2500 surface points for all further calculations unless otherwise stated. As shown in Fig. 5, the volume decreases with increasing number of surface points per atom. This behavior, which we term surface artifact, can be explained as follows. The decision, if a random point is inside a void or inside the envelope volume (see Fig. 1), is made based on the distance to the closest surface point. If the distance to the closest surface point is larger than the probe sphere radius, the point is inside a void. With fewer surface points, a random point which is located between two surface points might be treated as void point even if its real distance to the surface is less than the probe sphere radius and thus it should be considered as a point in the envelope volume. Since voids are included in the molecular volume, these misassigned points artificially increase the protein volume. However, as shown above, this effect only leads to a minor error. The number of Monte Carlo steps per Å^3 and the number of surface points per atom are the critical parameters for the runtime of the program. Table 1 gives a short overview of the runtime of the program in dependence of these two parameters. The runtime depends approximately linearly on the number of Monte Carlo steps with a slope of one. The dependence on the number of initial surface points is also linear but with a much smaller slope of about 0.01.

The relation between protein volume and number of atoms

We applied our algorithm to 15 enzymes between 896 and 20,835 atoms (see Table 2). In order to minimize the surface artifacts, we calculated the protein volume using 10,000 surface points per atom. For these proteins of different folds and molecular weights, we analyzed the volume of the voids, the volume of the protein and the ratios between these volumes. With one exception (2bgi) all structures show a similar ration between the protein volume and the number of atoms. The molecular volume is composed of the Van der Waals volume of the atoms and the volume of small voids between the atoms. Interestingly, the protein volume is directly correlated to the number of atoms, independent of the size or the folding of the protein (see Fig. 6). Linear regression leads to a slope of 8.04 Å^3/atom and a y-intercept of 102.9 Å^3. The y-intercept shows that the volume of the voids makes a significant contribution to the protein volume.

Cavities in proteins

The major goal of the above described algorithm is to find cavities in proteins. Identification of cavities in proteins is important for developing mechanistic models of the enzymatic activity, since cavities are often filled with water molecules that provide hydrogen bonds or are involved in proton transfer [31, 32]. The above described algorithm was applied to search cavities in three enzymes: Hen egg lysozyme, bacteriorhodopsin and the photosynthetic reaction center.

Table 1 Runtime of McVol (in seconds) for different parameter settings

MC steps per Å^3 box volume	Runtime [s] surface points per Atom								
	500	1000	1500	2000	2500	3000	4000	5000	10000
50	45	52	64	70	76	83	94	107	175
100	89	101	151	161	176	173	201	224	332
150	132	169	221	266	234	232	244	270	398
200	169	191	223	247	266	280	315	351	506
250	205	234	271	297	314	340	378	432	623

Table 2 Volume of 15 different proteins calculated by the program McVol

Protein	# atoms	Molecular volume [Å^3]	Volume/# atoms [Å^3]	vdW-Volume/ void-Volume
Bovine pancreatic tryp. inhibitor (1bpi) [40]	896	7325	8.175	3.648
Henn egg white Lysozyme (4lym) [34]	1967	16369	8.322	3.248
Bacterial BLUF photoreceptor (2byc) [41]	2262	17480	7.728	2.800
Bovine beta-lactoglobulin (1beb) [42]	2492	19668	7.892	2.646
Ferrodoxin NADP(H) reductase (2bgi) [43]	2716	31616	11.641	2.454
Bacteriorhodopsin (1c3w) [44]	3560	27483	7.720	2.788
Urate Oxidase (1r4u) [45]	4670	39054	8.363	3.155
Ammonuim transporter (2b2f) [46]	6140	45487	7.408	2.86
Alpha amylase (1bag) [47]	6446	53168	8.248	2.397
Cryptochrome (1np7) [48]	7842	62631	7.987	2.605
Glucose oxidase (1cf3) [49]	8803	73259	8.322	2.324
BM-40 FS/EC domain pair (1bmo) [50]	9145	72138	7.888	2.721
3-hydrobenzoate hydrolase (2dkh) [30]	9474	79876	8.431	3.027
Acetylene Hydratase (2e7z) [51]	11528	95304	8.267	2.363
Bacterial reaction center(2j8c) [26]	16738	138220	8.258	2.837
average			7.94±1.84	2.76±0.4

Hen egg lysozyme

NMR experiments identified three major cavities in hen egg lysozyme [33]. Each of these cavities is well defined by a set of amino acid side chains surrounding these cavities. We applied our algorithm to hen egg lysozyme (pdb-code 4lym [34]) using a probe sphere radius of 1.3 Å, 250 Monte-Carlo steps per Å^3 of the box volume and 2562 dots per atom on the dot surface. With this probe sphere radius we were not able to detect all of the experimentally reported cavities. Therefore we reduced the probe sphere radius to 1.1 Å. Applying our algorithm with the reduced probe sphere radius, we could reproduce the cavities proposed for hen egg lysozyme. The reduced probe sphere radius may be necessary since a water molecule is not a perfect sphere and the Bondi hydrogen radius may be too large for polar hydrogens.

