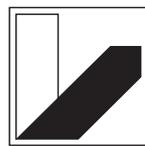

Structure-Function Relationship
of Archaeal Rhodopsin Proteins
analyzed by Continuum Electrostatics

D I S S E R T A T I O N

submitted to
the Faculty of Biology, Chemistry and Geoscience
of the University of Bayreuth, Germany



UNIVERSITÄT
BAYREUTH

for obtaining the degree of
Doctor of Natural Sciences

presented by
Edda Kloppmann
born in Lüneburg, Germany

Bayreuth 2010

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

Die vorliegende Arbeit wurde im Zeitraum von November 2002 bis November 2003 am IWR Heidelberg und von Dezember 2003 bis April 2007 am Lehrstuhl für Biopolymere an der Universität Bayreuth unter der Leitung von Professor G. Matthias Ullmann erstellt.

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

Ferner erkläre ich, dass ich nicht bereits anderweitig mit oder ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

Bayreuth, den 13.01.2010

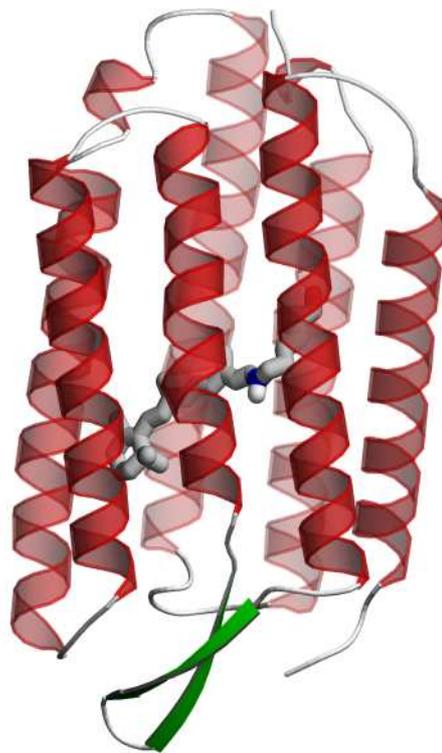
Prüfungsausschuss:

- | | |
|------------------|-------------------------------|
| 1. Gutachter: | Prof. Dr. G. Matthias Ullmann |
| 2. Gutachter: | Prof. Dr. Franz X. Schmid |
| Prüfungsvorsitz: | Prof. Dr. Thomas Hellweg |
| | Prof. Dr. Benedikt Westermann |

Einreichung des Promotionsgesuchs: 13.01.2010

Kolloquium: 04.05.2010

Structure-Function Relationship
of Archaeal Rhodopsin Proteins
analyzed by Continuum Electrostatics



A B S T R A C T

Rhodopsin proteins perform two cellular key functions: signaling of external stimuli and ion transport. Examples of both functional types are found in the family of archaeal rhodopsins, namely the proton pump bacteriorhodopsin, the chloride pump halorhodopsin and the photoreceptor sensory rhodopsin II. For these three membrane proteins, high-resolution X-ray structures are available, allowing a theoretical investigation in atomic detail. In this thesis, calculations are presented based on a continuum electrostatics approach using a finite-difference discretization of the Poisson-Boltzmann equation. The results can be divided into three parts.

One of the interesting features of rhodopsin proteins is the extreme range over which the absorption maximum of their chromophore retinal is tuned. This characteristic and the precision of the tuning mechanism is a fundamental requirement for color vision. Using the archaeal rhodopsins as model systems, this work aims at advancing the understanding of the inter-protein absorption shift. The presented results demonstrate that the electrostatic interactions of the protein with the retinal are a major determinant of the inter-protein shift. The differences in electrostatic potential that the proteins cause at the retinal could be assigned to seven residues. A generalized model of a quantum mechanical particle in a box including the electrostatic potential as a parameter allows a qualitative description of the absorption maxima.

Bacteriorhodopsin has become one of the most important model systems in the field of bioenergetics. This is due to its relative simplicity making it amenable to experimental and theoretical studies. Here, the probability of functionally relevant protonation states is calculated to characterize the available structures. The protonation behavior of the key residues of proton transfer and the correlation between the protonation of these residues is analyzed. The results show that with respect to the protonation the bR, K, L and M₁ intermediate state are well represented by the available structures, while the M₂, N and O intermediate state are less well represented.

An algorithm is introduced that determines a gap-free list of the lowest energy states. Such a list allows to analyze the ensemble of states accessible to a system in a certain energy range and, thus, can provide useful insight into the functional mechanism. The newly developed algorithm, termed X-DEE, is based on the dead-end elimination theorem. The X-DEE algorithm is applicable to a wide range of problems, for instance in protein design attempts. Here, X-DEE is successfully applied to bacteriorhodopsin to obtain gap-free lists of the lowest energy protonation states.

Z U S A M M E N F A S S U N G

Rhodopsine üben zwei zelluläre Schlüsselfunktionen aus, die Weiterleitung externer Signale und den Transport von Ionen. Beispiele beider Funktionstypen finden sich in der Familie der archaealen Rhodopsine: die Protonenpumpe Bacteriorhodopsin, die Chloridionenpumpe Halorhodopsin und den Photorezeptor Sensorisches Rhodopsin II. Ihre hochaufgelösten Röntgenkristall-Strukturen ermöglichen theoretische Untersuchungen auf atomarer Ebene. Die in der vorliegenden Arbeit durchgeführten Berechnungen basieren auf einem Kontinuum-Elektrostatik-Modell, das durch die Poisson-Boltzmann Gleichung gegeben ist, die mit Finite-Differenzen diskretisiert wird. Die Ergebnisse können in drei Abschnitte unterteilt werden.

Ein Charakteristikum der Rhodopsine ist der extreme Bereich, über den diese das Absorptionsmaximum ihres Chromophors Retinal einstellen. Diese Eigenschaft und die Präzision dieses Mechanismus ist eine grundlegende Voraussetzung des Farbsehens. Die archaealen Rhodopsine werden in der vorliegenden Arbeit als Modellsysteme genutzt, um die Verschiebung der Absorption zwischen den Proteinen zu untersuchen. Die vorliegenden Ergebnisse zeigen elektrostatische Wechselwirkungen des Proteins mit dem Retinal als entscheidenden Faktor für die Verschiebung der Absorption. Die Unterschiede des elektrostatischen Potentials, das die Proteine am Retinal verursachen, kann sieben Aminosäuren zugeschrieben werden. Ein verallgemeinertes Modell eines quantenmechanischen Teilchens in einer Box mit dem elektrostatischen Potential als Parameter ermöglicht eine qualitative Beschreibung der Absorptionsmaxima.

Bacteriorhodopsin ist eines der wichtigsten Modellsysteme der Bioenergetik. Da es ein relativ einfaches Protein ist, sind experimentelle und theoretische Studien gut durchführbar. In dieser Arbeit wird die Wahrscheinlichkeit funktionell relevanter Protonierungszustände berechnet, um die vorhandenen Strukturen genauer zu charakterisieren. Das Protonierungsverhalten der Schlüsselgruppen des Protonentransfers und die Korrelation der Protonierung dieser Gruppen wird analysiert. Die Ergebnisse zeigen, dass der bR-, K-, L- und M₁-Zustand bezüglich der Protonierung gut, während der M₂-, N- und O-Zustand weniger gut durch die vorhandenen Strukturen repräsentiert sind.

Ein Algorithmus wird eingeführt, der eine lückenlose Liste der Zustände niedrigster Energie ermittelt. Eine solche Liste ermöglicht die Untersuchung des gesamten Zustandsensembles, welches dem System in einem bestimmten Energiebereich zugänglich ist. Dadurch können hilfreiche Erkenntnisse über den funktionellen Mechanismus gewonnen werden. Der neu entwickelte Algorithmus, X-DEE genannt, basiert auf dem Theorem über die Dead-End-Eliminierung. Der X-DEE-Algorithmus ist in den unterschiedlichsten Gebieten verwendbar, unter anderem in Protein-Design-Ansätzen. In dieser Arbeit wird X-DEE erfolgreich angewandt, um lückenlose Listen der Protonierungszustände niedrigster Energie für Bacteriorhodopsins zu erhalten.

LIST OF PUBLICATIONS

‘When I use a word,’
Humpty Dumpty said in rather a scornful tone.
‘It means just what I choose it to mean –
neither more or less.’

Lewis Carroll, Alice in Wonderland

- [1] Edda Kloppmann, Torsten Becker and G Matthias Ullmann.
Electrostatic potential at the retinal of three archaeal rhodopsins: implications for their different absorption spectra. *Proteins*, 61: 953-965, 2005.
- [2] Edda Kloppmann, G Matthias Ullmann and Torsten Becker.
An extended dead-end elimination algorithm to determine gap-free lists of low energy states. *J Comput Chem*, 28: 2325-2335, 2007.
- [3] G Matthias Ullmann, Edda Kloppmann, Timm Essigke, Eva-Maria Krammer, Astrid R Klingen, Torsten Becker and Elisa Bombarda.
Investigating the mechanisms of photosynthetic proteins using continuum electrostatics. *Photosynth Res*, 97: 33-53, 2008.
- [4] Edda Kloppmann and G Matthias Ullmann.
Protonation states and correlation of the protonation behavior of the key residues of proton transfer in bacteriorhodopsin: a continuum electrostatic analysis of the photocycle intermediate structures. In preparation.

ACKNOWLEDGEMENTS

For all that has been, thanks.

For all that will be, yes.

Dag Hammarskjöld

This work would not have been possible without the support and encouragement of my colleagues, friends and family and I deeply value and appreciate the help of all who - directly or indirectly - contributed to this thesis.

First and foremost, I would like to sincerely thank my supervisor Matthias Ullmann for giving me the opportunity to work on a topic I enjoyed, for guiding the progress of my work with countless helpful and inspiring discussions, and for his unequivocal support.

Thanks to the whole BISB-group for making it a pleasure to work there. Special thanks go to Torsten Becker who collaborated on the solution for the generalized model of a particle in a box and on the development of the X-DEE algorithm. His ideas and his patience have greatly contributed to this work. Thanks to Timm Essigke, who installed and maintained our computer network, for his much appreciated technical help. Astrid Kligen and Elisa Bombarda have been splendid roommates and I thank both for all work-related and non-work-related discussions. Many, many thanks to Thomas Ullmann for proofreading this thesis and for his valuable comments, in particular with respect to the finite difference method. I am indebted to Siriporn Promsri who provided delicious food and comfortable shelter in times of need and, above all, is a wonderful friend.

Durba Sengupta and Lars Meinhold have been an inexhaustible source of knowledge in physics and chemistry (and membrane models). They proofread manuscripts and this thesis and, most importantly, never failed to make me laugh.

Sincere thanks are given to my parents Ursula and Udo Kloppmann for their support and for everything they taught me. I am indebted to my grandmothers Katharina Kloppmann and Luise Wilhelm and would like to thank them for being a wonderful example.

My heartfelt gratitude is due to Rainer Warnke for his invaluable help with mathematics and informatics, his encouragement, his unfailing support and everything else.

CONTENTS

Abstract	i
Zusammenfassung	iii
List of Publications	v
Acknowledgements	vii
Abbreviations	xiii
1 Introduction	1
1.1 The Rhodopsin Proteins	2
1.1.1 Ion Pumps	3
1.1.2 Photoreceptors	4
1.1.3 The Two Rhodopsin Protein Families	4
1.2 Halophilic Archaea	5
1.3 Archaeal Rhodopsins	5
1.3.1 The Proton Pump Bacteriorhodopsin	8
1.3.2 The Chloride Pump Halorhodopsin	13
1.3.3 The Photoreceptor Sensory Rhodopsin II	14
1.3.4 Change of Function for BR, HR and SRII	15
1.4 The Opsin Shift	15
1.5 Outline of the Thesis	17
2 Continuum Electrostatics	19
2.1 Building Computer Models	19
2.2 Conceptual Model	21
2.3 Electrostatic Potential in Dielectric Media	22
2.3.1 The Poisson-Boltzmann Equation	22
2.3.2 The Linearized Poisson-Boltzmann Equation	24
2.3.3 The Finite Difference Method	25
2.4 Protonation Equilibria in Proteins	27

2.4.1	The Intrinsic pK_a Value	29
2.4.2	The Interaction Energy $W_{\mu\nu}$	30
2.4.3	The Free Energy of a Protonation State	31
2.5	Calculation of Protonation Probabilities	31
2.5.1	Protonation Probability of a Single Site	32
2.5.2	Correlation Between Two Sites	32
2.5.3	Example: Correlation in DTPA	34
2.5.4	Probability of Protonation State Subsets	37
2.5.5	Metropolis Monte Carlo	38
3	High-Resolution Rhodopsin Protein Structures and Structure Preparation	41
3.1	Protein Structure Prediction	42
3.2	Experimental Protein Structures	42
3.2.1	Membrane Proteins	43
3.2.2	Crystallization of Membrane Proteins	43
3.2.3	Archaeal Rhodopsin Structures	45
3.3	Preparation of the X-Ray Structures	47
3.4	The Membrane Model	50
3.5	Calculation of the pK_{intr} Value and the Interaction Energy $W_{\mu\nu}$	51
4	The Inter-Protein Shift	55
4.1	Absorption Tuning in Rhodopsin Proteins	55
4.1.1	The Inter-Protein Shift in Archaeal Rhodopsins	56
4.1.2	QM/MM Calculations and Mutational Studies	57
4.1.3	Objective	58
4.2	Computational Details	58
4.3	Protonation State of BR, HR and SRII	59
4.4	Electrostatic Potential of BR, HR and SRII at the Retinal	60
4.5	Decomposition of the Electrostatic Potential	62
4.5.1	The Counterion Complex	62
4.5.2	The Retinal Binding Pocket	64
4.5.3	Potential of Individual Residues	66
4.6	Qualitative Model of the Inter-Protein Shift	70
4.6.1	Model of a Particle in a Box	71
4.6.2	Model of a Particle in a Box with Step Potential	71
4.7	Concluding Remarks	74

5 Protonation Probabilities and Correlations in BR	77
5.1 Proton Transfer of Bacteriorhodopsin	78
5.2 Computational Details	80
5.3 Protonation State Subsets	81
5.3.1 BR Structures	82
5.3.2 K Structures	83
5.3.3 L Structures	84
5.3.4 M Structures	85
5.3.5 N Structure	87
5.3.6 O-like Structure	88
5.3.7 Synopsis of Section 5.3	88
5.4 Protonation Probabilities of the Key Residues of Proton Transfer	89
5.4.1 Protonation Probability of Asp96	90
5.4.2 Protonation Probability of Asp115	90
5.4.3 Protonation Probability of the Retinal Schiff Base	91
5.4.4 Protonation Probability of Asp85	94
5.4.5 Protonation Probability of Asp212	95
5.4.6 Protonation Probability of Arg82	97
5.4.7 Protonation Probability of the Proton Release Group	98
5.5 Correlation of the Protonation Behavior	101
5.5.1 Correlation between Asp85 and the Retinal Schiff Base	102
5.5.2 Correlation between Asp96 and the Retinal Schiff Base	105
5.5.3 Correlation between Asp85 and Asp115	107
5.5.4 Correlation between Asp115 and the Retinal Schiff Base	108
5.5.5 Correlation between Glu194 and Glu204	111
5.5.6 Correlation between Asp85 and Glu194/Glu204	113
5.5.7 Correlation between Glu194/Glu204 and the Schiff Base	116
5.6 Implications for the Intermediate Structures	119
6 X-DEE: an Algorithm to Determine All States of Lowest Energy	121
6.1 Dead-End Elimination: an Algorithm to Determine the Lowest Energy State	122
6.1.1 The DEE Criterion	123
6.1.2 Objective to Obtain a List of the Lowest Energy States	125
6.2 X-DEE: an Algorithm to Determine a Gap-Free List of Lowest Energy States	126
6.2.1 Construction of a Search Basis \mathcal{B} for $\mathcal{M} \setminus \mathcal{L}$	127

6.2.2	Completeness of the Search Basis \mathcal{B}	129
6.3	Computational Details	131
6.4	Statistics of the Search Basis \mathcal{B}	131
6.4.1	Total Number of Search Keys	132
6.4.2	Properties of the Search Keys	133
6.4.3	Number of New Keys for Each New State in \mathcal{L}	134
6.5	Applicability of X-DEE	135
7	Application of X-DEE to Bacteriorhodopsin	137
7.1	Implementation of X-DEE	138
7.2	Computational Details	141
7.3	General Analysis of the Lowest Energy States	142
7.3.1	Convergence of the Partition Function	142
7.3.2	Energy Distribution of the Lowest Energy States	144
7.3.3	Probability of the Lowest Energy States	144
7.4	The Proton Pump Bacteriorhodopsin	146
7.4.1	Functionally Relevant Protonation States of BR	146
7.4.2	Distribution of Functionally Relevant Protonation States	147
7.4.3	Implications for the Proton Transfer in BR	148
7.5	Calculating the Kinetics of Proton Transfer	148
8	Concluding Remarks and Outlook	149
	Bibliography	157

A B B R E V I A T I O N S

ADP	adenosin diphosphate
ATP	adenosin triphosphate
BR	bacteriorhodopsin
CASP	critical assessment of techniques for protein structure prediction
CHARMM	chemistry at Harvard macromolecular mechanics
DEE	dead-end elimination
DTPA	diethylene triamine pentaacetic acid
EDTA	ethylene diamine tetraacetic acid
FAD	flavin adenine dinucleotide
FLOP	floating point operations per second
GPCR	G protein-coupled receptor
Hb. sal.	Halobacterium salinarum
HOMO	highest occupied molecular orbital
HR	halorhodopsin
HtrII	halobacterial transducer for sensory rhodopsin II
LUMO	lowest unoccupied molecular orbital
MEAD	macroscopic electrostatics with atomic detail
MSMS	maximal speed molecular surface
NAD ⁺	nicotinamide adenine dinucleotide
NMR	nuclear magnetic resonance
MCT	Monte Carlo titration
QM/MM	quantum mechanics/molecular mechanics
SRI	sensory rhodopsin I
SRII	sensory rhodopsin II
X-DEE	extended dead-end elimination

CHAPTER 1

INTRODUCTION

'What is the secret of life?' I asked.

'I forget,' said Sandra.

'Protein,' the bartender replied.

Kurt Vonnegut, Cat's Cradle

Theoretical methods have become a powerful tool in biophysics, biochemistry and biology complementing experimental research. This trend can be ascribed to the development of efficient algorithms and the enormous increase in computational power. For example, sophisticated algorithms are utilized to process the tremendous amount of data that has been accumulated by experimental research, *e.g.*, gene and protein sequences. The versatility of theoretical approaches to elucidate the function of biochemical and biological systems is described by numerous excellent textbooks published in the last decade [1–5]. The advance of theoretical methods is apparent both on the organismic and cellular as well as the molecular level.

Today, the complete genome of many organisms is known, among them the human DNA sequence. Systems biology gathers data, for example, on biochemical pathways, neuronal networks and the proteome of cells. The interpretation of these data relies heavily on computational methods. On the one hand this is due to the sheer amount of data to be analyzed. On the other hand, biological processes are often highly complex and correlated such that intuitive explanations for the workings of these systems in general cannot be given, but rely on a rigorous quantitative approach.

The structure-function relationship of molecules on the atomic level is accessible to experimental analysis only to a limited degree. Although the protein data bank provides over 60 000 high-resolution structures of proteins, a much lesser number of which belonging to different proteins, their functional mechanism is commonly not well understood [6]. Theoretical structural biology analyzes molecules in atomic detail. Using quantum mechanical approaches also chemical reactions can be modeled. Here, the interaction of theoretical and experimental research can provide useful insight into the underlying causes of an observed phenomena.

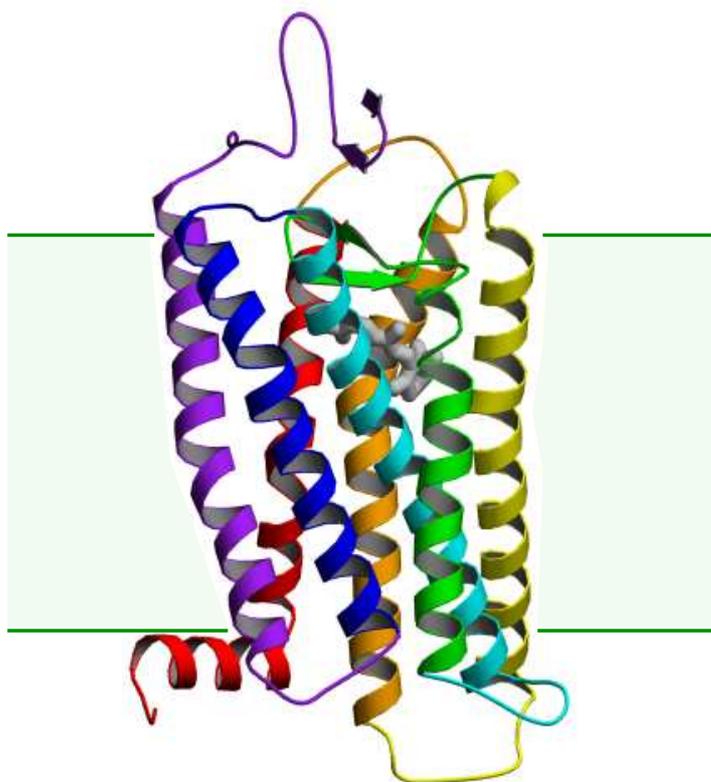


Figure 1.1. The visual pigment rhodopsin. The figure shows the high-resolution X-ray structure of bovine rhodopsin [7]. The membrane is indicated by the green slab. The retinal chromophore is depicted explicitly. In the ground state, rhodopsin binds 11-*cis* retinal which isomerizes after light absorption to all-*trans* retinal (*cf.* Figure 1.2). Rhodopsin signals the absorption of light to its transducer, the G protein transducin, and ultimately to the nervous system.

The present thesis aims at advancing the understanding of the functional mechanism of archaeal rhodopsins. For the analysis of these proteins a structure-based theoretical approach, namely continuum electrostatics given by the Poisson-Boltzmann equation, is employed. In the following section, the rhodopsin proteins are introduced. After briefly addressing these proteins in general, the discussion turns to the archaeal rhodopsins, in particular bacteriorhodopsin, halorhodopsin and sensory rhodopsin II, on which this thesis is focused. The chapter concludes with a general outline of this thesis.

1.1 THE RHODOPSIN PROTEINS

Rhodopsins are widespread in Nature. Over 300 rhodopsins have been described so far. They can be found in all three domains of life: archaea, bacteria and eukarya [8–11]. All rhodopsins share a common motif of seven transmembrane helices (*cf.* Figure 1.1) and bind the chromophore retinal to a lysine residue in the 7th helix, thereby forming a Schiff base. Retinal is a vitamin A aldehyde, consisting of a polyene chain and a β -ionone

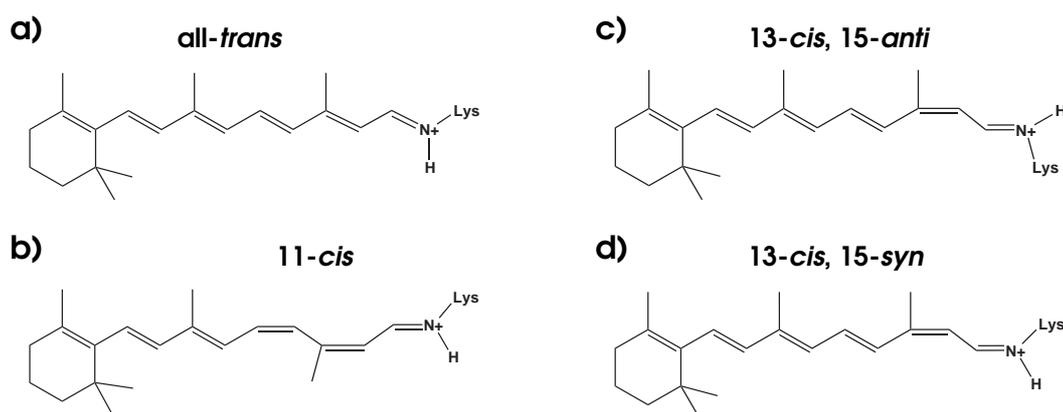


Figure 1.2. Retinal Schiff base conformations. The conjugated π -system of retinal can absorb one photon. This energy is utilized for the functional mechanism. Here, the retinal Schiff base is shown in its protonated form. **a)** In the ground state, before light absorption, archaeal rhodopsins bind all-*trans* retinal via a protonated Schiff base linkage to a lysine residue in the 7th helix. In the visual pigments, all-*trans* retinal occurs after light absorption. **b)** In the ground state, visual pigments bind 11-*cis* retinal. **c)** and **d)** In archaeal rhodopsins, upon light absorption, all-*trans* retinal isomerizes to 13-*cis* retinal. Both 13-*cis*, 15-*anti* and 13-*cis*, 15-*syn* can be observed. In these two conformations the orientation of the Schiff base proton differs. This feature is not easy to resolve by spectroscopic measurements.

ring. The retinal conformations observed in rhodopsins are depicted in Figure 1.2. Light absorption by retinal leads to its isomerization. Subsequent thermal relaxations govern conformational changes of the protein that drive the respective function of the rhodopsins. Rhodopsins perform two cellular key functions: the conversion of light energy into chemical energy and the reception of external stimuli, specifically light signals.

1.1.1 ION PUMPS

Several rhodopsins that function as ion pumps, *i.e.*, converting electromagnetic energy into the chemical energy of an ion gradient, are known. The proton pump proteorhodopsin, for example, was discovered a decade ago in marine plankton as the first bacterial rhodopsin [12–14]. Proteorhodopsin variants are abundant in marine plankton and it is assumed that proteorhodopsin-based phototrophy is globally a significant oceanic microbial process [15, 16]. Proteorhodopsin generates a proton gradient which is utilized to synthesize adenosine triphosphate (ATP), the universal energy storage molecule of the cell.

The bacterial proton pump xanthorhodopsin binds in addition to retinal a second chromophore, the carotenoid salinixanthin that acts as antenna pigment [17]. Light energy absorbed by the carotenoid is transferred to the retinal, thereby extending the wavelength range which the protein can utilize for function. Light-harvesting carotenoids are common in chlorophyll-based photosynthesis, but have previously not been observed for retinal-based ion pumps.

Other rhodopsin ion pumps exist such as a proton pump found in the fungus *Leptosphaeria maculans* [18]. Interesting are also rhodopsin ion channels that have been found in the green alga *Chlamydomonas reinhardtii* [19]. Channelrhodopsin-1 is a proton channel and channelrhodopsin-2 is a cation-selective ion channel that leads to cell depolarization after illumination [19].

1.1.2 PHOTORECEPTORS

Photoreceptors signal the detection of light. The best known rhodopsins are the visual pigments. These photoreceptors are found in the rod and cone cells of the vertebrate eye and signal the absorption of light to the nervous system. In 1967, the Nobel prize in physiology and medicine was rewarded to Ragnar Granit, Haldan K. Hartline and George Wald *for their discoveries concerning the primary physiological and chemical visual processes in the eye*.

Visual pigments belong to the protein family of G protein-coupled receptors (GPCRs) which participate in a multitude of physiological processes [20–22]. GPCRs constitute the largest and most structurally conserved family of signaling proteins. However, the only available high-resolution structures of GPCRs are solved from the rod cell bovine rhodopsin, depicted in Figure 1.1 [7, 23]. Thus, the visual pigments have become the paradigm of GPCRs and are extensively studied [24, 25].

Light signaling in single cell organisms is less complex and restricted to the detection of harmful and useful light conditions. Nevertheless, the ability of cells to detect these conditions and the resulting adaption of their movement, termed phototaxis, in principle constitutes vision. Examples of such phototactic sensory rhodopsins are the *Anabaena* sensory rhodopsin, the first sensory rhodopsin observed in bacteria and chlamyrodopsin found in green algae [26–28].

1.1.3 THE TWO RHODOPSIN PROTEIN FAMILIES

All rhodopsins share common structural features, *i.e.*, seven transmembrane helices and the chromophore retinal (*cf.* Figure 1.1 and 1.2), and they are activated by light absorption. Nevertheless, sequence alignment divides rhodopsins into two distinct protein families without any apparent relationship: the archaeal-type or type I rhodopsins, and the visual rhodopsins or type II rhodopsins [8].

Visual rhodopsins are found in the eyes throughout the animal kingdom, *e.g.*, the visual pigments. Furthermore, melanopsin, neuropsin and other so-called non-classical rhodopsins belong to this family [11, 29]. While it is known that they act as light sensors, the function of these non-classical rhodopsins is not well understood. It is likely that some of these rhodopsins function in a non-visual capacity such as the acute regulation of dermal pigmentation or the light-mediated synchronization of the circadian rhythm to the astronomical day.

The archaeal-type rhodopsins function as light driven ion pumps, *e.g.*, proteorhodopsin and xanthorhodopsin, and phototactic signaling proteins, *e.g.*, *Anabaena* sensory rhodopsin and chlmyrhodopsin. Archaeal-type rhodopsins are found in all three kingdoms of life: archaea, bacteria and eukarya [30, 31]. In the present work archaeal rhodopsins from halophilic archaea are investigated. These archaeal rhodopsins were the first discovered not belonging to the visual rhodopsins.

1.2 HALOPHILIC ARCHAEA

Halophilic archaea are the dominant microorganisms in hypersaline environments with salt concentrations exceeding 3 M, *e.g.*, the Dead Sea or saltern evaporation ponds [32–34]. These archaea adapt to the extreme conditions using the so-called *salt-in* strategy that permits them to grow even in saturated salt solutions of 5.2 M. They accumulate inorganic ions within the cell in concentrations osmotically at least equivalent to the salt concentration of their environment. The cytoplasm of halophilic archaea contains molar concentrations of KCl, while the main salt of the cell environment is commonly NaCl.

Osmotic balance in hypersaline environments can also be obtained by the accumulation of small organic molecules, glycerol, amino acids, sugar derivatives or other compounds in the cytoplasm. This *compatible-solute* strategy, *i.e.*, the accumulation of high concentrations of osmotic solutes, requires no special adaptations of the intracellular components to high salt concentrations.

In cells using the salt-in strategy to prevent osmotic stress, all cellular components are adapted to high salt concentrations. For instance, proteins are among other things protected by a high amount of acidic surface residues. In general, halophilic proteins depend on relatively high salt concentrations for their function. Halophilic archaea cannot grow, if the environmental salt concentration is below 3 M. The optimal salt concentration for the archaeon *Halobacterium salinarum* (*Hb. sal.*) is in the range of 4-5 M. A detailed review of the physiology of halophilic archaea can be found in Schäfer *et al.* [35].

1.3 ARCHAEL RHODOPSINS

The plasma membrane of halophilic archaea can contain specialized patches, the so-called purple membrane. The purple membrane consists of only 25% lipids and as much as 75% proteins in a near crystalline form. In the early 1970s, the purple membrane of *Hb. sal.* was found to contain only a single protein [36]. In reference to its similarity to vertebrate rhodopsins, this protein was named bacteriorhodopsin (BR). The prefix 'bacterio' was chosen since at that time archaea had not been recognized as an independent prokaryotic cell type quite distinct from bacteria. For the same reason the genus of halophilic archaea is called halobacteria. During the investigation of the physiology of *Hb. sal.*, three other archaeal rhodopsins were discovered: halorhodopsin (HR), sensory rhodopsin I (SRI) and sensory rhodopsin II (SRII).

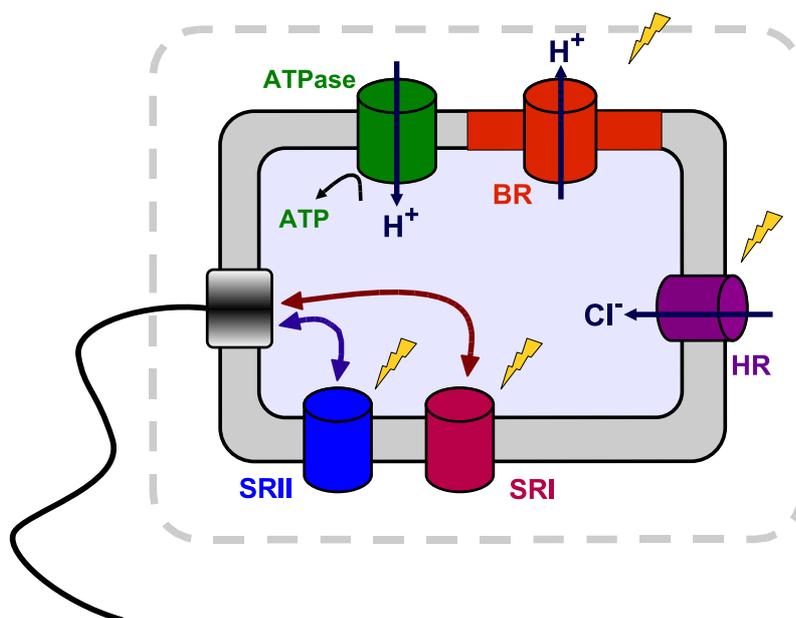


Figure 1.3. The halobacterial cell and archaeal rhodopsins. The proton pump bacteriorhodopsin (BR) generates a proton gradient that can be utilized by the ATPase to synthesize ATP. The chloride pump halorhodopsin (HR) pumps chloride ions into the cell to keep the osmotic pressure intact. The two sensory rhodopsins, SRI and SRII, regulate the phototactic response of the halobacterial cell. The flagellar bundle is represented by one flagellum. The cell wall is indicated by the dashed line.

A schematic representation of the halobacterial cell and its four archaeal rhodopsins is shown in Figure 1.3. BR functions as an outward proton pump, *i.e.*, protons are transferred from the cytoplasm to the extracellular environment. The resulting proton gradient can be utilized by an ATPase. HR is an ion pump like BR. Initially assumed to function as a sodium pump, it was soon shown that HR catalyzes the light-driven uptake of chloride ions [37].

The third discovered archaeal rhodopsin was SRI. It was immediately proposed that SRI functions as a photoreceptor [38]. For the halobacterial cell, SRI acts as an attractant for red light. Discrepancies between the absorption spectrum of SRI and the repellent effect of blue-green light on *Hb. sal.* led to the postulation of an additional photoreceptor and, finally, to the discovery of the fourth archaeal rhodopsin [39, 40]. With respect to its function as a repellent photoreceptor, the protein was initially termed phoborhodopsin, but is now commonly referred to as sensory rhodopsin II (SRII). Upon activation SRI and SRII relay the light signal to their respective transducer protein. In turn, the transducer activates the phosphorylation cascade that modulates the flagellar motor of *Hb. sal.*

Archaeal rhodopsins are relatively small proteins with a molecular mass of approximately 26 kDa. As depicted in the schematic in Figure 1.4, the seven transmembrane helices of archaeal rhodopsins are denoted by the letters A to G. As indicated in Figure 1.4, the chromophore retinal is bound to a lysine residue in the center of helix G protruding into the space between helix C to G, dividing the protein into a cytoplasmic and an

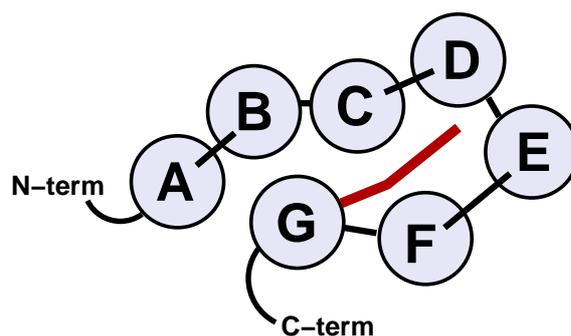


Figure 1.4. Schematic of archaeal rhodopsin viewed from the cytoplasmic side.

The seven transmembrane helices are denoted by the letters A – G. The N-terminus of the proteins protrudes into the extracellular space and the C-terminus into the cytoplasm. The position of the retinal Schiff base in between the helices is indicated in red. The loops connecting the helices are denoted AB-, BC-, CD-, DE-, EF- and FG-loop with respect to the two helices connected by this loop.

extracellular side. In all archaeal rhodopsins, retinal adopts an all-*trans* conformation in the ground state and isomerizes upon light absorption to 13-*cis* retinal (*cf.* Figure 1.2). During the last decade, high-resolution X-ray structures could be determined for BR [41–45], HR [46, 47] and SRII [48–50].

SEQUENCE ALIGNMENT OF BR, HR AND SRII

As shown by sequence homology, the archaeal rhodopsins belong to the same protein subfamily [51]. Figure 1.5 shows the sequence alignment of BR, HR and SRII based on the superposition of the high-resolution structures. The sequence identity of the archaeal rhodopsins ranges from 20% to 35%, but is significantly higher in the transmembrane region. This region is in general better conserved than the loop region. In particular, the residues of the retinal binding pocket which comprises the first shell of amino acids around the chromophore are highly conserved, emphasizing the importance of this region as the active center of the retinal proteins [51, 52]. These residues are highlighted in Figure 1.5.

THE RETINAL BINDING POCKET OF BR, HR AND SRII

The retinal binding pocket of BR, HR and SRII contains 24 residues that are depicted in Figure 1.6. Especially the aromatic residues, *i.e.*, tryptophan, phenylalanine and tyrosine, which define the shape of the pocket and stabilize the retinal in the protein environment are highly conserved (*cf.* Figure 1.5).

The retinal binding pocket of HR differs from that of BR and SRII most notably in the Schiff base region where a threonine residue, *i.e.*, Thr111, replaces the aspartate residue in BR and SRII, *i.e.*, Asp85 and Asp75, respectively. Thr111 is, contrary to the aspartate residue in BR and SRII, uncharged. However, in HR a chloride ion occupies approximately the position of the negatively charged O_{δ} atom of the aspartate residue, resulting in a similar charge distribution in this region for BR, HR and SRII.

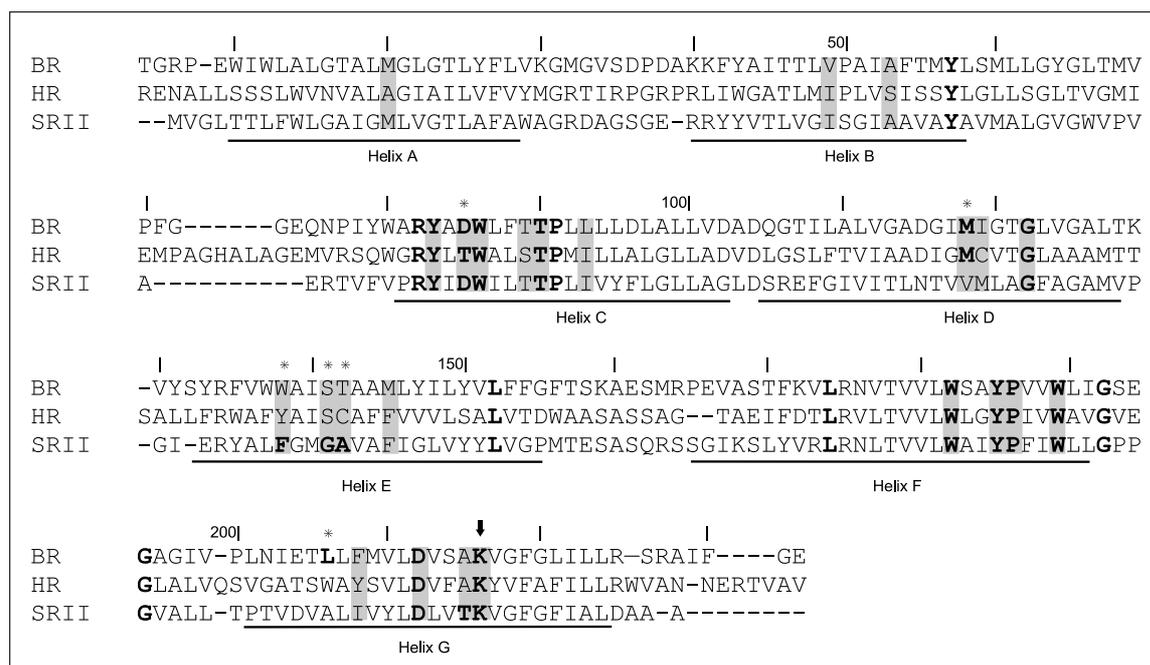


Figure 1.5. Sequence alignment for the archaeal rhodopsins. The alignment is obtained from the structural superposition of BR [41], HR [46], and SRII [48]. The numbering of the residues corresponds to BR. The 24 residues that form the retinal binding pocket are highlighted by gray boxes. The lysine residue in helix G that binds the retinal covalently to the protein is indicated by a black arrow. Residues that are conserved either in BR, HR or SRII are printed boldly and are marked by a star. Residues that are conserved throughout the archaeal rhodopsins are printed boldly. The information regarding the conservation is taken from Ref. [51]. The seven transmembrane helices A to G are indicated by black bars.

Close to the β -ionone ring, a serine residue conserved in BR and HR, *i.e.*, Ser141 and Ser168, respectively, is substituted in SRII by a glycine residue, *i.e.*, Gly130, causing a decrease in polarizability in this region. Above the Schiff base an alanine residue, *i.e.*, Ala215 and Ala241 in BR and HR, respectively, is replaced by a threonine residue, *i.e.*, Thr204, which leads to an increase in polarizability in the Schiff base region of SRII. The polarizability in the ring region of retinal is further lowered in SRII by placing a valine and an alanine residue, *i.e.*, Val108 and Ala131, where BR has a methionine and a threonine residue, *i.e.*, Met118 and Thr142, and HR a methionine and a cysteine residue, *i.e.*, Met144 and Cys169.