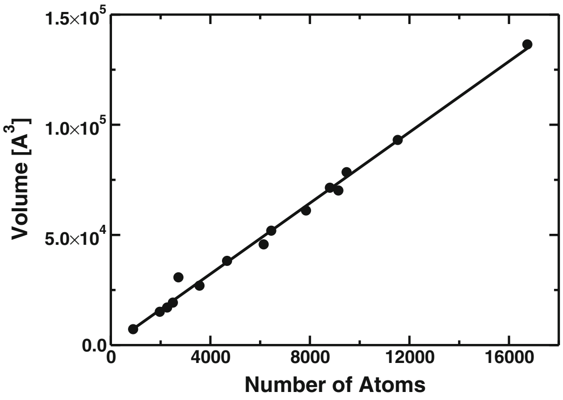


Fig. 6 Dependence of the molecular volume on the number of atoms. The red line is a regression of all points with a slope of 8.04 Å^3 and a y-intercept of 102.9 Å^3/atom

The experimentally determined cavities were found as two internal cavities and one cleft. The volumes of these cavities and the solvent accessible surfaces are listed in Table 3. The calculated volume of the first cavity is only approximated, since cavity I and the "hydrated cavity" as proposed by Otting et. al. [33] are merged to one cleft in our calculation. This cleft has three main clusters, each of equal size (see Fig. 7). The whole cleft has a size of 114 Å^3 therefore, cavity I was approximated to 38 Å^3. The "hydrated cavity" contains the water molecules 65, 70, and 75 in the pdb file 4lym. If cavity I is subtracted from the large cleft detected by our algorithm, the remaining volume of the "hydrated cavity" is 76 Å^3, which perfectly fits the three water molecules (see Fig. 7).

Table 3 Cavities found in the hen egg white lysozyme (4lym). The calculation was done with 250 MC steps per Å^3 box volume and 2500 surface points per atom

Cavity	Volume [Å^3]	SAS [Å^2]	Water molecules
I	38a	8.8	2
II	12	0.6	1
III	22	4.1	1
hydrated cavity	76	—	3

aVolume estimated from the cleft volume determined by McVol

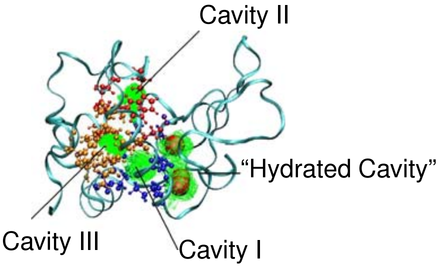


Fig. 7 Cavities found in Hen egg lysozyme. Colored residues show experimentally derived cavities. The large red speres represet three crystallographically resolved water molecules located in one large cleft

Bacteriorhodopsin

Water molecules are proposed in several proton transfer pathways through bacteriorhodopsin (BR) [35, 36]. Some of these water molecules are located near the retinal. We analyzed the cavities in the pdb-file 1c3w. We removed all experimentally derived water positions from the original file for this calculation. Our algorithm (applied with a probe sphere radius of 1.3 Å) was able to detect four cavities near the retinal. Cavity II perfectly fits the water molecules proposed to be involved in proton transfer. The calculated volumes and solvent accessible surfaces are shown in Table 4. The cavities are shown in Fig. 8. In addition, we compared the cavities found in BR with all experimentally derived water positions. Most of the experimentally derived water positions were also found as cavities by our algorithm. Lowering the probe sphere radius to 1.2 Å enabled us to find all experimentally derived water positions as cavities or clefts, except some positions which were on the surface of the protein and clearly not inside a cavity or a cleft. This result indicates that calculations with a probe sphere radius of 1.3 Å may not be able to identify all water filled cavities.

Photosynthetic reaction center

Many water molecules are participating in the proton transfer pathways in the photosynthetic reaction center [26, 37–39], but even in the x-ray structure with the highest resolution [26] not all cavities detected by McVol (using a probe sphere radius of 1.2 Å) are filled with water molecules. In addition to the crystallographically resolved water molecules, 35 cavities and surface clefts were found containing 103 water molecules. Some of these water molecules extend proposed proton transfer pathways connecting previously unconnected aminoacid sidechains participating in the proton transfer from the cytoplasmic site to the secondary quinone (Q_B). The location of the placed water molecules in the photosynthetic reaction center is shown in Fig. 9.

Table 4 Cavities found in the bacteriorhodopsin (1c3w) with a probe sphere radius of 1.3 Å. The calculation was done with 250 MC steps per $Å^3$ box volume and 2500 surface points per atom

Cavity	Volume [$Å^3$]	SAS [$Å^2$]	Water molecules
I	22	2.2	1
II	60	10.6	3
III	13	0.4	1
IV	43	9.0	2

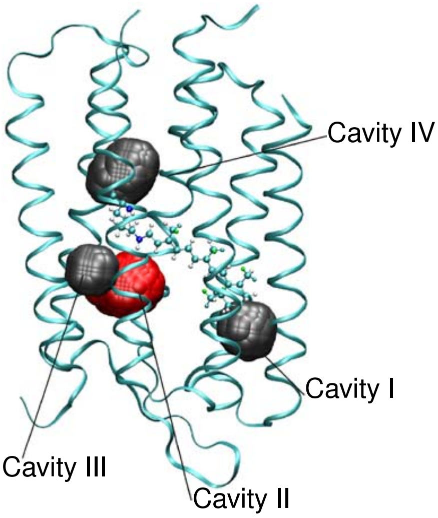


Fig. 8 Cavities found in bacteriorhodopsin. The red cavity fits three water molecules potentially involved in proton transfer in bacteriorhodopsin

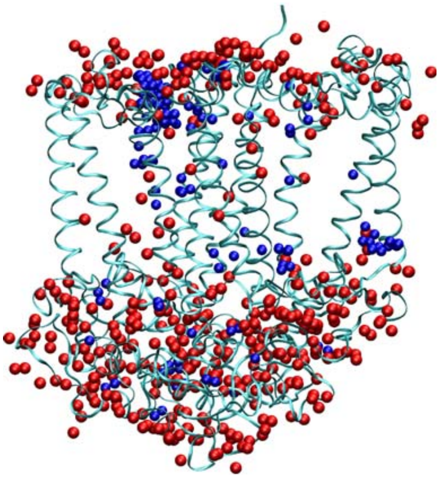


Fig. 9 Water molecules placed in the photosynthetic reaction center by the program McVol. Red spheres are crystallographically resolved water molecules, blue spheres are water molecules placed by McVol

Conclusion

In this work, we introduced a Monte Carlo algorithm for the calculation of protein volumes. Based on this algorithm, cavities inside the protein were located. The volume calculation are independent from any grid and therefore more accurate than the grid based methods developed so far.

The algorithm was applied to 15 proteins of different size. We found, that the ratio between the protein volume (including the volume of voids) and the number of atoms is almost the same for all sizes of proteins.

Our algorithm was able to reproduce experimentally derived cavities in the hen egg white lysozyme. Also the reported cavity volumes are in good agreement with our calculations. For bacteriorhodopsin, we could locate a cavity near the Schiff base maybe containing the water molecules important for the proton transfer process. An analysis of the cavities in the photosynthetic reaction center enabled us to place water molecules connecting originally separated proton transfer pathways through the protein. The Monte Carlo algorithm and the graph theoretical analysis of the protein volume, surfaces and cavities as well as the placement of water molecules is implemented in the program McVol. This program is able to calculate protein volumes, solvent accessible volumes and surfaces. McVol is available free of charge from our webpage http://www.bisb.uni-bayreuth.de/software.html.

Acknowledgements This work was supported by the DFG grant UL 174/7-1.