1.3.1 THE PROTON PUMP BACTERIORHODOPSIN

The chemiosmotic hypothesis proposed by Peter D. Mitchell in 1961 suggests that energy from redox-reactions, *i.e.*, electron transfer, is utilized to translocate protons across a membrane. In 1978, Mitchell's idea that the resulting electrochemical gradient of protons is the driving force of ATP synthesis has been rewarded with the Nobel prize in

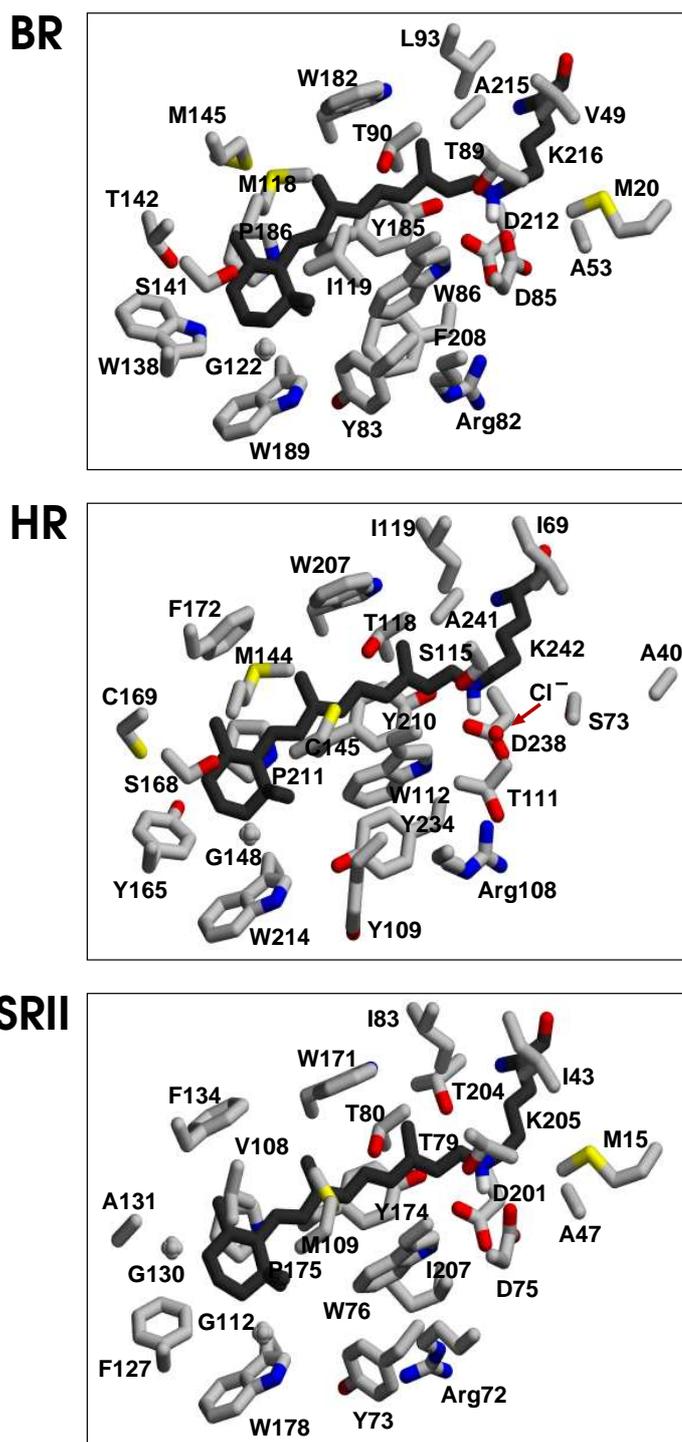


Figure 1.6. Retinal binding pocket. The side chains of the 24 residues that form the retinal binding pocket are depicted for BR, HR and SRII, respectively. These residues are located within 5 Å of the conjugated π -system of the retinal Schiff base. The retinal and the lysine residue which form the Schiff base are depicted in dark grey. Arg82 of BR, Arg108 of HR and Arg72 of SRII are not located in the 5 Å distance around the retinal. However, these arginine residues belong to the complex counterion of the positively charged retinal Schiff base and are, therefore, depicted. The retinal is stabilized by several aromatic ring systems. For example in BR: Tyr83, Trp86, Trp138, Trp182, Tyr185, P186 and W189.

chemistry for his contribution to the understanding of biological energy transfer through the formulation of the chemiosmotic theory.

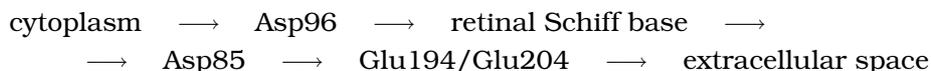
Two primary metabolic processes have been established to be in accordance with the chemiosmotic theory: respiration and chlorophyll-based photosynthesis [53]. With the discovery of BR, a novel and comparatively simple mechanism to generate a proton gradient was found. Light energy is used directly for proton transfer, *i.e.*, this retinal-based photosynthetic mechanism is independent of electron transfer.

BR constitutes the smallest known ion pump and is also one of the few whose structure could be resolved in atomic detail. Its structure and function are reviewed in Refs. [54–61]. The elucidation of its proton transfer mechanism may help to understand the functional mechanism of other ion pumps as well. BR has, thus, become a prototype for ion transport.

PROTON TRANSFER IN BR

Isomerization of retinal in BR initiates a sequence of events, termed photocycle. The initialization of the photocycle is light-dependent, while the following steps constitute thermal relaxations driven by the light energy taken up in the bR state.

The structure of BR is depicted in Figure 1.7. The key residues of the proton transfer are shown explicitly. The spatial order of the proton transfer from the cytoplasmic to the extracellular site is:



The two glutamate residues, Glu194 and Glu204, are part of the proton release group. When protonated, the proton is delocalized over the components of the proton release group, *i.e.*, Glu194, Glu204 and a water cluster between the glutamate residues [62, 63].

The residues Arg82, Asp115 and Asp212 are important for physiological proton transfer. They do, however, not change their protonation during the photocycle and act at most as transient proton acceptors. For example, the proton may be transferred from Asp85 to Glu194/Glu204 via Arg82. Asp115 is protonated under physiological conditions and located relatively far from the proton transfer channel close to the β -ionone ring of the retinal. It is implicated in the back-pressure effect that prevents alkalization of the cytoplasm due to an extreme pH gradient or an extreme membrane potential [64, 65].

Figure 1.8 shows the photocycle of BR which consists of a series of intermediates. The ground state of BR is termed bR. It should be noted that the abbreviation ‘BR’ refers to the protein bacteriorhodopsin, while ‘bR’ characterizes a state of BR. During the photocycle, the absorption maximum of the retinal changes. The different absorption maxima, listed in Table 1.1, allowed to characterize the bR, K, L, M, N and O intermediate state. As can be seen in Table 1.1, the absorption maximum of the M state is significantly blue-shifted. The M state is therefore also termed blue intermediate. The blue-shift is caused by the deprotonation of the retinal Schiff base, *i.e.*, the chromophore is neutral.

The change in absorption provided the first means to distinguish between the intermediate states and to analyze the kinetics of the photocycle. Later on, the intermediates

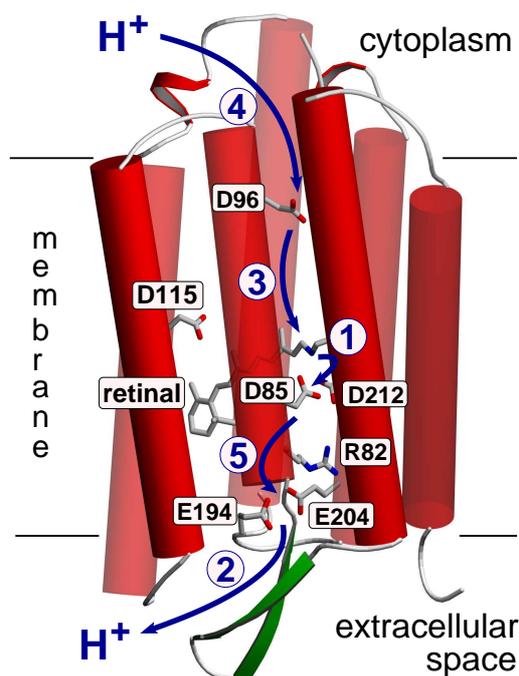


Figure 1.7. Proton transfer steps of bacteriorhodopsin. Light absorption in the ground state, *i.e.*, the bR state, triggers BR's photocycle, during which a proton is transferred from the cytoplasm to the extracellular space. In the BR structure, the key residues of proton transfer are depicted explicitly. The five sequential proton transfer steps are indicated by the arrows and numbers. After light absorption and retinal isomerization from all-*trans* to 13-*cis*, the proton is transferred from the retinal Schiff base to Asp85. Virtually at the same time, another proton is released from the proton release group, *i.e.*, Glu194 and Glu204, to the extracellular side. Then, the retinal Schiff base is protonated from Asp96, which is thereafter protonated from the cytoplasm. Finally, Asp85 deprotonates and the proton release group is protonated.

were characterized also with respect to conformational changes and different protonation of the residues involved in the proton transfer. This led to the separation of some intermediates into subspecies. The M intermediate, for example, is nowadays in general subdivided into the M_1 and M_2 state [66].

During the photocycle, the proton is not transferred from top to bottom, *i.e.*, starting at the cytoplasm. Instead, the proton transfer is initialized at the active center, *i.e.*, the retinal Schiff base region, where the light energy that drives the transfer is absorbed. The main events of the photocycle are listed in the following in chronological order:

- bR: In the ground state of BR, Asp96, the retinal Schiff base and the proton release group are protonated, while Asp85 is deprotonated.
- bR \rightarrow K: Absorption of red light leads to retinal isomerization from all-*trans* to 13-*cis*. The isomerization occurs in approximately 200 fs.

State	bR	K	L	M	N	O
λ_{\max}/nm	570	590	550	410	560	640

Table 1.1. Absorption maxima of the photocycle intermediate states of BR. During the photocycle, the absorption maximum of the retinal changes. This allows to characterize the K, L, M, N and O intermediate state.

- K → L: Localized structural rearrangements occur in the Schiff base region which prepare the proton transfer by spreading the absorbed energy from the retinal to the protein.
- L → M₁: The first transfer step occurs from the retinal Schiff base to Asp85, deprotonating the retinal Schiff base.
- M₁ → M₂: The proton release group, *i.e.*, Glu194 and Glu204, deprotonates. The proton is released to the extracellular space. Under physiological conditions the proton release to the extracellular space occurs virtually simultaneously to the first proton transfer step. In the M intermediate state, large rearrangements of the protein structure occur, *i.e.*, the cytoplasmic part of helix F tilts outward allowing water molecules to enter the opening cytoplasmic channel.
- M₂ → N: The third proton transfer step occurs between Asp96 and the retinal Schiff base. The distance between Asp96 and retinal Schiff base is larger than 10 Å and this region of BR is relatively apolar. The water molecules that entered the cytoplasmic region in the M state may provide a hydrogen bonded chain along which proton transfer can occur.
- N → O: In the fourth step, Asp96 is protonated from the cytoplasm. Thus, one proton is now translocated from the cytoplasm to the extracellular space. Retinal isomerizes back to its all-*trans* conformation.
- O → bR: To return to the ground state, the last proton transfer step translocates the proton from Asp85 to the proton release group. The cycle is now complete.

The kinetics of the transitions between the photocycle intermediates is highly complex and the above described photocycle is a simplified description. Despite extensive studies, the kinetics of the photocycle and the classification of the intermediates is still discussed [67–74]. Although the key residues of proton transfer and the sequence of transfer events have been known for a long time, open questions as to the exact proton transfer pathway remain [75–78]. The sequence of events described here constitutes a generally accepted working model for proton transfer under physiological conditions.

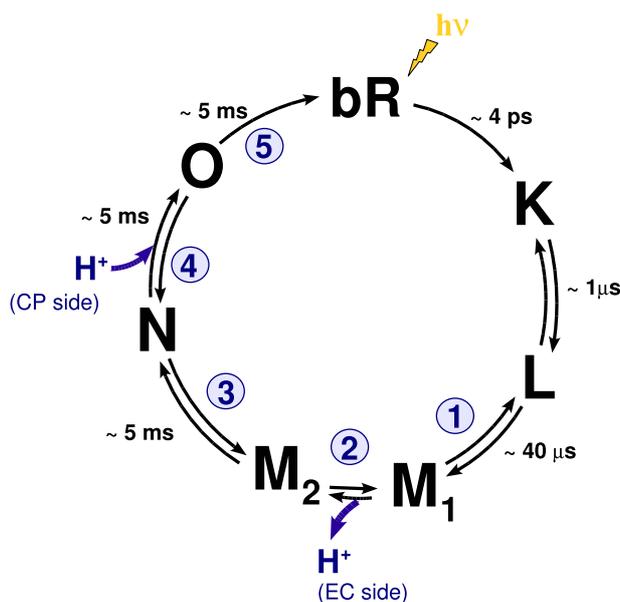


Figure 1.8. The photocycle of bacteriorhodopsin. The photocycle intermediate states, in sequential order: K, L, M₁, M₂, N and O, arise from thermal relaxation steps during which the light energy is utilized for proton transfer. The time gaps between the intermediates are denoted. The five proton transfer steps depicted in Figure 1.7 are indicated by the blue numbers. The proton release to the extracellular side during the transition from the M₁ to the M₂ state is indicated, as well as the proton uptake from the cytoplasm during the transition from the N to the O state.

1.3.2 THE CHLORIDE PUMP HALORHODOPSIN

To cope with the osmotic stress inherent to high salt concentrations, halophilic archaea maintain isomolar intracellular salt concentrations [79]. Since the membrane potential is negative at the cytoplasmic side of the membrane, passive influx of cations is possible, whereas anions require active transport. HR, that primarily mediates the import of chloride ions into the cell, but also of bromide, iodide and nitrate, plays an important role in the net salt uptake. HR is the only known light-driven anion pump [80–82].

The adaption to anion transport is reflected in the structure of HR. The aspartate residue Asp85 of BR is replaced by a threonine residue. This mutation creates a binding site for the negatively charged chloride ion, where in BR one of the O _{δ} atoms of Asp85 is located (*cf.* Figure 1.6 and 1.9). The photocycle of HR is similar to that of BR with regard to the isomerization of retinal and the general conformational changes of the protein. However, no changes in the protonation state of HR are observed under physiological conditions [83, 84]. Since the blue shift of the absorption maximum is caused by the deprotonation of the retinal Schiff base, no M state or blue intermediate exists for HR.

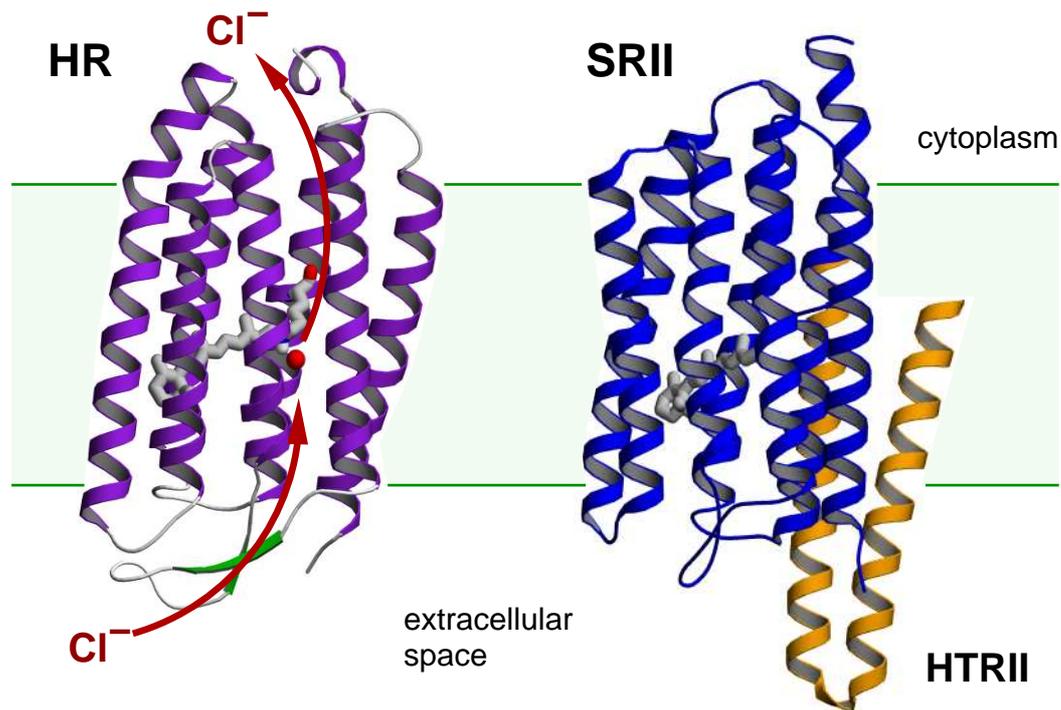


Figure 1.9. Halorhodopsin and sensory rhodopsin II. The structures of HR and SRII are depicted [46, 50]. The chromophore retinal is shown explicitly. For HR, the chloride in its binding site on the extracellular side of the retinal Schiff base is shown in red. For SRII, the resolved two helices of its transducer protein HtrII are shown in orange. The membrane is indicated by the green slab.

1.3.3 THE PHOTORECEPTOR SENSORY RHODOPSIN II

The repellent receptor SRII is typical for sunlight avoidance receptors: its absorption maximum of approximately 500 nm is close to the energy peak in the solar radiation spectrum at the surface of the earth. The activation of SRII by sunlight generates a signal that leads the halobacterial cell away from the harmful near ultraviolet light. SRII is complexed with a membrane-embedded transducer termed halobacterial transducer for SRII (HtrII) as depicted in Figure 1.9. Upon receptor activation, HtrII modulates a cytoplasmic phosphorylation cascade controlling the flagellar motor [85, 86].

The photocycle of SRII resembles that of BR. However, in the native membrane, no net-transport of protons can be detected, instead SRII circulates protons in its extracellular channel [87]. In the second half of the photocycle large conformational changes occur, in particular the cytoplasmic side of archaeal rhodopsins opens up due to helix angle tilting. In BR, this mechanism is believed to allow water molecules to enter the cytoplasmic side of the protein permitting proton transport through this apolar region. It is suggested that helix tilting causes signal transduction by means of helix-helix interaction of SRII with HtrII, invoking the same mechanism used by BR for ion-transport for a different purpose. Hence, a general functional mechanism for all archaeal rhodopsins is proposed [88, 89].

1.3.4 CHANGE OF FUNCTION FOR BR, HR AND SRII

The functional mechanism of the archaeal rhodopsins BR, HR and SRII is proposed to be very similar [8, 88–92]. It is possible to transform the function of BR, HR and SRII, respectively, into that of another archaeal rhodopsin. For example, the Asp85Thr mutant of BR functions as a chloride pump [93, 94]. It could be shown that this mutation creates a binding site for chloride equivalently to the chloride binding site in HR. The Asp85Thr mutant actively translocates chloride ions from the extracellular space into the cytoplasm. Thus, the interaction of the retinal Schiff base with either an aspartate or a threonine residue determines the ionic specificity of BR. This indicates that BR and HR share a common transport mechanism.

For the chloride pump HR, azide-dependent proton pumping can be observed [83, 84]. An azide binding to the chloride binding site below the Schiff base and another azide shuttling between the Schiff base and the cytoplasmic surface, fulfill the function of Asp85 and Asp96, in BR (*cf.* Figure 1.7 and 1.9). This azide-binding results in active proton transport from the cytoplasm to the extracellular space. Evidently, HR contains all structural requirements of a proton pump with the exception of an internal proton acceptor and donor.

During the physiological photocycle of the photoreceptor SRII, the retinal Schiff base deprotonates. However, if SRII and its transducer HtrII are dimerized, no proton translocation across the membrane can be observed. Instead, SRII cycles protons in its extracellular channel [95]. When not associated with its transducer, SRII was shown to pump protons, indicating that key elements of the proton transport mechanism of BR have been conserved in SRII [85].

1.4 THE OPSIN SHIFT

One of the most notable properties of the chromophore retinal among the various chromophores utilized by photoreceptors is the extreme variance of its absorption maximum. In polar solution, a protonated retinal Schiff base has an absorption maximum of 440 nm [96]. In rhodopsins, the protein environment tunes the absorption maximum of retinal over a wide range from the ultra violet to the far red [97, 98]. The magnitude of the shift in absorption is known as the opsin shift. The opsin shift describes the shift between the absorption maximum of a protonated retinal Schiff base in polar solution and in the protein environment of the rhodopsins [99]. The name originates from the term for the apoprotein: opsin. The exact regulation of the absorption maxima by the opsins over such a wide range enables organisms not only to detect light, but to discriminate between different light conditions. Thus, light of specific wavelengths can be used for signaling which is the underlying mechanism of color vision.

Human color vision is based on three light-sensitive pigments that correspond to one of the three colors: blue, green or red. The absorption spectra of these three pigments are shown in Figure 1.10. Blue rhodopsin absorbs maximally at 420 nm, green rhodopsin at 530 nm and red rhodopsin at 570 nm. The ability to discriminate between a seemingly infinite number of colors is based on the additive response of all three receptor types.