References

1. Kawabata T, Go N (2007) Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites. Structure 68:516–529
2. Liang J, Edelsbrunner H, Woodward C (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Prot Sci 7:1884–1897
3. Thornton JM, Todd AE, Milburn D, Borkakoti N, Orengo CA (2000) From structure to function: approaches and limitations. Nat Struct Biol 7:991–994
4. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol 161:269–288
5. Delaney JS (1992) Finding and filling protein cavities using cellular logic operations. J Mol Graph 10:174
6. Warshel A (2002) Molecular dynamics simulations of biological reactions. Acc Chem Res 35:385–395
7. Till MS, Essigke T, Becker T, Ullmann GM (2008) Simulating the proton transfer in gramicidin a by a sequential dynamical Monte Carlo method. J Phys Chem, B 112:13401–13410
8. Bondar AN, Elstner M, Suhai S, Smith JC, Fischer S (2004) Mechanism of primary proton transfer in bacteriorhodopsin. Structure 12:1281–1288
9. Edelsbrunner H, Mucke EP (1990) Simulation of simplicity - a techique to cope with degenerate cases in geometric algorithms. ACM Trans Graph 9:66–104
10. Xie L, Bourne PE (2007) A robust and efficient algorithm for the shape description of protein structures and its application in predicting ligand binding sites. Bioinformatics 8
11. Hendlich M, Rippmann F, Barnickel G (1997) LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. J Mol Graph 15:359
12. Levitt DG, Banaszak LJ (1992) POCKET - A computer-graphics method for identifying and displaying protein cavities and their surrounding amino acids. J Mol Graph 10:229–234
13. Laskowski RA (1995) SURFNET - A program for visualizing molecular surfaces, cavities and intermolecular interactions. J Mol Graph 13:323
14. Allen MP, Tildesley DJ (1989) Computer simulation of liquids. Oxford University Press
15. Lee B, Richards FM (1971) Interpretation of protein structures - estimation of static accessibility. J Mol Biol 55:379
16. Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M (1995) The double cubic lattice method - efficient approaches to numerical-integration of surface-area and volume and to dot surface contouring of molecular assemblies. J Com Chem 16:273–284
17. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recipes in C, 2nd edn. Cambridge University Press, Cambridge UK
18. Sedgewick (R) Algorithms in C++, part 5. Addison-Wesley, Boston
19. M. Masuya and J. Doi. Detection and Geometric Modeling of Molecular Surfaces and Cavities Using Digital Mathematical Morphological Operations. *J Mol Graph*, 13:331, 1995.
20. Peters KP, Fauck J, Frommel C (1996) The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. J Mol Biol 256:201–213
21. Ruppert J, Welch W, Jain AN (1997) Automatic identification and representation of protein binding sites for molecular docking. Prot Sci 6:524–533
22. Brady GP, Stouten PFW (2000) Fast prediction and visualization of protein binding pockets with PASS. J Comput Aided Mol Design 14:383–401
23. Venkatachalam CM, Jiang X, Oldfield T, Waldman M (2003) Ligandfit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J Mol Graph 21:289–307
24. Laurie ATR, Jackson RM (2005) Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics 21:1908–1916
25. Calimet N, Ullmann GM (2004) The influence of a transmembrane ph gradient on protonation probabilities of bacteriorhodopsin: the structural basis of the back-pressure effect. J Mol Bio 339(3):571–589
26. Koepke J, Krammer E-M, Klingen AR, Sebban P, Ullmann GM, 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
27. Klingen AR, Palsdottir H, Hunte C, Ullmann GM (2007) Redox-linked protonation state changes in cytochrome bc1 identified by Poisson-Bolt zmann electrostatics calculations. Biochem Biophys Acta 1767:204–221
28. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminatha S, Karplus M (1983) CHARMM - A programm for macromolecular energy, minimization, and dynamics calculations. J Com Chem 4:187–217
29. Bondi A (1964) Van der Waals volumes and radii. J Phys Chem 68:441
30. Hiromoto T, Fujiwara S, Hosokawa K, Yamaguchi H (2006) Crystal structure of 3- hydroxybenzoate hydroxylase from comamonas testosteroni has a large tunnel for substrate and oxygen access to the active site. J Mol Biol 364:878–896

31. Borodich AI, Ullmann GM (2004) Internal hydration of protein cavities: studies on BPTI. Phys Chem 6:1906–1911
32. Takano K, Funahashi J, Yamagata Y, Fujii S, Yutani K (1997) Contribution of water molecules in the interior of a protein to the conformational stability. J Mol Biol 274:132–142
33. Otting G, Liepinsh E, Halle B, Frey U (1997) NMR identification of hydrophobic cavities with low water occupancies in protein structures using small gas molecules. Nat Struct Biol 4:396–404
34. Kodanadapani R, Suresh CG, Vijayan M (1990) Crystal-structure of low humidity tetragonal lysozyme at 2.1A resolution. J Biol Chem 265:16126–16131
35. Lanyi JK (2006) Proton transfers in the bacteriorhodopsin photocycle. Bioch et Biophys Acta - Bioener 1757:1012–1018
36. Heberle J (2000) Proton transfer reactions across bacteriorhodopsin and along the membrane. Bioch et Biophys Acta - Bioener 1458:135–147
37. Okamura MY, Paddock ML, Graige MS, Feher G (2000) Proton and electron transfer in bacterial reaction centers. Biochim Biophys Acta 1458:148–163
38. Stowell MHB, McPhillips TM, Rees DC, Soltis SM, Abresch E, Feher G (1997) Light-induced structural changes in photosynthetic reaction center: implications for mechanism of electron-proton transfer. Science 276:812–816
39. Paddock ML, Feher G, Okamura MY (2003) Proton transfer pathways and mechanism in bacterial reaction centers. FEBS Lett 555:45–50
40. Parkin S, Rupp B, Hope H (1996) Structure of bovine pancreatic trypsin inhibitor at 125K: definition of carboxyl-terminal residues Gly57 and Ala58. Acta Crystallogr, Sect.D 52:18–29
41. Jung A, Domratcheva T, Tarutina M, Wu Q, Ko WH, Shoeman RL, Gomelsky M, Gardner KH, Schlichting L (2005) Structure of a bacterial BLUF photoreceptor: insights into blue light-mediated signal transduction. Proc Natl Acad Sci USA 102:12350–12355
42. Brownlow S, Cabral JHM, Cooper R, Flower DR, Yewdall SJ, Polikarpov I, North ACT, Sawyer L (1997) Bovine Beta-lactoglobulin at 1.8 angstrom resolution - still an enigmatic lipocalin. Structure 5:481–495
43. Nogues I, Perez-Dorado I, Frago S, Bittel C, Mayhew SG, Gomez-Moreno C, Hermoso JA, Medina M, Cortez N, Carrillo N (2005) The Ferredoxin-NADP(H) reductase from rhodobacter capsulatus: molecular structure and catalytic mechanism. Biochemistry 44:11730–11740
44. Luecke H, Schobert B, Richter HT, Cartailler JP, Lanyi JK (1999) Structure of bacteriorhodopsin at 1.55 angstrom resolution. J Mol Biol 291:899–911
45. Retailleau P, Colloc'h N, Vivares D, Bonnete F, Castro B, El Hajji M, Mornon JP, Monard G, Prange T (2004) Complexed and ligand-free high-resolution structures of Urate Oxidase (Uox) from aspergillus flavus: a reassignment of the active-site binding. Acta Crystallogr, Sect.D 60:453–462
46. Andrade SLA, Dickmanns A, Ficner R, Einsle O (2005) Crystal structure of the archaeal ammonium transporter amt-1 from archaeoglobus fulgidus. Proc Natl Acad Sci USA 102:14994–14999
47. Fujimoto Z, Takase K, Doui N, Momma M, Matsumoto T, Mizuno H (1998) Crystal structure of a catalytic-site mutant alpha-amylase from bacillus subtilis complexed with maltopentaose. J Mol Biol 277:393–407
48. Brudler R, Hitomi K, Daiyasu H, Toh H, Kucho K, Ishiura M, Kanehisa M, Roberts VA, Todo T, Tainer JA, Getzoff ED (2003) Identification of a new cryptochrome class: structure, function, and evolution. Mol Cell 11:59–67
49. Wohlfahrt G, Witt S, Hendle J, Schomburg D, Kalisz HM, Hecht HJ (1999) 1.8 and 1.9 angstrom resolution structures of the penicillium amagasakiense and aspergillus niger glucose oxidases as a basis for modelling substrate complexes. Acta Crystallogr, Sect.D 55:969–977
50. Hohenester E, Maurer P, Timpl R (1997) Crystal structure of a pair of follistatin-like and EF-hand calcium-binding domains in BM-40. EMBO J 16:3778–3786
51. Seiffert GB, Ullmann GM, Messerschmidt A, Schink B, Kroneck PMH, Einsle O (2007) Structure of the non-redox-active tungsten/[4Fe : 4S] enzyme acetylene hydratase. Proc Natl Acad Sci USA 104:3073–3077