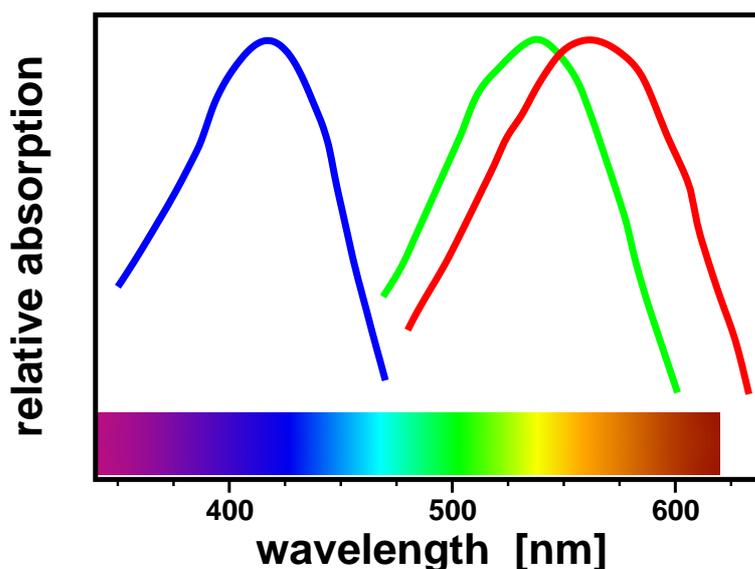


Figure 1.10. Absorption spectra of human color pigments. The graph shows the absorption spectra of the blue, green and red rhodopsin. From the intensity of the response of one pigment to light and the additive response of two or three pigments, the color is determined. The visible light of the electromagnetic spectrum is indicated.

Variants of the proton pump proteorhodopsin found in marine bacteria differ in the magnitude of the opsin shift. The different absorption maxima of proteorhodopsin allow these bacteria to utilize light of different wavelengths for energy conversion. The bacteria can grow in habitats with varying light conditions [100].

Halophilic archaea are able to differentiate between harmful, high-energy blue light and useful red light whose energy is utilized by BR and HR for ion transport. Two photoreceptors, SRI and SRII, regulate the motility pattern of the cells to guide them into regions optimal for growth. As shown in Figure 1.11, the absorption spectrum of the repellent photoreceptor for blue light, SRII, is significantly blue-shifted compared to that of BR and HR.

Light absorption by a chromophore corresponds to the absorption of a photon by an electron of this molecule. The physical basis behind the absorption of light is electronic excitation. The energy difference between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) corresponds to the absorption maximum. In the ground state S_0 , the chromophore absorbs a photon and an electron is promoted from the HOMO to the LUMO, *i.e.*, the chromophore is in the first excited state S_1 . The energetically favored promotion of an electron will always be between S_0 and S_1 . Thus, this energy difference determines the absorption maximum of a chromophore and the photon energy corresponds to the S_0 - S_1 transition energy or excitation energy [104].

Spectral tuning in rhodopsins has been studied for decades both experimentally and theoretically [105–112]. However, despite these efforts, the structural basis of the tuning mechanism is only partly understood. The protein matrix can influence the excitation energy of a chromophore by interacting preferably with the S_0 -state, thereby increasing the

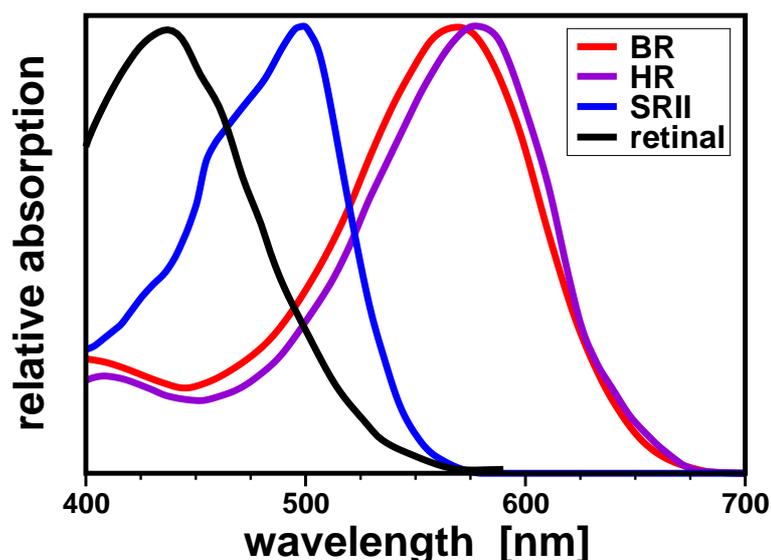


Figure 1.11. Absorption spectra of archaeal rhodopsins. Experimental absorption spectra of BR, HR and SRII [87, 101, 102]. The black line shows the absorption spectrum of the protonated retinal Schiff base in methanol solution [103].

S_0 - S_1 -transition energy, or preferably with the S_1 -state, thereby stabilizing the S_1 -state and lowering the S_0 - S_1 -transition energy. BR, HR and SRII lower the S_0 - S_1 -transition energy. The absorption maximum of their retinal chromophore is, therefore, red-shifted to light of lower energy compared to the absorption maximum of 440 nm of the protonated retinal Schiff base in polar solution. The opsin-shift of BR and HR is, however, much higher than that of SRII.

1.5 OUTLINE OF THE THESIS

This thesis aims at advancing the understanding of rhodopsin proteins by theoretical calculations. Since high-resolution X-ray structures suitable for theoretical studies are available, the archaeal rhodopsins, BR, HR and SRII, have been chosen as model systems. These proteins are representatives of both ion pumps and photoreceptors and aspects of both functional mechanisms, *i.e.*, ion transport and light sensing, are analyzed.

All calculations performed in this work are based on continuum electrostatics given by the Poisson-Boltzmann equation. Chapter 2 introduces the theoretical framework and the numerical methods used in this thesis.

Chapter 3 concentrates on the high-resolution X-ray structures of the archaeal rhodopsins. The preparation of these structures for the performed electrostatic calculations are described and the system parameters used, *e.g.*, atomic partial charges, are given.

An analysis of the absorption shift of archaeal rhodopsins is presented in Chapter 4. First, the electrostatic potential that BR, HR and SRII cause at the retinal is compared. Thereafter, the potential is decomposed to assign the observed differences to individual

protein residues. A generalized model of a quantum mechanical particle in a box is used to qualitatively describe the absorption maxima in dependence of the electrostatic potential.

In Chapter 5, the probability of functionally relevant protonation states of multiple BR structures is presented. Furthermore, the protonation behavior and the correlation of the key residues of proton transfer is analyzed.

Chapter 6 introduces a novel algorithm, termed extended dead-end elimination (X-DEE), which is effective in generating gap-free lists of lowest energy states. X-DEE is applicable to various systems. In this work, X-DEE is implemented for protonation state calculations. A first application is presented in Chapter 7, where gap-free lists of protonation states of BR are analyzed.

A conclusion with respect to the presented results is given at the end of the Chapters 4 to 7. In Chapter 8, a general conclusion of this work is drawn. An outlook is given with respect to further investigations and new applications of the methods used in this thesis.

CHAPTER 2

CONTINUUM ELECTROSTATICS

Truth is much too complicated to allow anything but approximations.

John von Neumann

Electrostatic interactions play a central role in the stabilization and function of biomolecules. They are involved in virtually all biological processes such as molecular recognition, ion transport and enzyme catalysis [53, 113–116]. Electrostatics also have a major impact in absorption processes of light. Polar and charged groups constitute a main part of the building blocks of biological macromolecules. In addition, they are often complexed with ions, *e.g.*, Mg^{2+} ions stabilize the DNA and RNA macro-ions and Ca^{2+} ions are involved in signaling cascades. An understanding of electrostatic interactions is, thus, of major interest in structural biology.

In this thesis a theoretical approach is used to analyze the electrostatics of archaeal rhodopsins. Mathematical models and computer calculations can complement experimental research when direct measurements are either impractical or impossible. To successfully describe the system under investigation a suitable model has to be formulated based on physical laws and translated into computer language. In the next section the steps involved in building a computer model will briefly be discussed.

The main part of the chapter presents the conceptual, mathematical and computational models employed in this work. Continuum electrostatics based on the Poisson-Boltzmann equation is used to model the protein-membrane-solvent system of the archaeal rhodopsins. Here, the underlying theory and the numerical solution of the Poisson-Boltzmann equation are outlined. Subsequently, the calculation of protonation equilibria from electrostatic potentials is discussed. Last, the Metropolis Monte Carlo algorithm is explained.

2.1 BUILDING COMPUTER MODELS

Exact mathematical descriptions of processes in Nature seldom exist. Instead, the emphasis in formulating models lies on developing useful approximations that allow to gain

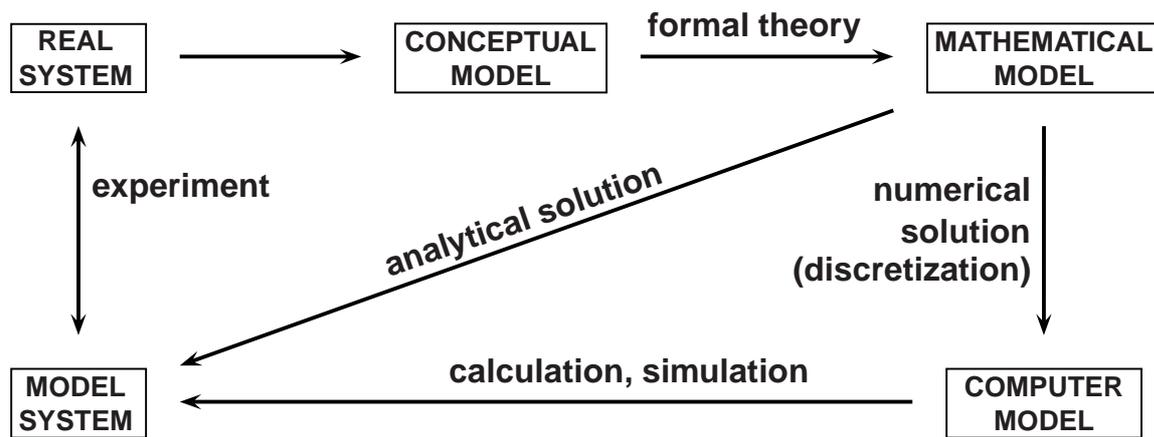


Figure 2.1. Computer models. Theory as well as experimental research utilizes models to explain observed phenomena with scientific laws. This diagram shows the main steps towards a computer model of a given system. Computational calculations and simulations give insight into the system under investigation.

insight into the described process. As Samuel Karlin stated, ‘*the purpose of models is not to fit the data but to sharpen the questions*’.

The diagram in Figure 2.1 shows the principal steps of building computational models. The first and crucial challenge lies in the formulation of a conceptual model that translates the real system into the idealized model system. For this purpose, the problem should be clearly posed, *i.e.*, the first step is to ask ‘What is the Question?’ The model should describe the phenomena under investigation, reproduce already obtained knowledge, *e.g.*, experimental data, and allow further predictions for the system under observation. The conceptual model should be as complicated as necessary, but also as simple as possible.

The conceptual model provides a qualitative understanding of the real system. To obtain a quantitative picture, formal physical theory is applied to translate the conceptual into a mathematical model. The same principles as for the formulation of the conceptual model, *e.g.*, available data should be reproduced and it should be possible to predict the outcome of new experiments, apply also to the mathematical model. The level of theory should be chosen with respect to the size of the system, the time-span to be observed and the phenomenon to be analyzed.

An analytical solution of the mathematical model is generally feasible only for few and simple cases. More complex applications require a discretization of the mathematical model leading to the computer model. The resulting calculated data is verified with available experimental data whenever possible and analyzed with respect to phenomena that are as yet and may never be accessible to experimental research.

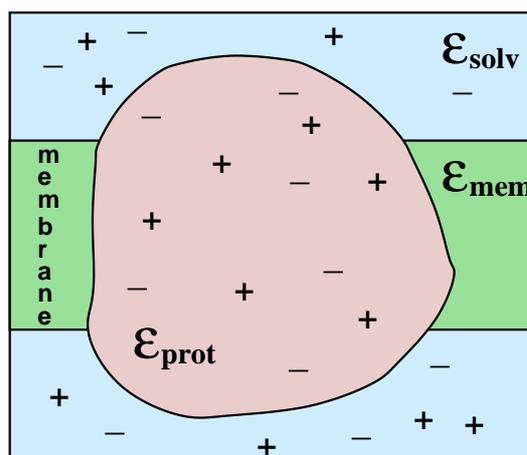


Figure 2.2. Continuum electrostatics. The protein is modeled as a region with low permittivity ϵ_{prot} containing fixed point charges that represent the partial charges of the protein atoms. The aqueous solution is represented by a region with a high permittivity ϵ_{solv} and mobile charges (dissolved ions). The apolar part of the membrane is modeled by a low permittivity ϵ_{mem} which is in general equal to ϵ_{prot} .

2.2 CONCEPTUAL MODEL

The theoretical methods that are employed to model the electrostatics in biochemical systems can be roughly classified into two categories: (i) explicit solvent methods that describe all atoms of the system explicitly, including the solvent and dissolved ions, and (ii) implicit solvent methods which use a continuum model to describe the solvent [117, 118]. Implicit solvent models approximate the average influence of the solvent on the solute. While using a macroscopic description of the solvent, solutes can be described in atomic detail. Here, an implicit solvent method is used to describe the electrostatics of a complex protein-membrane system in aqueous solution. A schematic representation of the continuum model is shown in Figure 2.2.

The protein is surrounded by solvent, *i.e.*, water and ions, and is embedded into a membrane. Polarization effects are represented by the permittivity that describes the influence of an electric field on a dielectric medium and *vice versa*. It is determined by the ability of the medium to polarize in response to the field, thereby reducing the total electric field inside the medium. Polarizability is the relative tendency of a charge distribution to be distorted from its normal shape by an external electric field. Thus, the permittivity accounts for the electrostatic screening of charges in different media. The protein is described with a low permittivity, representing its low polarizability, with fixed point charges that represent its atomic partial charges. The aqueous phase is described with a high permittivity, representing its high polarizability, and contains mobile point charges, *i.e.*, dissolved ions. The membrane is described with a low permittivity, representing its apolar phase.

More elaborate continuum models of membranes exist where slabs of different dielectric continua represent the hydrophilic and the hydrophobic parts of the membrane [119].

However, when the effects to be studied are not close to the membrane-protein interface, the continuum model described here is well justified.

2.3 ELECTROSTATIC POTENTIAL IN DIELECTRIC MEDIA

The electrostatic potential of biological molecules in dielectric media is routinely calculated using the Poisson-Boltzmann equation [120]. This equation combines two fundamental physical equations: The Poisson equation which provides an expression for the electrostatic potential and the Boltzmann equation which describes the distribution of particles in response to a field, *e.g.*, the mean probability of finding an ion at a point in space in an electrostatic field. In the following subsections, the Poisson-Boltzmann equation and its numerical solution are introduced. A general introduction to classical electrostatics and an in-depth description can be found in Ref. [121].

2.3.1 THE POISSON-BOLTZMANN EQUATION

The simplest case of classical electrostatics is given when considering charges in vacuum. The electrostatic potential in vacuum can be expressed by the Poisson equation which defines the electrostatic potential $\phi(\mathbf{r})$ originating from the charge density $\rho(\mathbf{r})$ at a spacial point \mathbf{r} :

$$\nabla\nabla\phi(\mathbf{r}) = -\frac{4\pi}{\varepsilon_0} \rho(\mathbf{r}) \quad , \quad (2.1)$$

where ∇ is the gradient operator with respect to the spatial coordinates, $\rho(\mathbf{r})$ is the charge density and ε_0 is the permittivity of vacuum.

Due to the relative permittivity of dielectric media, the electrostatic interactions among charges in a uniform medium of gas, liquid or solid are usually weakened compared to the interactions of the same charges in vacuum. If a dielectric medium is introduced into an electrostatic field, this field induces a dipole moment in the atoms or molecules placed into it. The electrostatic field \mathbf{E} is given by the negative gradient of the potential:

$$\mathbf{E}(\mathbf{r}) = -\nabla\phi(\mathbf{r}) \quad . \quad (2.2)$$

If the molecule has a net dipole moment, the electrostatic field increases the dipole moment and the dipole aligns with the field. The electric field generated by these induced dipoles is directed opposite to the inducing field. As a consequence, the overall field is weakened. This property of the external field can be described implicitly by reducing the magnitude of the electrostatic potential by a constant factor known as the relative permittivity of the medium ε_r . The Poisson equation takes the form:

$$\nabla\nabla\phi(\mathbf{r}) = -\frac{4\pi}{\varepsilon_0 \varepsilon_r} \rho(\mathbf{r}) \quad . \quad (2.3)$$

In a homogeneous dielectric medium, the potentials are then $1/\varepsilon_r$ of the corresponding potentials given by Eq. (2.1) in vacuum for the same charge distribution.

The relatively apolar protein in aqueous, *i.e.*, polar, solution is not well described by an uniform permittivity. To account for the differences in the dielectric properties of the protein and its environment, a dielectric displacement vector \mathbf{D} is introduced:

$$\mathbf{D}(\mathbf{r}) = \varepsilon(\mathbf{r}) \mathbf{E}(\mathbf{r}) \quad , \quad (2.4)$$

where $\varepsilon(\mathbf{r})$ is the spatially varying dielectric constant given by $\varepsilon(\mathbf{r}) = \varepsilon_0 \varepsilon_{\mathbf{r}}(\mathbf{r})$. Together with Eq. (2.2) and Eq. (2.4), the Poisson equation for a medium with a spatially varying dielectric constant $\varepsilon(\mathbf{r})$ is obtained:

$$\nabla(\varepsilon(\mathbf{r}) \nabla\phi(\mathbf{r})) = -4\pi \rho(\mathbf{r}) \quad . \quad (2.5)$$

The low dielectric constant of the protein region $\varepsilon_{\text{prot}}$ is typically set between 2 and 4 to take into account the electronic polarization and the orientational polarizability of a semi-rigid molecule, while the solvent (water) has a high dielectric constant, usually $\varepsilon_{\text{solv}} = 80$, which is due to the electrostatic field generated by the large permanent dipole moment of the water molecules and to their high flexibility in the liquid phase [122, 123]. Thus, the electrostatic interactions are more shielded in aqueous solution than inside a protein.

The charge density $\rho(\mathbf{r})$ of the system has two contributions, the fixed charges of the protein $\rho_{\text{prot}}(\mathbf{r})$ and the mobile charges of the solvent $\rho_{\text{solv}}(\mathbf{r})$:

$$\rho(\mathbf{r}) = \rho_{\text{prot}}(\mathbf{r}) + \rho_{\text{solv}}(\mathbf{r}) \quad . \quad (2.6)$$

The mean distribution of the mobile solvent ions in the potential can be described by a Boltzmann distribution:

$$\rho_{\text{solv}}(\mathbf{r}) = \sum_{i=1}^K N_A c_i^{\text{bulk}} Z_i e_o \exp\left(-\frac{Z_i e_o \phi(\mathbf{r})}{k_B T}\right) \quad , \quad (2.7)$$

where K is the number of ion species in solution, N_A is the Avogadro number, c_i^{bulk} is the concentration of the ion species, Z_i is the formal charge of the ion, e_o is the elementary charge, k_B is the Boltzmann constant and T the temperature. When substituting Eq. (2.6) and Eq. (2.7) into Eq. (2.5), the Poisson-Boltzmann equation is obtained:

$$\nabla(\varepsilon(\mathbf{r}) \nabla\phi(\mathbf{r})) = -4\pi \left(\rho_{\text{prot}}(\mathbf{r}) + \sum_{i=1}^K N_A c_i^{\text{bulk}} Z_i e_o \exp\left(-\frac{Z_i e_o \phi(\mathbf{r})}{k_B T}\right) \right) \quad . \quad (2.8)$$

A major limitation of the Poisson-Boltzmann approach arises from the mean-field approximation of the interaction of dissolved ions. Furthermore, ions are treated as point-charges. Nevertheless, when being aware of these limitations, the Poisson-Boltzmann equation offers a suitable description of the electrostatic potential in protein-membrane-solvent systems.

For a wide variety of problems, the solution of Eq. (2.8) can be well approximated by the solution of its linearized form. Consequently, the linearized form of the Poisson-Boltzmann equation is often used for calculations [124].

2.3.2 THE LINEARIZED POISSON-BOLTZMANN EQUATION

The Poisson-Boltzmann equation, derived in the previous section, is non-linear. Thus, fundamental properties of linear systems, *e.g.*, the superposition principle, do not apply. The principle of superposition states that a linear combination of solutions to the system is again a solution to the same linear system. The superposition principle applies to linear systems of algebraic equations, linear differential equations, or systems of linear differential equations. The linearization of a non-linear system may, therefore, facilitate the analysis of the system.

The Poisson-Boltzmann equation can be linearized by employing Taylor expansion of the exponential function up to the linear term:

$$\sum_{i=1}^K N_A c_i^{\text{bulk}} Z_i e_o \exp\left(-\frac{Z_i e_o \phi(\mathbf{r})}{k_B T}\right) \cong \sum_{i=1}^K N_A c_i^{\text{bulk}} Z_i e_o - \sum_{i=1}^K N_A c_i^{\text{bulk}} Z_i^2 e_o^2 \frac{\phi(\mathbf{r})}{k_B T} \quad (2.9)$$

The first term of the Taylor expansion vanishes because the solvent is electroneutral, *i.e.*, the sum of all ionic charges in solution is assumed to be zero. The terms following the linear term are neglected for small potentials, *i.e.*, for $e_o \phi(\mathbf{r})/k_B T \ll 1$, which is in general true for protein systems. The linearized Poisson-Boltzmann equation is obtained as:

$$\nabla(\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r})) = -4\pi \left(\rho_{\text{prot}}(\mathbf{r}) - \sum_{i=1}^K N_A c_i^{\text{bulk}} Z_i^2 e_o^2 \frac{\phi(\mathbf{r})}{k_B T} \right) \quad (2.10)$$

It is convenient to define the ionic strength of the solution as:

$$I = \frac{1}{2} \sum_{i=1}^K N_A c_i^{\text{bulk}} Z_i^2 \quad (2.11)$$

The ionic strength I of a solution is a function of the concentration of all ions present in a solution. I is proportional to the square of the Debye parameter $\bar{\kappa}$ which characterizes the shielding due to mobile charges:

$$\bar{\kappa}^2 = \frac{8\pi e_o^2 I}{k_B T} \quad (2.12)$$

The ionic strength I and $\bar{\kappa}$ are parameters of the ionic solution and consequently zero in the protein region. Therefore, the linearized Poisson-Boltzmann equation can be written as:

$$\nabla(\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r})) = -4\pi \left(\rho_{\text{prot}}(\mathbf{r}) - 2e_o^2 \frac{\phi(\mathbf{r})}{k_B T} I \right) = -4\pi \rho_{\text{prot}}(\mathbf{r}) + \bar{\kappa}^2 \phi(\mathbf{r}) \quad (2.13)$$

For the linearized Poisson-Boltzmann equation the superposition principle applies as long as the dielectric boundaries of the molecule remain unchanged. The superposition principle states that the total electrostatic potential of a system of charges is the sum of the potentials of the individual charges. Thus, the electrostatic potential of a system with N charges is the superposition of the potentials induced by the individual charges.

2.3.3 THE FINITE DIFFERENCE METHOD

Analytical solutions of the linearized Poisson-Boltzmann equation exist only for simple, symmetric geometries such as spherical bodies. Proteins, however, are usually of irregular shape and, thus, discretization methods are used to transfer the equation from a continuous partial differential equation into discrete problems. The linearized Poisson-Boltzmann equation can be solved numerically for example using finite elements or finite differences. In the following the finite difference solution of the linearized Poisson-Boltzmann equation is described. In finite differences, the atomic partial charges of the molecule, the permittivity and the ionic strength are discretized on a grid. In numerical analysis, finite differences are one of the simplest and most powerful means to approximate a differential operator, and are extensively used to solve differential equations.

Figure 2.3 shows a flowchart illustrating the steps of the numerical solution of the linearized Poisson-Boltzmann equation:

- (i) To calculate the electrostatic potential the solvent accessible surface has to be calculated from the atomic coordinates by rolling a probe sphere over the van der Waals radius of the atoms. Usually, the probe sphere radius is taken to be 1.4 Å, representing a water molecule. The solvent accessible surface allows to differentiate between solvent and protein region, thereby, ϵ_{solv} and ϵ_{prot} can be assigned. The ion exclusion layer is derived by extending the atomic radii by a value representing the ion radius. The region inside the ion exclusion layer is ion-free and consequently $\bar{\kappa}$ is zero in this region.
- (ii) In finite differences, the space is subdivided into a lattice. The atomic point charges are assigned to the grid points of the box they occupy, *i.e.*, to the eight nearest grid points, by a triangular weighting scheme:

$$q_{\text{grid}} = q(1 - a/h)(1 - b/h)(1 - c/h) \quad , \quad (2.14)$$

where q is the charge, h is the grid spacing, a , b and c are the distances between charge q and the grid point in x , y , z -direction, and q_{grid} is the resulting grid charge. The so-called grid artifact arises from the interaction of the split charges with each other. However, when energy differences are calculated, the grid artifact is canceled. The electrostatic potential and mobile charges, represented by the parameter $\bar{\kappa}$, are also mapped to the grid points and the permittivity is defined on the grid lines. The dielectric constant is set according to whether the grid line is in the protein or in the solvent region. Grid lines that cross the protein-solvent barrier, *i.e.*, the solvent accessible surface, are assigned an intermediate dielectric constant.

- (iii) The size of the initial grid should be chosen such that an analytic solution gives a good approximation of the boundary potential, *i.e.*, the boundary of the grid should be in the solvent region and relatively distant from any solutes. Grid points at the boundary of the grid are special in that they have fewer neighboring grid points. Boundary conditions, *e.g.*, analytical solutions of the potential using for instance the Debye-Hückel theory, are assigned to these grid points in the initial step of the calculation.

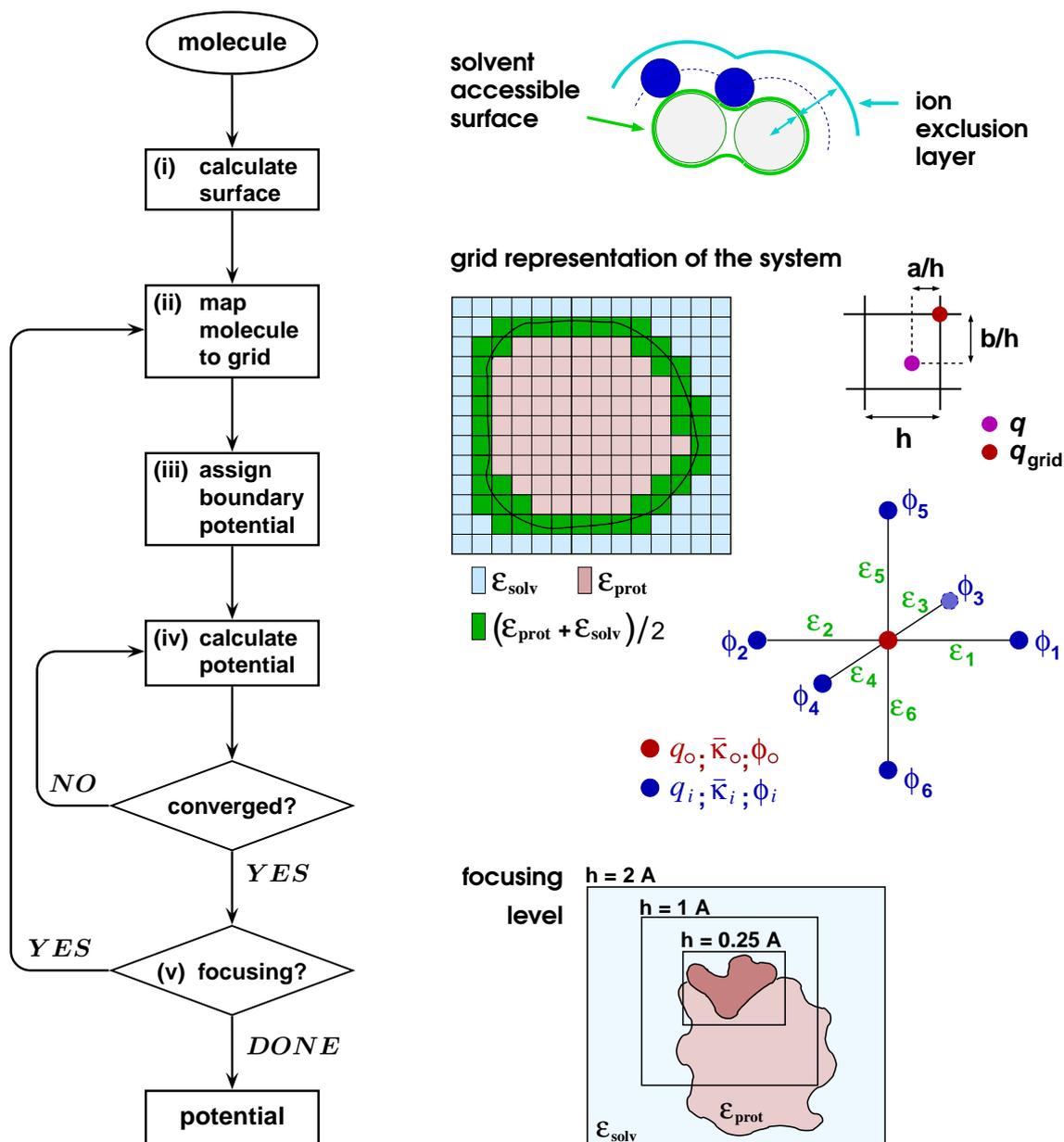


Figure 2.3. Calculating the electrostatic potential. (i) From the atom coordinates of the molecule, the solvent accessible surface is calculated by rolling a probe sphere over the van der Waals surface of the atoms. Usually, the radius of the probe sphere is 1.4 \AA representing a water molecule. Extending the atomic radii of the solute atoms by a value representing the ion radius, the ion exclusion layer is derived. (ii) A grid with spacing h is superposed on the solvent-solute system. The ionic strength, represented by $\bar{\kappa}$ (cf. Eq. (2.12)) is assigned to the grid points outside the ion exclusion layer. Charges q are assigned to the grid points using Eq. (2.14), a , b and c denote the distance between q and a grid point in x , y and z direction, respectively. The dielectric constant $\epsilon(\mathbf{r})$ is assigned to the grid lines. The boundary between molecule and solvent is smoothed by using an average dielectric constant in this region. (iii) The boundary potential is derived using an analytical solution, e.g., Debye-Hückel theory. (iv) The electrostatic potential ϕ is calculated at the grid points from the linearized Poisson-Boltzmann equation (cf. Eq. (2.13)) by using finite differences (cf. Eq. (2.15)). The calculation is repeated until self-consistency, i.e., until the potentials at all grid points have converged. (v) Several focusing levels with decreasing grid spacing and size may be used.

(iv) The electrostatic potential at all other grid points is given by:

$$\phi_o = \frac{\left(\sum_{i=1}^6 h \varepsilon_i \phi_i \right) + 4\pi q_o}{\left(\sum_{i=1}^6 h \varepsilon_i \right) + h^3 \bar{\kappa}_o^2}, \quad (2.15)$$

where ϕ_o is the potential at the grid point in question. The sum is taken over the 6 neighboring grid points. Eq. (2.15) is solved in an iterative cycle until self-consistency is reached. A finite difference scheme is said to be convergent, if all of its solutions in response to initial conditions and excitations converge point wise to the corresponding solutions of the original differential equation as the grid spacing approaches zero. The potential at any off-grid location are obtained by interpolation.

(v) The accuracy of the finite difference method depends on the grid spacing h . In principle, the smaller the grid spacing, the more accurate the calculated potentials will be, but the higher the computational cost. By focusing, *i.e.*, using increasingly higher resolved grids which are focused on the region of interest, the computational cost can be kept relatively low, while allowing to calculate the potentials with great accuracy. The boundary conditions of the first, coarse grid have to be assigned using for instance the Debye-Hückel theory (step (iii)). From the second focusing level onwards, the boundary potential can be initialized from the electrostatic potentials calculated using the coarser grid. Often, not only the grid spacing decreases with focusing level, but also the grid size, thereby further increasing the computational efficiency.

2.4 PROTONATION EQUILIBRIA IN PROTEINS

Acid-base titration is the deprotonation process of a protonatable site with increasing pH. Proteins contain protonatable amino acids: glutamate, aspartate, cysteine and tyrosine are acidic and arginine, lysine and histidine are basic. Consequently, the charge distribution of a protein depends on pH. In particular, for proteins whose function is associated with charge transfer processes, an understanding of their protonation behavior in dependence of pH is of importance. In the following the theory of calculating protonation equilibria in proteins is outlined [125, 126].

In general the protonation probability of an independent protonatable site in aqueous solution can be derived from the classical Henderson-Hasselbalch equation:

$$\text{pH} = \text{p}K_a + \log \frac{[A^-]}{[HA]} \quad \text{as} \quad \langle x \rangle = \frac{\exp(-\ln 10(\text{pH} - \text{p}K_a))}{1 + \exp(-\ln 10(\text{pH} - \text{p}K_a))}, \quad (2.16)$$

where $[HA]$ and $[A^-]$ are the concentration of the protonated and deprotonated form of the protonatable site, respectively; and $\langle x \rangle$ is the protonation probability of this site. The $\text{p}K_a$

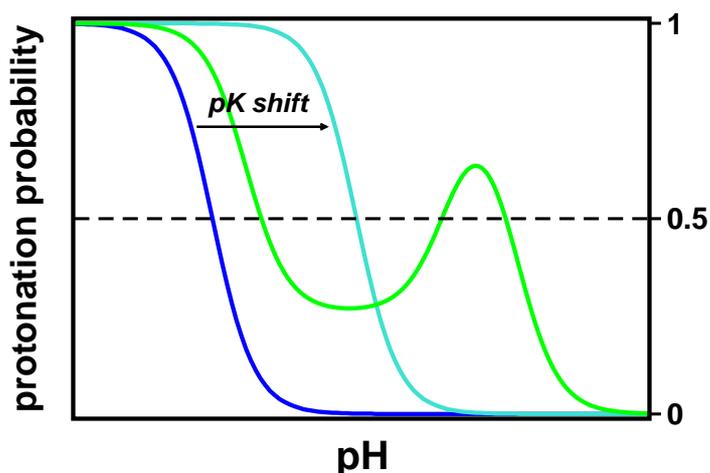


Figure 2.4. Exemplary titration curves. The blue curve shows the typical sigmoid titration of a model compound in solution. Upon transferring the site from aqueous solution into a protein environment the change in solvation and background energy (cf. Eq. (2.18) and Eq. (2.19), respectively) may shift the pK_a value (cyan curve). The shifted pK_a value is defined as pK_{intr} . Interaction of a protonatable site in a protein with other protonatable sites may lead to complex, non-sigmoid titration behavior. An example of such a non-sigmoid titration behavior shows the central nitrogen atom of DTPA (green curve) [127]. The dashed line indicates a protonation probability of 0.5, commonly described as pK_a value. Obviously, for a group with complex protonation behavior as shown by the green curve a single pK_a value cannot be defined.

value of an isolated site is defined as the pH value at which the protonation probability of this site is 0.5. Figure 2.4 shows the typical sigmoid shape of the titration curve. As can be seen, the pK_a value is sufficient to describe the titration of an isolated site in aqueous solution.

The behavior of a protonatable site in a protein may be considerably more complex and can often not be described by a single pK_a value. To calculate protonation equilibria in proteins, model compounds are defined that represent the protonatable amino acids in aqueous solution. The pK_a value of the model compound in aqueous solution, pK_{mod} , is usually known from experiment [128] or derived from high level quantum mechanical calculations [129, 130].

As depicted in Figure 2.5, three factors change the protonation behavior of a protonatable site in a protein compared to that of the model compound:

1. The change in the dielectric environment: water is significantly more polar than the protein environment and the electric field induced by the site in aqueous solution is stronger and, thus, the overall field is weakened compared to the protein environment. In other words, the screening of charges is stronger in aqueous solution than in the protein. The difference in interaction with the dielectric environment is defined as the Born solvation free energy $\Delta\Delta G_{Born}$.
2. Interaction of the site with the so-called background charges of the protein environment: the background energy contribution $\Delta\Delta G_{back}$ accounts for the interaction of

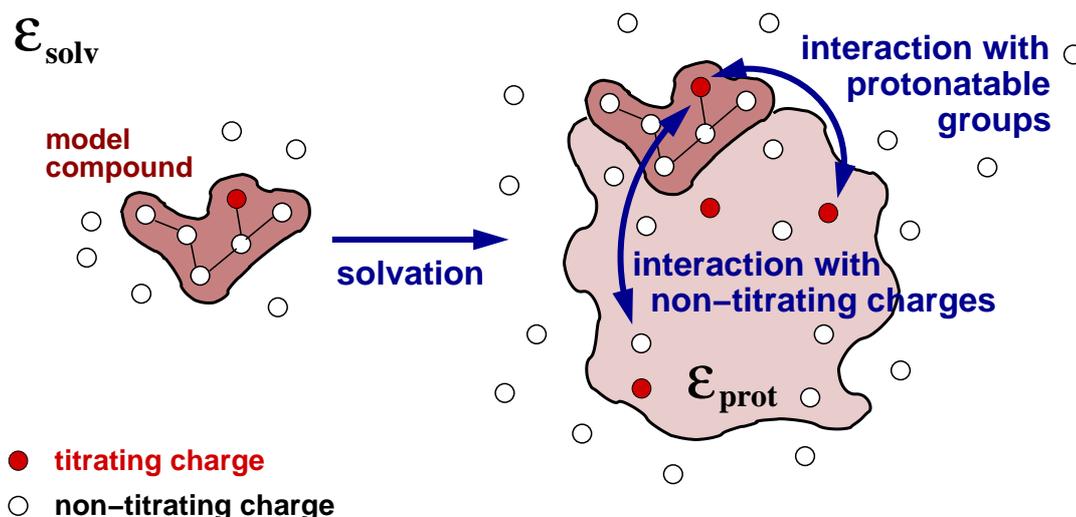


Figure 2.5. Titration in a protein environment. The model compound is the protonatable site in solution represented by ϵ_{solv} . The protonation behavior of such a model compound can be either experimentally determined or calculated from high level quantum chemical calculations. The behavior of a protonatable site in a protein environment (represented by ϵ_{prot} , ϵ_{solv} , partial charges of the protein atoms and ions in solution) compared to the model compound is influenced by the change in solvation energy (Eq. (2.18)), the interaction with non-titrating charges of the protein and ions in solution (Eq. (2.19)) and the interaction with other protonatable sites in the protein (Eq. (2.20)).

the model compound charges with the background charges in the protein. The background charges are the partial charges of atoms that are independent of pH and the charges of other protonatable groups in their so-called reference protonation form.

3. The pairwise interaction energy between the site and other protonatable sites in the protein $W_{\mu\nu}$.

The Born solvation and the background energy may lead to a shifted $\text{p}K_{\text{a}}$ value which is defined as the intrinsic $\text{p}K_{\text{a}}$ value: $\text{p}K_{\text{intr}}$. The $\text{p}K_{\text{intr}}$ value can be calculated from electrostatic potentials and hence from the Poisson-Boltzmann equation as shown in the next subsection. The interaction energy between two protonatable sites $W_{\mu\nu}$ may lead to non-sigmoid titration behavior as shown in Figure 2.4. The calculation of $W_{\mu\nu}$ from electrostatic potentials is outlined in Subsection 2.4.2.

2.4.1 THE INTRINSIC $\text{p}K_{\text{a}}$ VALUE

The intrinsic $\text{p}K_{\text{a}}$ value, $\text{p}K_{\text{intr}}$, of a protonatable site in a protein environment is defined as the $\text{p}K_{\text{a}}$ value this site would have, when all other protonatable sites are fixed in their neutral state:

$$\text{p}K_{\text{intr}} = \text{p}K_{\text{mod}} - \frac{\Delta \Delta G_{\text{Born}} + \Delta \Delta G_{\text{back}}}{\ln 10 RT}, \quad (2.17)$$

where pK_{mod} is the pK_a value of the protonatable site in aqueous solution (model compound), R is the gas constant and T the temperature. The Born solvation energy $\Delta\Delta G_{\text{Born}}$ arises from the interaction of the partial charges $Q_{i,\mu}$ of the protonatable site μ with its reaction field, *i.e.*, the polarization $Q_{i,\mu}$ induces in the dielectric environment:

$$\begin{aligned} \Delta\Delta G_{\text{Born}} = & \frac{1}{2} \sum_{i=1}^N Q_{i,\mu}^{\text{p}} \left[\phi_{\text{prot}}(\mathbf{r}_i, Q_{\mu}^{\text{p}}) - \phi_{\text{mod}}(\mathbf{r}_i, Q_{\mu}^{\text{p}}) \right] \\ & - \frac{1}{2} \sum_{i=1}^N Q_{i,\mu}^{\text{u}} \left[\phi_{\text{prot}}(\mathbf{r}_i, Q_{\mu}^{\text{u}}) - \phi_{\text{mod}}(\mathbf{r}_i, Q_{\mu}^{\text{u}}) \right] . \end{aligned} \quad (2.18)$$

The sum is taken over the N atoms of site μ that have different charges in the protonated and deprotonated form, as denoted by the superscripts p and u, respectively. The electrostatic potentials ϕ are calculated at the position \mathbf{r}_i of atom i . The dielectric boundary was either derived from the protein or the model compound as denoted by the subscripts prot and mod, respectively. The solution of the linearized Poisson-Boltzmann equation was then obtained assigning either the charges of the protonated form of site μ (Q_{μ}^{p}) or the deprotonated form (Q_{μ}^{u}).