4.4 Manuscript C

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

Received 7 January 2009, accepted 8 March 2009

J. Mol. Biol. (2009) 388, 631 - 643

DOI:10.1016/j.jmb.2009.03.020

doi:10.1016/j.jmb.2009.03.020 J. Mol. Biol. (2009) **388**, 631–643

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*

Received 7 January 2009; received in revised form 5 March 2009; accepted 8 March 2009
Available online
13 March 2009

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.

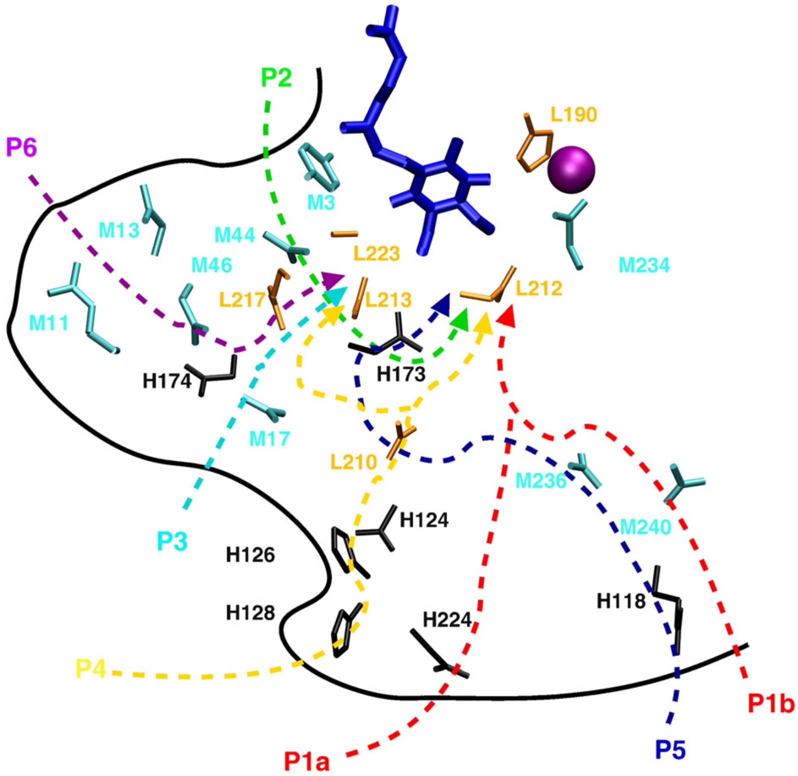


Fig. 1. Key residues for proton transfer to Q_B. All residues are colored according to their subunit (M = cyan, L = orange and H = black). Only side chains are shown. The proposed proton-transfer pathways P1 (red), P2 (green) and P3 (light blue),[4] P4 (yellow),[3] P5 (dark blue) and P6 (purple)[6] are 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.