The background energy $\Delta\Delta G_{\text{back}}$ is due to the interaction of the partial charges $Q_{i,\mu}$ of the protonatable site μ with the partial charges q_j of non-titrating sites and with the neutral form of all other protonatable residues in the proteins:

$$\begin{aligned} \Delta\Delta G_{\text{back}} = & \sum_{j=1}^N q_j \left[\phi_{\text{prot}}(\mathbf{r}_i, Q_{\mu}^{\text{p}}) - \phi_{\text{prot}}(\mathbf{r}_i, Q_{\mu}^{\text{u}}) \right] \\ & - \sum_{j=1}^N q_j \left[\phi_{\text{mod}}(\mathbf{r}_i, Q_{\mu}^{\text{p}}) - \phi_{\text{mod}}(\mathbf{r}_i, Q_{\mu}^{\text{u}}) \right] . \end{aligned} \quad (2.19)$$

The first summation is over the N charges of the protein that belong to atoms of non-titrating sites and to atoms of other protonatable residues in their neutral form. The second summation is over the N charges of the protonatable site μ that are independent of the protonation form of μ .

The Born solvation energy and the background energy can shift the pK_a value as expressed by the pK_{intr} value of a protonatable site in a protein, but their contribution cannot lead to non-sigmoid titration behavior.

2.4.2 THE INTERACTION ENERGY $W_{\mu\nu}$

In order to calculate the full ionization behavior of a protein as a function of pH, the interaction between the protonatable sites has to be taken into account. The interaction energy between a pair of protonatable sites μ and ν is given by:

$$W_{\mu\nu} = \sum_{i=1}^N \left[Q_{i,\mu}^{\text{p}} - Q_{i,\mu}^{\text{u}} \right] \left[\phi_{\text{prot}}(\mathbf{r}_i, Q_{\nu}^{\text{p}}) - \phi_{\text{prot}}(\mathbf{r}_i, Q_{\nu}^{\text{u}}) \right] . \quad (2.20)$$

The sum is taken over all atomic partial charges of site μ that differ in their charge in the protonated and deprotonated form as denoted by the superscripts p and u, respectively. The electrostatic potentials at position r_i of atom i is calculated in dependence on the charge distribution of the protonatable site ν in its protonated and deprotonated form, respectively. $W_{\mu\nu}$ is the additional energy required to add a proton to site μ due to a proton on site ν . The self-interaction energy of a site is assumed to be zero.

From $\text{p}K_{\text{intr}}$ and $W_{\mu\nu}$ the full protonation behavior of a protonatable site in a protein can be calculated. As depicted in Figure 2.4, the interaction energy between protonatable sites in a protein can lead to non-sigmoid titration behavior.

2.4.3 THE FREE ENERGY OF A PROTONATION STATE

Assuming that each protonatable site has two possible protonation forms, protonated and deprotonated, the total number of protonation states for a protein with N protonatable sites is 2^N . Each protonation state n of a system with N protonatable sites can be expressed as a vector $\vec{x}_n = (x_{1,n}, \dots, x_{\mu,n}, \dots, x_{N,n})$, where $x_{\mu,n}$ denotes the protonation form of site μ in protonation state n .

The free energy of a protonation state G_n of a protein with N protonatable sites depends on the protonation form of each protonatable site x_μ and its $\text{p}K_{\text{intr}}$ value, on the interaction between each pair of protonatable sites $W_{\mu\nu}$ and on pH:

$$G_n = RT \ln 10 \sum_{\mu=1}^N \left((x_{\mu,n} - x_\mu^\circ) (\text{pH} - \text{p}K_{\text{intr},\mu}) \right) + \frac{1}{2} \sum_{\mu=1}^N \sum_{\nu=1}^N \left(W_{\mu\nu} (x_{\mu,n} - x_\mu^\circ) (x_{\nu,n} - x_\nu^\circ) \right) \quad (2.21)$$

The first term of Eq.(2.21) represents the intrinsic protonation energy and the second term the interaction energy between the pair of sites μ and ν . The value of $x_{\mu,n}$ is 1 if site μ is protonated and 0 if site μ is deprotonated in protonation state n . The reference protonation form of the site μ , *i.e.*, its neutral form, is given by x_μ° , where $x_\mu^\circ = 1$ for acids and $x_\mu^\circ = 0$ for bases. For site ν , the protonation form $x_{\nu,n}$ and the reference protonation form x_ν° is defined equivalently.

2.5 CALCULATION OF PROTONATION PROBABILITIES

In the previous section, the factors determining the protonation of groups in a protein environment were discussed. This section focuses on the calculation of protonation probabilities. First, the protonation probability for an individual site in proteins is derived. Then, the correlation between the protonation of two protonatable sites is discussed. Protonation probabilities and correlation are exemplified by calculations done on the small molecule DTPA. Lastly, the protonation probabilities for subsets of protonation states is outlined. These probabilities, *i.e.*, thermodynamic properties, are derived from the Boltzmann distribution of states.

Obtaining the Boltzmann distribution requires, in principle, the calculation of all possible protonation state energies G_n of the system. However, already for protein systems with relatively few protonatable sites N , the calculation of all possible 2^N protonation states is seldom feasible. Commonly, the Metropolis Monte Carlo method is employed to approximate the Boltzmann distribution of states as described in the last subsection of this Chapter, Subsection 2.5.5.

2.5.1 PROTONATION PROBABILITY OF A SINGLE SITE

The protonation probability $\langle x_\mu \rangle$ of the site μ at a certain pH value is given by a thermodynamic average over all possible protonation states of the protein. The partition function Z of a system with 2^N states is derived as the sum over all Boltzmann factors:

$$Z = \sum_{n=1}^{2^N} \exp(-G_n/RT) \quad , \quad (2.22)$$

where G_n is the free energy of the protonation state n as defined by Eq. (2.21) in the previous subsection. Using the partition function to normalize the Boltzmann factor of state n , the probability P_n for this state is derived:

$$P_n = \frac{1}{Z} \exp(-G_n/RT) \quad . \quad (2.23)$$

The total of these probabilities P_n is referred to as the Boltzmann distribution. The protonation probability $\langle x_\mu \rangle$ of site μ can then be expressed in terms of the Boltzmann distribution:

$$\langle x_\mu \rangle = \sum_{n=1}^{2^N} x_{\mu,n} P_n = \frac{1}{Z} \sum_{n=1}^{2^N} x_{\mu,n} \exp(-G_n/RT) \quad , \quad (2.24)$$

where $x_{\mu,n}$ is 1 or 0 depending upon site μ being protonated or deprotonated in the protonation state n . A summand becomes zero if x_μ is deprotonated.

Plotting the protonation probability $\langle x_\mu \rangle$, *i.e.*, the expected value for x_μ , against pH, the titration curve of site μ is obtained. Due to the interaction of the protonatable sites with each other and because the charges of these sites are themselves pH dependent, titration curves of individual sites in proteins may be non-sigmoid. An example of a complex titration curve is given in Figure 2.4.

2.5.2 CORRELATION BETWEEN TWO SITES

In statistics, the correlation is a measure for the strength and direction of a linear relationship between two random variables. Well known and widely used is the Pearson product-moment correlation coefficient [131]. This correlation coefficient is obtained by dividing the covariance of the two variables by the product of their standard deviations. Accordingly, the correlation coefficient $c_{\mu\nu}$ of the protonation of two sites μ and ν is given

by:

$$c_{\mu\nu} = \frac{\text{cov}(x_\mu, x_\nu)}{\sigma_\mu \sigma_\nu} , \quad (2.25)$$

where $\text{cov}(x_\mu, x_\nu)$ is the covariance of x_μ and x_ν ; σ_μ and σ_ν are the standard derivations of x_μ and x_ν , respectively. The covariance, a measure of how much two variables change together, can be expressed as:

$$\text{cov}(x_\mu, x_\nu) = \langle x_\mu x_\nu \rangle - \langle x_\mu \rangle \langle x_\nu \rangle , \quad (2.26)$$

where $\langle x_\mu \rangle$, and $\langle x_\nu \rangle$ are given by Eq. (2.24). The probability $\langle x_\mu x_\nu \rangle$ of the two sites μ and ν both being protonated can be calculated analogously by replacing x_μ in Eq. (2.24) by $x_\mu x_\nu$:

$$\langle x_\mu x_\nu \rangle = \sum_{n=1}^{2^N} x_{\mu,n} x_{\nu,n} P_n . \quad (2.27)$$

A summand equals zero if one or both sites are deprotonated.

The range of the covariance for two protonatable sites can be derived as follows: In Eq. (2.26), $\langle x_\mu \rangle$, $\langle x_\nu \rangle$ and $\langle x_\mu x_\nu \rangle$ are substituted by Eq. (2.24) and Eq. (2.27), respectively. Letting $W_{\mu\nu}$ (cf. Eq. (2.21) and Eq. (2.23)) approach positive and negative infinity, it can be shown that the maximum value for the covariance is 0.25 and the minimum value is -0.25.

In the special case that the two variables are identical, *i.e.*, $x_\nu = x_\mu$:

$$\text{cov}(x_\mu, x_\mu) = \langle x_\mu x_\mu \rangle - \langle x_\mu \rangle \langle x_\mu \rangle = \langle x_\mu \rangle - \langle x_\mu \rangle^2 = \sigma_\mu^2 \quad \text{for } x_\mu = 1 \text{ or } 0 . \quad (2.28)$$

Thus, the covariance reduces to the variance σ_μ^2 . The square root of the variance is known as the standard deviation σ_μ . The standard deviation is a measure of the dispersion of the data around its mean. The correlation coefficient $c_{\mu\nu}$ of the protonation of two sites μ and ν can then be determined by:

$$c_{\mu\nu} = \frac{\text{cov}(x_\mu, x_\nu)}{\sigma_\mu \sigma_\nu} = \frac{\langle x_\mu x_\nu \rangle - \langle x_\mu \rangle \langle x_\nu \rangle}{\sqrt{(\langle x_\mu \rangle - \langle x_\mu \rangle^2)(\langle x_\nu \rangle - \langle x_\nu \rangle^2)}} . \quad (2.29)$$

The correlation coefficient $c_{\mu\nu}$ takes values between -1 and 1. Positive correlation, *i.e.*, $c_{\mu\nu} > 0$, indicates that the protonation of site μ favors the protonation of site ν . The behavior of the two groups is similar. Negative correlation, *i.e.*, $c_{\mu\nu} < 0$ indicates that the protonation of site μ disfavors the protonation of site ν . The behavior of the two groups is contrary to each other. A correlation coefficient $c_{\mu\nu}$ of 0 indicates that the protonation of the two sites is uncorrelated, *i.e.*, their behavior is independent of each other. Correlation between two random variables is a necessary but not sufficient condition for causation.

2.5.3 EXAMPLE: CORRELATION IN DTPA

Diethylene triamine pentaacetic acid (DTPA) is an elongated version of ethylene diamine tetraacetic acid (EDTA) and can chelate heavy metal ions, *e.g.*, iron or plutonium. The chemical structure of DTPA is depicted in Figure 2.6 a. DTPA contains three protonatable nitrogen atoms. In the structure, the two terminal nitrogen atoms, N_t , are highlighted in blue and the central nitrogen atom, N_c , is highlighted in green. In the following, N_t denotes both of the terminal nitrogen atoms, N_{t1} and N_{t2} . N_{tx} denotes one specific terminal nitrogen atom, N_{t1} or N_{t2} . In the considered pH range between 2 and 12, the five carboxyl groups of DTPA are completely deprotonated.

POPULATION OF STATES

With respect to the protonation of the three nitrogen atoms, DTPA can assume $2^3 = 8$ protonation states. The eight protonation states are characterized by the vector $\vec{x} = (x_{N_{t1}}, x_{N_c}, x_{N_{t2}})$, where x can be 1 or 0, specifying whether N_c , N_{t1} and N_{t2} are protonated or deprotonated. All state energies G_n have been calculated from Eq. (2.21) and the partition function Z has been derived from Eq. (2.22). The probability for each of these states is calculated from Eq. (2.23).

In Figure 2.6 b, the population of DTPA's protonation states is plotted against pH. At pH = 2, all three nitrogen atoms are protonated and only the state $(1_{N_{t1}}, 1_{N_c}, 1_{N_{t2}})$ is populated. From pH > 2 to pH = 6.7, DTPA gradually loses one proton and the three states where two protons are bound to DTPA become populated: $(1_{N_{t1}}, 0_{N_c}, 1_{N_{t2}})$, $(1_{N_{t1}}, 1_{N_c}, 0_{N_{t2}})$, and $(0_{N_{t1}}, 1_{N_c}, 1_{N_{t2}})$. These three states have a maximum probability at pH = 6.7, where the probability is 0.73 for $(1_{N_{t1}}, 0_{N_c}, 1_{N_{t2}})$, and 0.13 for $(1_{N_{t1}}, 1_{N_c}, 0_{N_{t2}})$ and $(0_{N_{t1}}, 1_{N_c}, 1_{N_{t2}})$, respectively. Thus, the probability that both N_{t1} and N_{t2} are protonated and N_c is deprotonated is significantly higher than the probability that N_c and only one of the terminal nitrogen atoms is protonated.

At pH > 7, DTPA gradually loses another proton and the three states, where only one proton is bound to DTPA, become populated, namely: $(0_{N_{t1}}, 1_{N_c}, 0_{N_{t2}})$, $(1_{N_{t1}}, 0_{N_c}, 0_{N_{t2}})$, and $(0_{N_{t1}}, 0_{N_c}, 1_{N_{t2}})$. These three states have their maximum probability at pH = 9.5, where the probability is 0.6 for $(0_{N_{t1}}, 1_{N_c}, 0_{N_{t2}})$, and 0.06 for both $(1_{N_{t1}}, 0_{N_c}, 0_{N_{t2}})$ and $(0_{N_{t1}}, 0_{N_c}, 1_{N_{t2}})$. Thus, the probability that N_c is protonated and both N_{t1} and N_{t2} are deprotonated is significantly higher than the probability that either of the terminal nitrogen atoms is protonated.

At pH > 8, the protonation probability of all three nitrogen atoms decreases rapidly and the state $(0_{N_{t1}}, 0_{N_c}, 0_{N_{t2}})$ becomes populated, where all three nitrogen atoms are deprotonated. Finally at pH = 12, this state is exclusively populated.

PROTONATION PROBABILITY

The protonation probabilities for N_c and N_{tx} have been calculated from Eq. (2.24). The probability for two nitrogen atoms to be protonated at the same time has been derived from Eq. (2.27). In Figure 2.6 c, the protonation probabilities are plotted against pH.

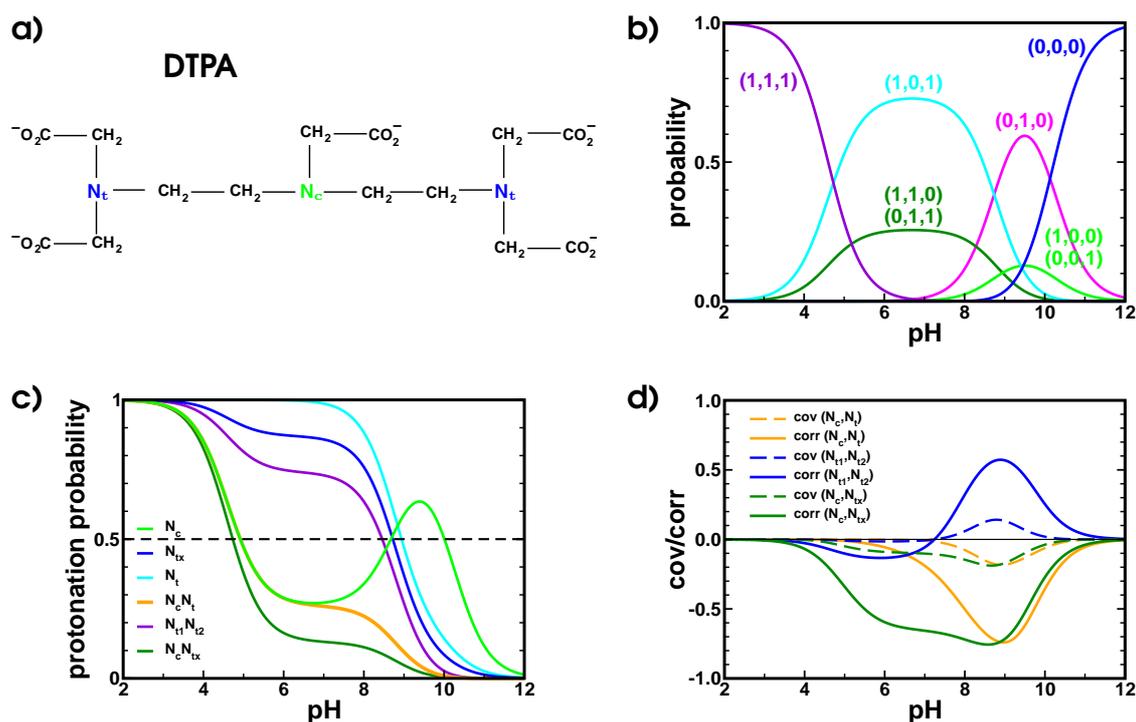


Figure 2.6. Correlation in DTPA. **a)** Structure of DTPA. The molecule has three protonatable nitrogen atoms. The two terminal nitrogen atoms, N_t , and the central nitrogen atom, N_c , are labeled in blue and green, respectively. **b)** The protonation states of DTPA. The probability of all 8 protonation states is plotted against pH. For each state, the state vector $\vec{x} = (x_{N_{t1}}, x_{N_c}, x_{N_{t2}})$ is given, where 1 and 0 indicate whether the respective nitrogen atom is protonated or deprotonated. **c)** The protonation probability is plotted against pH: (i) for the central nitrogen atom, N_c , and (ii) one terminal nitrogen atom, N_{tx} . Additionally, the probability is plotted that: (iii) either or both of the terminal nitrogen atoms are protonated, N_t , (iv) the central and at least one of the terminal nitrogen atoms are protonated at the same time, $N_c N_t$, (v) both terminal nitrogen atoms are protonated, $N_{t1} N_{t2}$, and (vi) the central and one specific terminal nitrogen atom are protonated at the same time, $N_c N_{tx}$. **d)** The covariance and the correlation between the central nitrogen atom and the terminal nitrogen atoms, N_c and N_t , between the two terminal nitrogen atoms, N_{t1} and N_{t2} , and between the central and one specific terminal nitrogen, N_c and N_{tx} , are plotted against pH. The line at 0 separates positive and negative covariance and correlation.

At pH=2, all nitrogen atoms of DTPA are protonated. (i) With increasing pH, N_c deprotonates, and at pH=6.7, the protonation probability of N_c reaches its first minimum of 0.27. When the pH value increases further, the protonation probability of N_c increases again up to about 0.64 at pH=9.4. Thereafter, N_c deprotonates completely. (ii) The probability curve of N_{tx} (N_{t1} or N_{t2}) is not as unusual. When the pH value increases above 2, N_{tx} slowly deprotonates. Between pH 5.5 and pH 7, the protonation probability stagmates at about 0.85. Thereafter, N_{tx} deprotonates completely. (iii) The probability that N_{t1} or N_{t2} or both are protonated, N_t , is greater than 0.99 at $\text{pH} \leq 6.7$. Thereafter, the probability rapidly decreases to 0, resulting in a sigmoid curve. (iv) Given that either N_{t1} or N_{t2} or both are protonated at $\text{pH} < 6.5$, the probability that the central and one of the terminal nitrogen atoms are protonated at the same time, $N_c N_t$, depends in this pH

range on the protonation probability of N_c . Thus, at $\text{pH} < 6.5$ the progression of these two curves is the same. At $\text{pH} > 6.5$, the probability decreases and is equal to 0 at $\text{pH} = 10$. (v) At $\text{pH} > 3$, the probability that both N_{t1} and N_{t2} are protonated at the same time is obviously less than the probability that N_{t1} or N_{t2} or both are protonated. It is also less than the protonation probability of N_{tx} (one specific terminal nitrogen atom), though the shapes of the curves resemble each other. (vi) The probability that the central and one specific terminal nitrogen atom are protonated at the same time, $N_c N_{tx}$, is smaller than the probability that the central nitrogen atom and either of the terminal nitrogen atoms are protonated. The curve progression is, however, alike.

COVARIANCE AND CORRELATION

The covariance and the correlation have been calculated from Eq. (2.26) and Eq. (2.29), respectively: All parameters of these equations are plotted in Figure 2.6 c. Figure 2.6 d depicts the covariance and the correlation between N_{t1} and N_{t2} , between N_c and N_t , *i.e.*, between the central nitrogen atoms and both of the terminal nitrogen atoms, and also between N_c and N_{tx} , *i.e.*, between the central and one specific terminal nitrogen atom.

N_{t1} and N_{t2} are slightly negatively correlated in the pH range from 3 to 7.2, but show a significant positive correlation in the pH range from 7.2 to 12 with a maximum correlation of nearly 0.6 at $\text{pH} \approx 9$. A positive correlation indicates that N_{t1} and N_{t2} behave similarly in this pH range.

N_c and N_t are negatively correlated in the pH range from 4 to 12 with a minimum correlation of -0.75 at $\text{pH} = 9$. A negative correlation indicates that the behavior of N_c and N_t is dissimilar. That is, the protonation of N_c disfavors the protonation of N_t and *vice versa*.

N_c and N_{tx} are negatively correlated in the pH range from 2 to 12. At $\text{pH} < 8$, the negative correlation between N_c and N_{tx} is significantly more pronounced than that between N_c and N_t . In the pH range from 8 to 11, the correlation between N_c and N_{tx} closely resembles that between N_c and N_t .

In Figure 2.6 d, it can be seen that the correlation is the normalized covariance and the curve progression of the covariance closely resembles that of the correlation. In the special case presented here, the covariance takes values between -0.25 and 0.25. This is due to the fact that the protonation behavior of two sites is compared, which can only be 1 (protonated) or 0 (deprotonated).

DISCUSSION

From the population of protonation states, it is obvious that symmetric states are significantly favored over asymmetric states (*cf.* Figure 2.6 b). When two protons are bound to DTPA, mainly the state $(1_{N_{t1}}, 0_{N_c}, 1_{N_{t2}})$ is populated. Thus, N_c is deprotonated, while both N_{t1} and N_{t2} are protonated. When only one proton is bound to DTPA, mainly the state $(0_{N_{t1}}, 1_{N_c}, 0_{N_{t2}})$ is populated. Thus, N_c is protonated, while both N_{t1} and N_{t2} are deprotonated.

This protonation behavior of DTPA coincides with the unusual protonation probability curve of N_c (*cf.* Figure 2.6 c). Though N_c is partly deprotonated in the pH range from 6 to 8, its protonation probability increases significantly at higher pH values. At $\text{pH} = 9.4$ a

maximum is reached, after which N_c deprotonates. Thus, in the pH range where DTPA binds two protons, the protonation probability of N_{t1} and N_{t2} is significantly higher than that of N_c . While in the pH range where DTPA binds only one proton, the protonation probability of N_c is higher than that of N_{t1} and N_{t2} .

In the pH range from 3 to 7.2, the correlation between N_{t1} and N_{t2} is slightly negative (cf. Figure 2.6 d). The fact that in this pH range N_{t1} and N_{t2} are more likely to be protonated at the same time (in the symmetric state), does not effect the correlation behavior (cf. Figure 2.6 b). Both N_{t1} and N_{t2} , individually, have a high protonation probability and also the probability of N_{t1} and N_{t2} being protonated at the same time is high (cf. Figure 2.6 c). The protonation of both N_{t1} and N_{t2} in the symmetric state ($1_{N_{t1}}, 0_{N_c}, 1_{N_{t2}}$) is, therefore, not correlated. Rather in the asymmetric states, N_{t1} and N_{t2} compete against each other for the proton. Thus, the protonation of N_{t1} disfavors the protonation of N_{t2} and *vice versa* and the correlation between N_{t1} and N_{t2} is – if only slightly – negative.

The positive correlation between N_{t1} and N_{t2} in the pH range from 7.2 to 12 is due to the population of ($0_{N_{t1}}, 1_{N_c}, 0_{N_{t2}}$) (cf. Figure 2.6 b and 2.6 d). That is, the deprotonation of N_{t1} favors the deprotonation of N_{t2} and *vice versa*. In this pH range, the negative correlation between N_c and the terminal nitrogen atoms N_t is caused by the same effect. Since the symmetric state ($0_{N_{t1}}, 1_{N_c}, 0_{N_{t2}}$) is preferred, the protonation of N_c disfavors the protonation of N_{t1} and N_{t2} .

At $\text{pH} < 8$, the negative correlation between N_c and one specific terminal nitrogen atom N_{tx} is significantly more pronounced than that between N_c and N_t (cf. Figure 2.6 b and 2.6 d). At $\text{pH} < 8$, the protonation probability of N_t is close to 1, *i.e.*, at least one of the terminal nitrogen atoms is protonated. Hence, there is no variance in the behavior of N_t . Therefore, the covariance and accordingly the correlation between this site and another will be small. Thus, even though the behavior of N_c and N_t is dissimilar, *i.e.*, N_c has a protonation probability of less than 0.5 and N_t is protonated, their behavior is not correlated. In contrast, N_{tx} begins to deprotonate at $\text{pH} = 4$. In the pH range from 5 to 8, its protonation probability is approximately 0.85. Thus, although N_{tx} is likely to be protonated in this pH range, there is some variance in its behavior. Therefore, the protonation behavior of N_c and N_{tx} is significantly negatively correlated.

Correlation between two sites can occur when there is variance in their behavior. At low pH values, all nitrogen atoms are protonated and at high pH values, they are deprotonated. Correlation behavior can be seen only in between the extreme pH values, where the nitrogen atoms have some probability to be either protonated or deprotonated. Interestingly, the example of DTPA demonstrates that positive correlation can occur despite exclusively repulsive interaction in molecules which have a complicated interaction network, *i.e.*, the positive correlation between N_{t1} and N_{t2} .

2.5.4 PROBABILITY OF PROTONATION STATE SUBSETS

Often only the protonation of some residues is of interest, while the behavior of other protonatable sites may be neglected. A subset S of all protonation states can be characterized by a vector \vec{x}_{sub} that describes only the protonation form of the protonatable sites of interest.

Given for example a system with three sites that each can adopt two forms. Each state of this system is defined by a state vector $\vec{x}_n = (x_1, x_2, x_3)$. Assuming only the behavior of site 1 and 3 is of interest: A subset \mathcal{S} of states can be characterized by $\vec{x}_{\text{sub}} = (x_1, x_3)$, where x_1 and x_3 specify the protonation form of interest for site 1 and 3, respectively. In this case, \mathcal{S} characterizes two states that are identical at site 1 and 3, but differ in the protonation form of site 2. The vector \vec{x}_{sub} , thus, describes a subset of the complete space of protonation states.

For a given protonation state n described by the state vector \vec{x}_n and for a subset \mathcal{S} characterized by \vec{x}_{sub} the following set function is defined:

$$\begin{aligned} \delta_n(\vec{x}_{\text{sub}}) &= 1 & \text{for } \vec{x}_n \in \mathcal{S} \\ \delta_n(\vec{x}_{\text{sub}}) &= 0 & \text{for } \vec{x}_n \notin \mathcal{S} . \end{aligned} \quad (2.30)$$

To derive the probability for the system to be in subset \mathcal{S} , one can substitute x_μ by $\delta_n(\vec{x}_{\text{sub}})$ in Eq. (2.24):

$$\langle \vec{x}_{\text{sub}} \rangle = \sum_{n=1}^{2^N} \delta_n(\vec{x}_{\text{sub}}) P_n = \frac{1}{Z} \sum_{n=1}^{2^N} \delta_n(\vec{x}_{\text{sub}}) \exp(-G_n / RT) . \quad (2.31)$$

The summand is zero for the case that the protonation state \vec{x}_n with energy G_n is not part of the subset \mathcal{S} described by \vec{x}_{sub} . Thus, only energies of states that are part of \mathcal{S} contribute to the sum. Eq. (2.31) yields the probability of the subset \mathcal{S} of protonation states characterized by \vec{x}_{sub} .

2.5.5 METROPOLIS MONTE CARLO

Monte Carlo methods comprise any method that uses statistical sampling. Their probabilistic nature is reflected by naming the method after the casinos of Monte Carlo. The potential of computers for statistical sampling in mathematical physics was first recognized by Stanislaw Ulam. Together with John von Neumann and Nicholas Metropolis, he developed a formal methodology applicable to a wide variety of problems [132]. The Metropolis Monte Carlo method employed in this thesis provides an approximation of the Boltzmann distribution of the states [133]. A flowchart of the sampling procedure, which is summarized in the following, is presented in Figure 2.7.

THE METROPOLIS CRITERION

Input of the algorithm is a randomly chosen protonation state with the protonation state energy G_{old} . Then a random site μ is chosen and its protonation form is changed. The state energy of the new state, G_{new} , is determined and the change in energy between old and new state is calculated: $\Delta G = G_{\text{new}} - G_{\text{old}}$.

The new state is accepted, if $\Delta G \leq 0$. For the case that $\Delta G > 0$, a random number r is chosen with $0 \leq r \leq 1$ such that all r are evenly distributed between 0 and 1. The new state is accepted, if $\exp(-\Delta G/k_B T) \geq r$. If $\exp(-\Delta G/k_B T) < r$, the new state is rejected and the

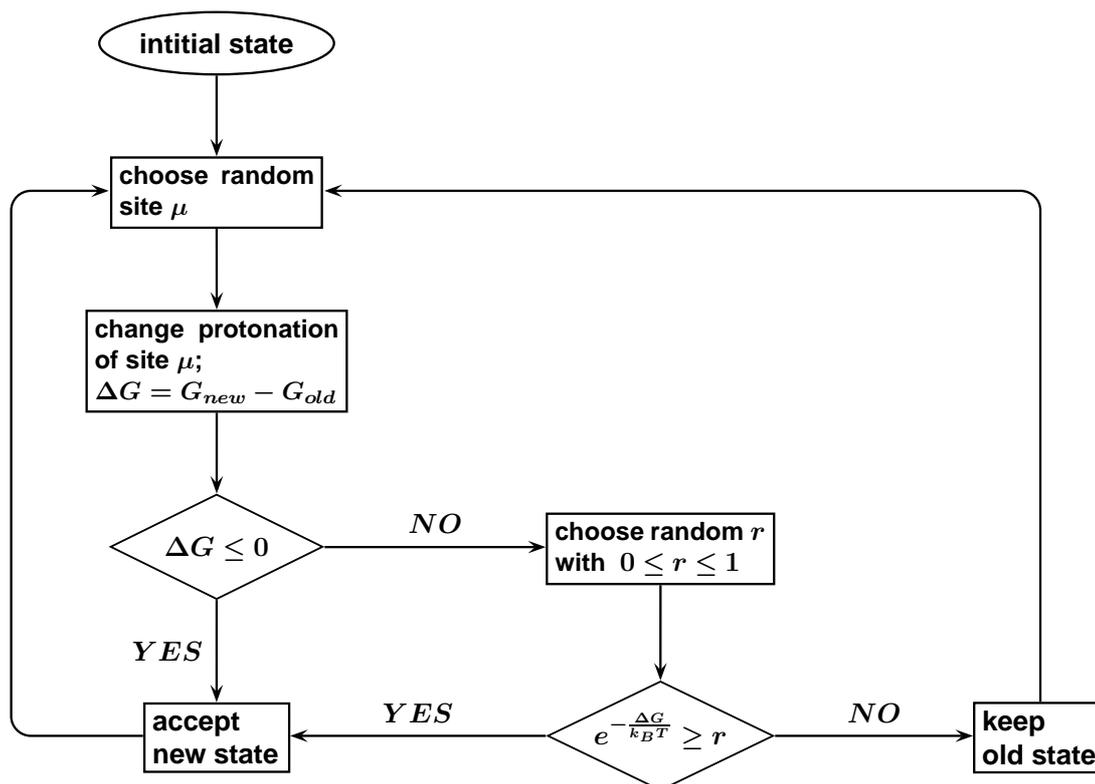


Figure 2.7. Metropolis Monte Carlo. From an initial random protonation state a random site is chosen and its protonation is changed. The new state is accepted if the energy difference between old and new state is smaller than zero. If this is not the case, the state is still accepted, if the Boltzmann factor of the change in energy is larger than a number randomly chosen between 0 and 1. The Metropolis criterion is iteratively applied until either a certain number of steps is performed or convergence of the sampled properties is achieved.

system remains in the old state. This Metropolis criterion ensures that the sampling of the phase space is biased towards the Boltzmann distribution, *i.e.*, the important region of the phase space.

DOUBLE AND TRIPLE MONTE CARLO MOVES

Metropolis Monte Carlo can approximate the Boltzmann distribution of states only, if the algorithm in principle allows to reach the complete search space. For some systems, regions of the search space may be separated by a high energy barrier that can only be overcome with a very small probability. Then, the system may not or too seldom traverse from one region to the other during the Monte Carlo sampling. In such cases, the Boltzmann distribution is not approximated. To prevent this problem for the calculation of protonation probabilities, convergence is ensured by introducing double or triple Monte Carlo moves. In such moves, the protonation form of two or three sites, respectively, is changed simultaneously in one Monte Carlo step. Double moves are introduced, when a pair of sites has an interaction energy $W_{\mu\nu}$ above a certain threshold. Equivalently, triple

moves can be introduced, when the interaction energies of site μ with two other sites ν and γ both lie above a certain threshold.

Considering a pair of sites μ and ν for which the states with one proton, *i.e.*, $\vec{x}_a = (0_\mu, 1_\nu)$ and $\vec{x}_b = (1_\mu, 0_\nu)$, have a low energy. Furthermore, it shall be assumed that the intermediate states of the transition from \vec{x}_a to \vec{x}_b or *vice versa*, *i.e.*, $\vec{x}_c = (1_\mu, 1_\nu)$ and $\vec{x}_d = (0_\mu, 0_\nu)$, have a very high energy. Then, starting the Metropolis Monte Carlo sampling from either \vec{x}_a or \vec{x}_b , would render the sampling of the other state highly improbable. This sampling problem can be overcome by changing the protonation form of both sites μ and ν simultaneously in a so-called double move.

DERIVATION OF PROBABILITIES FROM METROPOLIS MONTE CARLO

The Metropolis Monte Carlo method samples a set of states which approximates the Boltzmann distribution of all states. Thus, the protonation probability as well as the correlation coefficient can be approximated employing this method. Assuming M states are sampled, the protonation probability $\langle x_\mu \rangle$ defined by Eq. (2.24) can approximately be computed by:

$$\langle x_\mu \rangle = \frac{1}{M} \sum_{i=1}^M x_{\mu,i} \quad . \quad (2.32)$$

Equivalently, the protonation probability $\langle x_\mu x_\nu \rangle$ given by Eq. (2.27) is approximated as:

$$\langle x_\mu x_\nu \rangle = \frac{1}{M} \sum_{i=1}^M x_{\mu,i} x_{\nu,i} \quad , \quad (2.33)$$

and the probability of protonation substates given by Eq. (2.31) can be approximately computed by:

$$\langle \vec{x}_{\text{sub}} \rangle = \frac{1}{M} \sum_{i=1}^M \delta_i(\vec{x}_{\text{sub}}) \quad . \quad (2.34)$$

Furthermore, Eq. (2.32) and Eq. (2.33) allow to compute the correlation coefficient $c_{\mu\nu}$ of the protonation of two sites μ and ν by the formula of Eq. (2.29):

$$c_{\mu\nu} = \frac{\text{cov}(x_\mu, x_\nu)}{\sigma_\mu \sigma_\nu} = \frac{\langle x_\mu x_\nu \rangle - \langle x_\mu \rangle \langle x_\nu \rangle}{\sqrt{(\langle x_\mu \rangle - \langle x_\mu \rangle^2)(\langle x_\nu \rangle - \langle x_\nu \rangle^2)}} \quad . \quad (2.35)$$

In this work, probabilities are computed utilizing the Metropolis Monte Carlo method and computing the equations given above. The parameters used are detailed in the next chapter.

CHAPTER 3

HIGH-RESOLUTION RHODOPSIN PROTEIN STRUCTURES AND STRUCTURE PREPARATION

*Could the search for ultimate truth really have revealed
so hideous and visceral looking an object?*

Max Perutz, The Hemoglobin Molecule

The dynamics of a protein and, therefore, its function, *e.g.*, ligand binding, substrate catalysis or the transmission of signals, is determined by the protein structure. An important step towards understanding the functional mechanism of a protein is, thus, the determination of its three-dimensional structure in atomic detail. Due to the complexity and diversity of proteins this remains one of the fundamental challenges in protein science both in experimental and in theoretical approaches. The main goal is to find structure of lowest energy which may then provide insight into the functional mechanism of the protein. Furthermore, the protein folding mechanism which is as yet only partly understood is extensively researched. A theoretical understanding of protein folding may significantly advance further attempts to predict protein structures from amino acid sequences.

In the following sections, theoretical and experimental protein structure determination will be briefly discussed. Thereafter, details of the calculations performed on the experimental X-ray structures of bacteriorhodopsin (BR), halorhodopsin (HR) and sensory rhodopsin II (SRII) are presented. The necessary preparations of the structures used for the calculations will be described in Subsection 3.2.3. Section 3.5 covers the electrostatic calculation of the pK_{intr} value and the interaction energy matrix $W_{\mu\nu}$ (*cf.* Chapter 2, Subsections 2.4.1 and 2.4.2).

3.1 PROTEIN STRUCTURE PREDICTION

Due to the difficult and time-consuming task of experimental protein structure determination, major efforts are made to develop and improve theoretical techniques for protein structure prediction. Attempts in *de novo* protein structure prediction from the amino-acid sequence have met with varying success. For this task, large computational resources like those offered by powerful supercomputers or distributed computing, *i.e.*, grid computing, projects are needed.

The world's largest distributed computing projects is Folding@Home devised by Vijay Pande at Stanford University. In 2009, Folding@home officially reached a performance level above 5 native petaFLOP being the first computing system to do so. Folding@home aims at understanding the dynamics of the protein folding process [134, 135]. Initially the folding of small molecules was studied, for example the folding pathway of the c-terminal β -hairpin from protein G [136]. Since then, the folding of more complex proteins has been simulated: for example, the effect of mutations on the tumor suppressor protein p53 [137]. Another distributed computing project is Rosetta@home that was developed by David Baker's group at the University of Washington [138]. The main goal of Rosetta@home is the prediction of the lowest energy structure of proteins. Rosetta, a knowledge-based force field, is one of the top performing methods for *de novo* prediction in the Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiment. One of the hallmarks was the design of a protein called Top7 with a novel sequence and topology [139]. Top7 was experimentally shown to be stable and the X-ray structure closely resembles the design model.

Homology modeling, also termed comparative or knowledge-based modeling, develops a three-dimensional model from a protein sequence based on the structures of homologous proteins [140, 141]. This approach is based on the observation that the three-dimensional structures of homologous proteins are conserved to a greater extent than their primary structures. Furthermore, it appears that the number of tertiary structural motifs is limited. It has been suggested that there are only approximately 2000 distinct protein folds in Nature, though there are many millions of different proteins. However, despite the encouraging developments in protein structure prediction, to date high-quality, high-resolution structures suitable for theoretical investigations of the functional mechanism are still mainly determined by experimental methods.

3.2 EXPERIMENTAL PROTEIN STRUCTURES

Currently, the Brookhaven Protein Data Bank stores over 60 000 protein structures, of which the majority is solved by X-ray crystallography [6]. NMR (nuclear magnetic resonance) spectroscopy [142] renders approximately 13 % of all resolved protein structures. Despite its importance in protein structure determination, NMR will here not be further discussed, since all calculations of this work are performed on X-ray structures.

Using X-ray scattering the first protein structure was resolved in atomic detail in 1958 by John Kendrew, *i.e.*, myoglobin at about 2.5Å [143]. Kendrew commented that '*perhaps*

the most remarkable features of the molecule are its complexity and its lack of symmetry. The arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates.' Two years later, Max Perutz whose ideas led to major advances of the X-ray crystallization method succeeded in resolving the structure of hemoglobin at 5.5 Å [144]. His first impression of the protein structure precedes this chapter. For their pioneering work, Max Perutz and John Kendrew were awarded the Nobel prize in chemistry *for their studies of the structures of globular proteins* in 1962. Today, X-ray crystallography is the predominant method of protein structure determination in atomic detail.

3.2.1 MEMBRANE PROTEINS

Out of the approximately 8 000 unique protein structures only about 100 belong to membrane proteins. The sparse number of available membrane protein structures stands in direct opposition to their crucial role in the cellular metabolism [53, 145]. Membrane proteins mediate between the cell and its environment and between organelle and cytoplasm. For example, they transport metabolites and ions across the membrane, thereby providing the substances necessary for growth. Most importantly, they convert energy stored in electron-carriers like flavin adenine dinucleotide (FAD) and nicotinamide adenine dinucleotide (NAD⁺) and also electromagnetic energy into an electrochemical gradient of protons across the membrane. The so-called proton motive force is then utilized by another membrane protein to synthesize ATP. Hence, membrane proteins supply the energy for all metabolic processes. Furthermore, membrane proteins that act as receptors transmit signals across the membrane, *e.g.*, signals from neurotransmitters, light or mechanical stimuli. In accordance with their vital importance, about one-third of all known genes codes for proteins having at least one transmembrane segment.

The rhodopsin transmembrane proteins of halophilic bacteria have become prototypical for the investigation of membrane proteins. For one, this is due to their function as energy converters and as (photo-)receptors, thereby exemplifying two major groups of membrane proteins. Furthermore, they are relatively small, simple and stable proteins which made their crystallization and the subsequent determination of their three-dimensional structure possible.

3.2.2 CRYSTALLIZATION OF MEMBRANE PROTEINS

The X-ray diffraction pattern of well-ordered protein crystals can be translated into their three-dimensional structure. Protein crystals are, however, not easy to grow and the preparation of diffraction-quality crystals remains the bottleneck of X-ray structure determination. One of the major problems is the search for the right detergent and conditions, since crystallization requires a homogeneous and monodisperse solution of the protein at a relatively high concentration. Due to their unique solvent, these problems are even more pronounced in the case of membrane proteins and account for the few structures that could be resolved to date.