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.

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 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

			Exchanged to (%)			
Subunit	Residue of *R. sphaeroides*	Conservation (%)	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

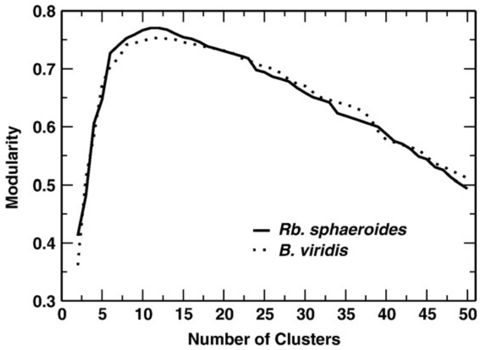


Fig. 2. Modularity of the clustering in dependence on the number of clusters for the Q_B network of the bRCs from *B. viridis* (dotted line) and from *R. sphaeroides* (continuous line).

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

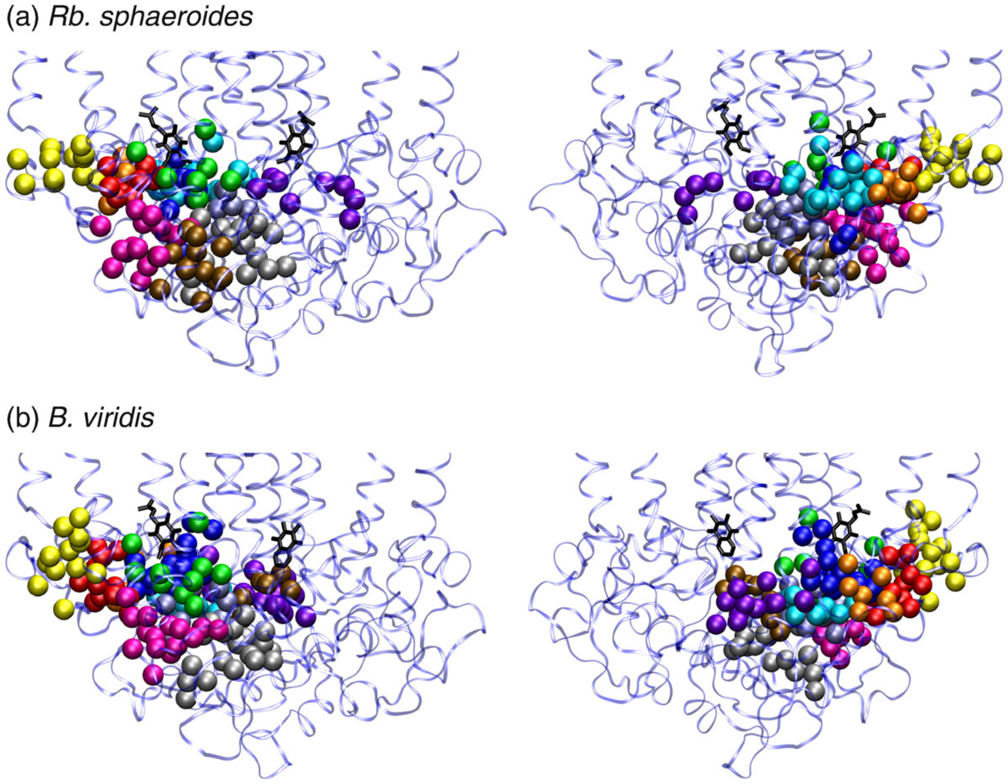


Fig. 3. Clusters of the Q_B network. The colors of the participating groups refer to the clusters (1 = green, 2 = magenta, 3 = red, 4 = yellow, 5 = blue, 6 = cyan, 7 = orange, 8 = violet, 9 = ice blue, 10 = gray and 11 = ocher). The clusters are shown for the bRCs of (a) *R. sphaeroides* and (b) *B. 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]

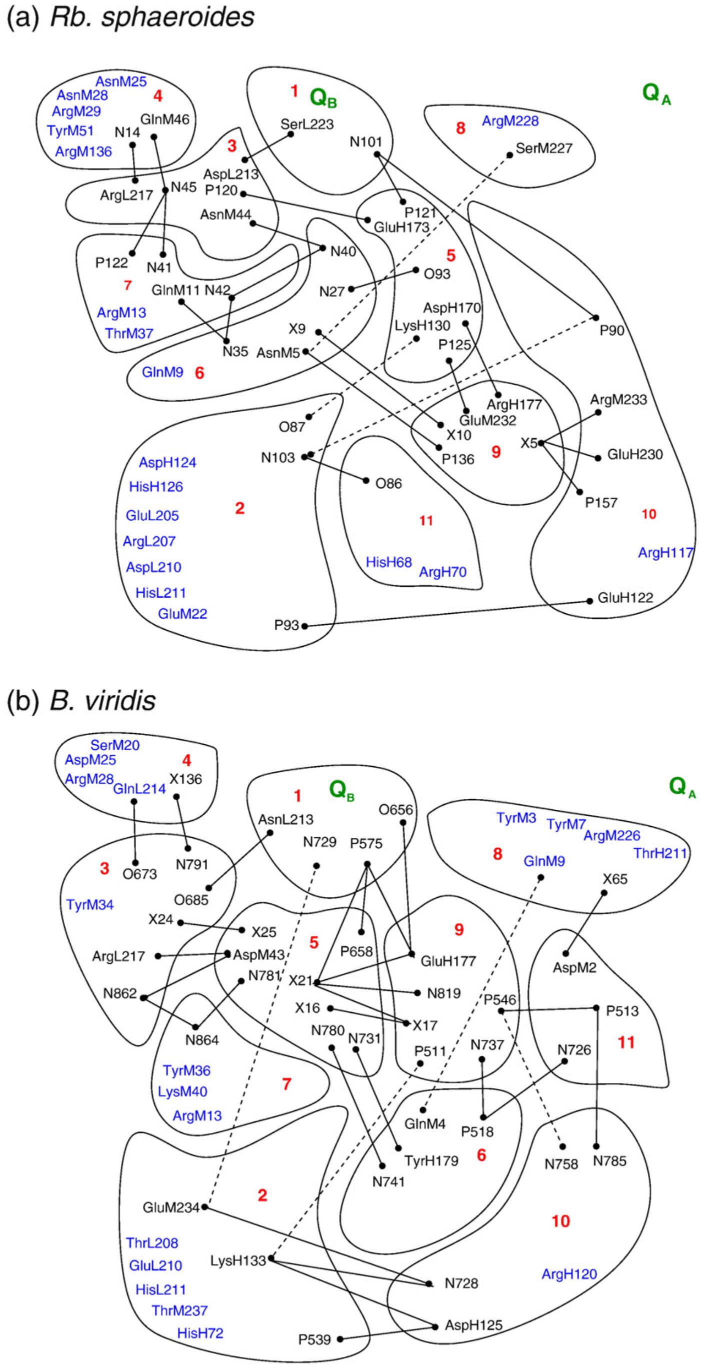


Fig. 4. Clusters of the Q_B network for the bRCS of (a) *R. sphaeroides* and (b) *B. viridis*. Cluster numbers are shown in red. The connections between the clusters and possible proton entry points (blue) are shown. Connections are shown as continuous lines or as dashed lines if a connection crosses 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.

Table 3. Possible proton entry points

	Protein residues			
	R. sphaeroides		*B. viridis*	
Subunit	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.

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 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 Å^3.

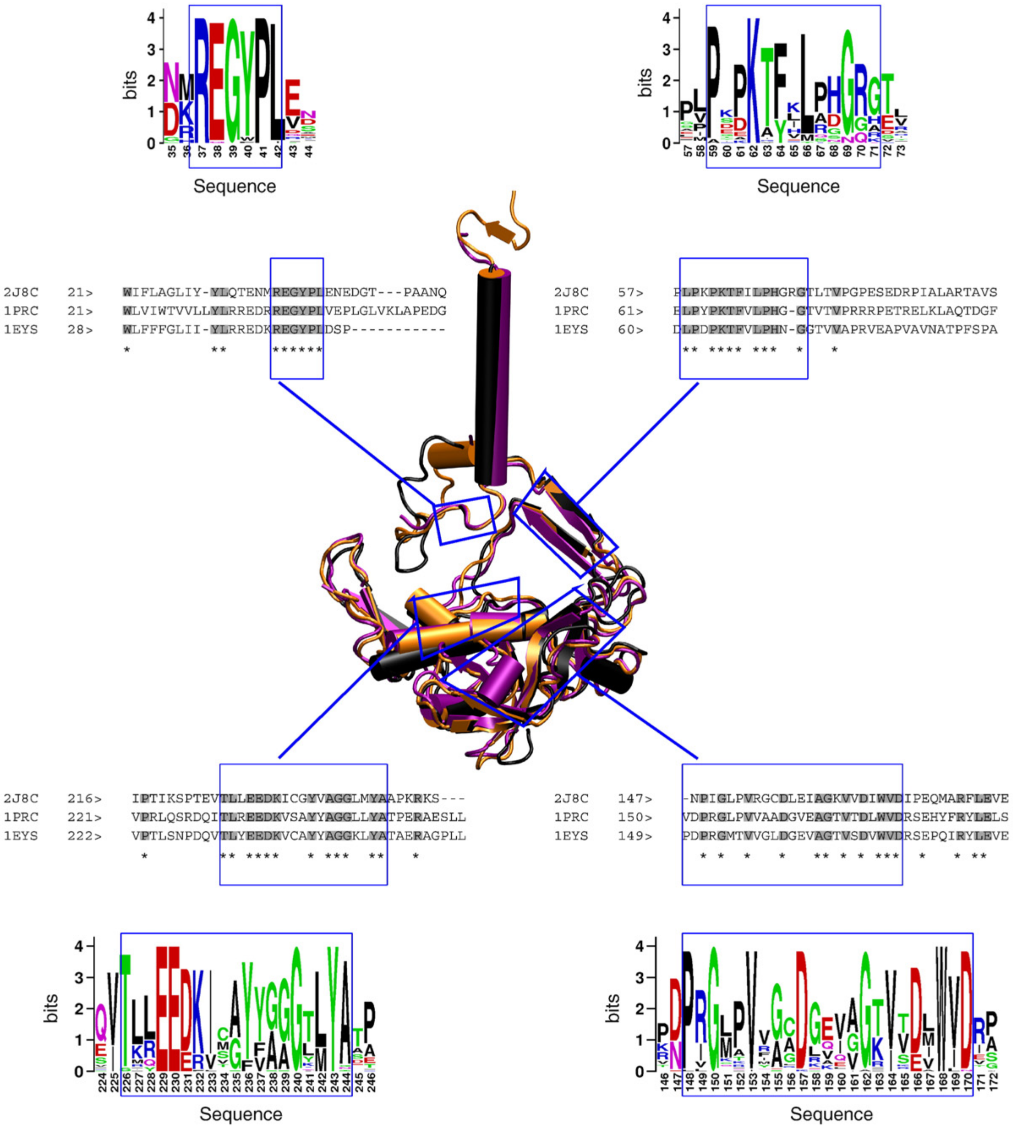

bits
Sequence
2J8C 21> WIFLAGLIY-YLQTENMREGYPLENEDGT---PAANQ
1PRC 21> WLVIWTVVLLYLRREDRREGYPLVEPLGLVKLAPEDG
1EYS 28> WLFFFGLII-YLRREDKREGYPLDSP-----------
* ** ******
2J8C 57> PLPKPKTFILPHGRGTLTVPGPESEDRPIALARTAVS
1PRC 61> ELPYPKTFVLPHG-GTVTVPRRRPETRELKLAQTDGF
1EYS 60> DLPDPKTFVLPHN-GGTVVAPRVEAPVAVNATPFSPA
** **** *** * *
2J8C 216> IPTIKSPTEVTLLEEDKICGYVAGGLMYAAPKRKS---
1PRC 221> VPRLQSRDQITLREEDKVSAYYAGGLLYATPERAESLL
1EYS 222> VPTLSNPDQVTLYEEDKVCAYYAGGKLYATAERAGPLL
* ** **** * *** ** *
2J8C 147> -NPIGLPVRGCDLEIAGKVVDIWVDIPEQMARFLEVE
1PRC 150> VDPRGLPVVAADGVEAGTVTDLWVDRSEHYFRYLELS
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.

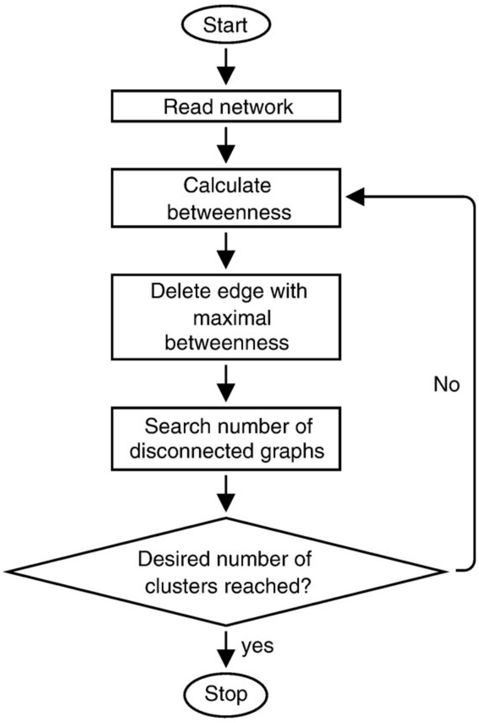

Start
Read network
Calculate betweenness
Delete edge with maximal betweenness
Search number of disconnected graphs
Desired number of clusters reached?
No
yes
Stop

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.

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 connections 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

This work was supported by the Deutsche Forschungsgemeinschaft through grant UL174/7-1 and by the Deutscher Akademischer Austauschdienst German/French PROCOPE through bilateral travel grants 11438QH and D/0502198.

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.

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 endglültig nicht bestanden habe.

Bayreuth, den 8. Februar 2009

Mirco S. Till