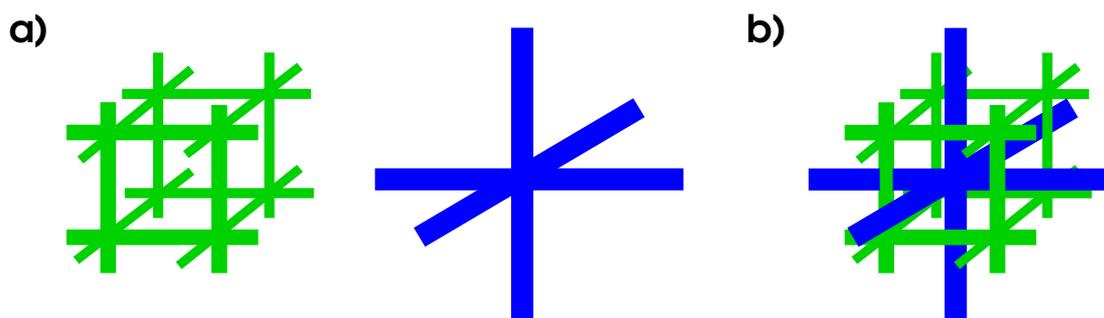


Figure 3.1. Lipidic cubic phase. The lipidic cubic phase is composed of a bicontinuous surface that creates two mutually exclusive solvent compartments. **a)** Depicted are two labyrinths, green and blue, that are congruent to each other. These two labyrinths represent the two solvent compartments. **b)** Overlay of the two labyrinths. The space between the two labyrinths, *i.e.*, solvent compartments, is filled with the continuous bilayer (not shown). Thereby, the continuous bilayer separates the two solvent compartments from each other. A detergent-solubilized protein such as bacteriorhodopsin will integrate itself into the bilayer. The incorporated protein is believed to act as a reservoir for growing crystals.

Trans-membrane proteins have both hydrophobic and hydrophilic regions on their surfaces. The former interact with the non-polar chains of lipids and the latter are exposed to the aqueous solution. Due to the large hydrophobic surface region, they are not soluble in aqueous solution and denature. In the 1970s, membrane proteins were successfully solubilized by adding detergents to an aqueous solution which subsequently led to the first crystallizations of membrane proteins. Finally in 1984, the first structure of a membrane protein, a photosynthetic reaction center, could be determined [146]. This breakthrough was rewarded with the Nobel prize in chemistry received by Johann Deisenhofer, Robert Hubert and Hartmut Michel in 1988 *for the determination of the three-dimensional structure of a photosynthetic reaction center*. The determination of other membrane protein structures progressed relatively slowly, and is limited to proteins that do not denature easily. Therefore, improvements of the detergent micelle-based crystallization as well as entirely new approaches are sought continuously.

LIPIDIC CUBIC PHASE

In 1996, the revolutionary concept of crystallizing membrane proteins in lipidic cubic phases was introduced by the group of Ehud M. Landau [147]. Mixing a long-chain lipid, *e.g.*, monoolein, with a small amount of water results in curved continuous bilayers extending in three dimensions which are penetrated by interconnected aqueous channels, *i.e.*, a lipidic cubic phase. An idea of such a lipidic cubic phase can be gained from the schematic depicted in Figure 3.1. A detergent-solubilized protein can integrate itself into the bilayer which ultimately leads to its crystallization. Since its introduction, this method has yielded well-ordered crystals and thereafter X-ray structures of several membrane proteins. Crystallization of membrane proteins in lipidic cubic phases is described in detail in Refs. [148] and [149].

3.2.3 ARCHAEAL RHODOPSIN STRUCTURES

The lipidic cubic phase provided the means to crystallize proteins of the archaeal-type rhodopsin family. During the last decade approximately 50 high-resolution X-ray structures of archaeal rhodopsins could be resolved. The results presented in this thesis are based on calculations performed on BR, HR and SRII structures. The structures chosen for the calculations are listed in Table 3.1. The amino acid sequences of BR and HR are taken from the halophilic archaeon *Hb. sal* and the sequence of SRII from *Natronomonas pharaonis*. The BR structures will be referred to by their intermediate state and PDB code as listed in Table 3.1, *e.g.*, M:1kg8. The HR and SRII structure will be referred to as such.

BACTERIORHODOPSIN

BR is organized in trimers, which are densely packed in the purple membrane. The trimeric structure appears to be stabilized by direct protein-protein interactions, whereas each BR trimer is surrounded by a single layer of lipids such that no direct protein-protein contact exists between neighboring trimers [160, 161]. BR is unusual among membrane proteins since the majority of its mass is membrane-embedded, existing as highly ordered two-dimensional crystalline arrays. The protein content of the purple membrane is as high as 75%.

Several high-resolution structures of the bR state, *i.e.*, the ground state, are available for BR. Efforts have been made to obtain high-resolution structures of every intermediate state. Illumination of the BR crystals thereby triggering the photocycle, in principle, allows to trap the other intermediate states. A major difficulty of this approach lies in the differentiation between the intermediate states due to the strong overlap of their absorption spectra. Only in the case of the M intermediate, *i.e.*, the so-called blue intermediate, a wavelength can be chosen, where no other photocycle intermediate absorbs light. Despite these difficulties, these efforts have resulted in structures of the K, L, M and N photocycle intermediate.

In general, however, the crystals used for data collection do not contain a perfectly pure ensemble of one photocycle intermediate. For example, the M intermediate is commonly subdivided into the M₁ and M₂ state. The M₁ and M₂ state differ in their protonation state, but not in their absorption properties. Furthermore, they succeed each other quickly. It is, therefore, difficult to obtain a pure crystal of either state. Although some M structures are categorized as M₁ or M₂, most are published simply as M structures. Likewise in this thesis, the structures will be classified as M structures without further subdivision.

Upon lowering the temperature which is necessary for X-ray data collection, the O intermediate state disappears. Hence, no structure of this intermediate state exists. Instead, an O-like structure is available. Its coordinates were obtained from the Asp85Ser mutant of BR. Substituting Asp85 in BR for a threonine or a serine, converts BR to a HR-like chloride pump, since these mutants can bind halide ions near the retinal Schiff base. In the absence of halide ions and in the ground state, the Asp85Ser mutant of BR resembles the O intermediate state of wild type BR. In particular, the retinal Schiff base is in its all-*trans* conformation and the Schiff base and Asp96 are protonated, while the mutant

	STATE	PDB ENTRY	RES. [Å]	CONDITIONS T [K]	pH	MUTATION	REMARK	REF.
BR	bR	1c3w ^a	1.55	100	5.6			[42]
		1f50 ^a	1.70	100	5.6	E204Q		[150]
		1iw6	2.30	100	5.2			[151]
		1qhj	1.9	100	5.6			[41]
	K	1ixf	2.60	100	5.2			[151]
		1m0k ^a	1.43	100	5.6			[152]
		1qko	2.1	110	5.6			[153]
		1qkp	2.1	110	5.6			[153]
	L	1e0p	2.1	110	5.6			[154]
		1o0a ^a	1.62	100	5.6			[155]
		1vjm	2.3	100	5.6		1vjm.a, 1vjm.b	[156]
	M	1cwq	2.25	100	5.6			[157]
		1f4z ^a	1.80	100	5.6	E204Q		[150]
		1kg8 ^a	2.00	100	5.6		M ₁	[43]
		1kg9 ^a	1.81	100	5.6	E204Q	M ₁ , 'mock-trapped'	[43]
		1m0m ^a	1.43	100	5.6			[158]
		1p8h ^a	1.52	100	5.6		M ₁ ;	[45]
	N	1p8u ^a	1.62	100	5.6	V49A		[45]
	O	1jv7 ^b	2.25	100	5.6	D85S	O-like structure	[159]
HR	ground state	1e12	1.8	100	7.0			[46]
SRII	ground state	1jgj	2.4	100	5.3		from <i>N. pharaonis</i>	[48]

^a The electron density of the EF loop could not be resolved and the PDB entry contains no coordinates for this loop.

^b The PDB entry contains no coordinates for the BC loop.

Table 3.1. X-ray structures of the archaeal rhodopsins. For the investigation of the absorption tuning mechanism in Chapter 4, the BR structure 1qhj and the HR and SRII structures are used. The protonation behavior of all intermediate structures of BR listed in this table is analyzed in Chapter 5. For the calculations presented in Chapter 6, the M state structure 1kg9 was used. For the analysis of all lowest energy states, discussed in Chapter 7, calculations were performed on the bR state structure 1c3w and the M state structure 1kg8.

is neutral at the former Asp85 position. The structure of this Asp85Ser mutant, solved in the absence of halide ions, is, therefore, assumed to represent the structure of the O intermediate.

For BR, more structures are available than listed in Table 3.1. For instance, BR structures were determined, where the aim was not the investigation of the functional mechanism, but general protein stability. The majority of the structures not listed in Table 3.1 are, however, ground state structures. Calculations performed on these bR state struc-

tures showed that they are nearly identical in their behavior. To avoid redundancy, data is presented for only four bR state structures.

HALORHODOPSIN AND SENSORY RHODOPSIN II

In crystallized form HR is organized in trimers like BR, whereas the functional unit of SRII is a heterodimer formed together with its transducer HTRII. For the calculations presented in this thesis, the ground state structures of HR and SRII listed in Table 3.1 are used.

3.3 PREPARATION OF THE X-RAY STRUCTURES

The X-ray structures listed in Table 3.1 have to be prepared for the electrostatic calculations presented in this thesis. The resolution is for example not high enough to reveal the positions of hydrogen atoms. These can only be detected if the structure is resolved to a value below 1 Å. Care is taken to keep the modifications of the structures to a minimum to avoid biasing the results of the electrostatic calculations. For instance, the coordinates stored in the Protein Data Bank are in general not altered. A minimization of the coordinates is in most cases restricted to those added to the structure, *i.e.*, hydrogen atom coordinates or the atom coordinates of unresolved residues.

MINIMIZATION OF THE POTENTIAL ENERGY FUNCTION

All energy minimizations are done using CHARMM (Chemistry at Harvard Macromolecular Mechanics) [162]. In the CHARMM force field, atoms are represented as charged point masses. Its fundamental potential function V is separated into bonded and non-bonded terms. The bonded terms V_{bonded} describe the bonds, angles and bond rotations in a molecule and the non-bonded terms $V_{\text{non-bonded}}$ account for interactions between non-bonded atoms, *i.e.*, atoms separated by three or more covalent bonds:

$$\begin{aligned} V_{\text{total}} &= V_{\text{bonded}} + V_{\text{non-bonded}} \\ &\text{where} \\ V_{\text{bonded}} &= V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihedral}} + V_{\text{improper}} \\ &\text{and} \\ V_{\text{non-bonded}} &= V_{\text{vdW}} + V_{\text{elec}} \end{aligned}$$

where V_{bond} is the bond stretching energy term, V_{angle} is the angle bending energy term, V_{dihedral} accounts for the distortion around a bond, *i.e.*, the dihedral or torsion angle, and V_{improper} is the distortion term, *i.e.*, the energy of so-called improper torsions. The improper torsion angle can be observed for four connected atoms where the torsion angle is not defined by four angles connected sequentially. A schematic representation of the bonded atom interactions is given in Figure 3.2. V_{vdW} is the van der Waals interaction

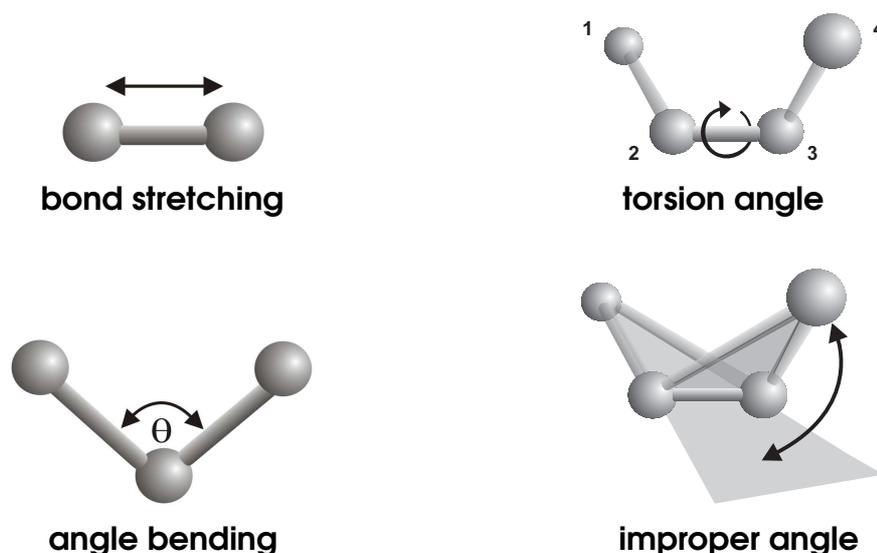


Figure 3.2. Schematic representation of the bonded terms. In the CHARMM force field, the bonded terms include the bond stretching between two covalently bound atoms, the angle bending between three covalently bound atoms and the torsion angle between four covalently bound atoms. Furthermore, the improper angle between four atoms is taken into consideration. The graphics are kindly provided by Torsten Becker.

energy modeled using a 6-12 Lennard-Jones potential, where the repulsive forces fall off with distance with the square of which the attractive forces decrease. V_{elec} is the electrostatic energy between a pair of atoms represented by the Coulomb potential.

The potential function is optimized using the steepest descent and subsequently the adopted basis Newton-Raphson method. Steepest descent is the simplest way to optimize a function. In each step of steepest descent, the coordinates are adjusted in the negative direction of the gradient, *i.e.*, in the direction opposite to its most significant increase. The adopted basis Newton-Raphson method in principle expands the function to the second order and attempts to find the point where the gradient of the second-order approximation is zero.

ATOMIC PARTIAL CHARGES

The atomic partial charges for the standard amino acids are taken from the CHARMM22 parameter set of the CHARMM force field [163]. Since the CHARMM22 parameter set does not contain partial charges for the retinal Schiff base these are taken from Ref. [75].

MISSING COORDINATES

Coordinates for residue atoms that could not be determined from the electron density map are generated from internal coordinates of the CHARMM22 topology using the ICBUILD routine of CHARMM. Coordinates generated from internal coordinates are minimized with the steepest descent and subsequently the adopted basis Newton-Raphson

method until a tolerance gradient criterion is fulfilled. Using the parameter TOLGRD, the minimization is stopped, if the gradient is less or equal to 10^{-7} .

As indicated in Table 3.1, some BR structures miss coordinates for the residues 157–161. These residues are part of the highly flexible cytoplasmic EF-loop. The coordinates for the missing residues of the EF-loop are taken from the 1qhj structure [41] after superposition of the backbone C_{α} atoms using the Kabsch algorithm [164]. The EF-loop and its flanking residues, *i.e.*, residues 150–176, are then carefully relaxed, while all other heavy atom coordinates are kept fixed. During the minimization, harmonic constraints are used to limit coordinate changes. A force constant of 1 kcal/(molÅ²) restrained the residues 157–161 and the nearby Met163 residue. Higher restraints are applied to the residues flanking the EF-loop, *i.e.*, residues 150–154, 162 and 164–176. Their backbone and side chain atoms are constrained by a force constant of 5 and 2 kcal/(molÅ²), respectively. Only the side chains of Arg175 and Lys172 are restrained with a weaker force constant of 0.1 kcal/(molÅ²). The steepest descent subsequently followed by the adopted basis Newton-Raphson method is used for the minimization until a tolerance gradient criterion is fulfilled. Using the parameter TOLGRD, the minimization is stopped, if the gradient is less or equal to 10^{-7} . The O-like structure of BR, 1jv7, is lacking coordinates for the BD-loop. These are generated in equivalence to the procedure described above for the EF-loop.

The N- and C-termini could not be resolved by X-ray crystallography due to their high flexibility. Their absence has, however, no significant influence on the electrostatic calculations performed during this thesis. This is due to the solvent exposure of the termini that screens their charges. Furthermore, they are relatively far from all functionally important regions that are analyzed here. Neutral blocking groups are added to the main-chain termini, *i.e.*, an acetylated N-terminus and an N-methylamide C-terminus. Mutants are converted back to the wild type by replacing the mutated with the respective wild type amino acid side chain and minimizing the new coordinates.

HYDROGEN ATOMS

For all structures, hydrogen atoms are constructed using the HBUILD routine of CHARMM and their positions are optimized with the steepest descent and subsequently with the adopted basis Newton-Raphson routine until a tolerance gradient criterion is fulfilled. Using the parameter TOLGRD, the minimization is stopped, if the gradient is less or equal to 10^{-7} . All non-hydrogen atom coordinates are constraint during the minimization. The hydrogen atom positions are minimized with all internal water molecules present. For the electrostatic calculations, the water molecules are deleted and represented implicitly by a continuum that accounts for reorientation of the water dipoles in response to the electrostatic field.

IONS

The chloride ion which is resolved for the HR structure in the chloride binding site below the retinal Schiff base is included in all calculations. Additionally, the electron density of HR revealed a potassium ion which is not considered in the calculations. No biological

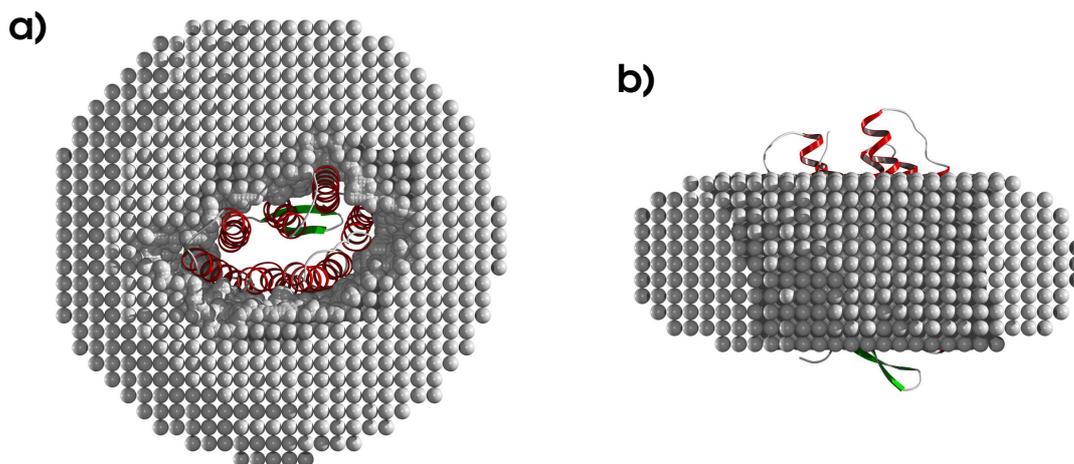


Figure 3.3. Membrane model. Depicted is the torus of tightly packed dummy atoms around BR. The dummy atoms are uncharged and have a radius of 1.5 Å. **a)** View from the cytoplasm. **b)** View along the membrane plane, the cytoplasm is at the top. These figures are kindly provided by Nicolas Calimet.

function can be attributed to this potassium ion. Furthermore, it is located relatively far from the retinal Schiff base. The presence of this ion is not discussed in the publication of this structure, possibly since it may in fact be a water molecule [46].

LIPIDS

HR binds a palmitic acid which is located approximately parallel to helix D. In its deprotonated form, the palmitic acid is negatively charged. It is, therefore, included in the electrostatic calculations as a protonatable site. Other lipids resolved for archaeal rhodopsin structures are neutral and are, therefore, not considered in the electrostatic calculations. The lipids surrounding the archaeal rhodopsins are represented by a membrane model as described in the next section.

3.4 THE MEMBRANE MODEL

The archaeal rhodopsins are transmembrane proteins. To model the apolar properties of the membrane environment a ring of uncharged dummy atoms is generated around the protein representing the hydrophobic core of the membrane. The membrane model is depicted in Figure 3.3. The ring of uncharged, apolar dummy atoms resembles a micelle of lipids surrounding the protein. The dummy atoms with a radius of 1.5 Å contact each other in a close-packed, face-centered array fashion. The thickness of the membrane model is 31 Å which is compatible with experimental measurements of membrane thickness, if the polar heads of the lipids, *i.e.*, about 5 Å, are not considered [165]. The density of the dummy atoms increases along the protein surface to avoid artificial high-dielectric cavities. The membrane is described in detail in Ref. [64].

RESIDUE	pK_{mod}	REFERENCE FORM
arginine	12.0	0
aspartic acid	4.0	1
cysteine	9.1	1
glutamic acid	4.4	1
histidine (N_{δ})	7.0	0
histidine (N_{ϵ})	6.6	0
lysine	10.4	0
tyrosine	9.6	1
retinal Schiff base	7.0	0
palmitic acid	4.7	1

Table 3.2. The pK_{mod} values and the reference form of the protonatable sites.

Using these pK_{mod} values, the pK -shift upon transfer of the model compound into the protein-membrane environment can be calculated [125, 126, 166]. In the MEAD program package which is used for the calculations presented in this thesis, the reference protonation form of each residue is defined as the uncharged form. If the reference form is 1, the neutral form is protonated, if it is 0, the neutral form is deprotonated.

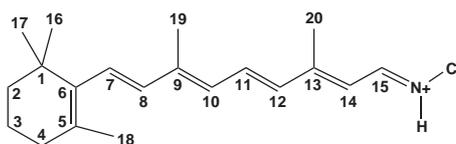
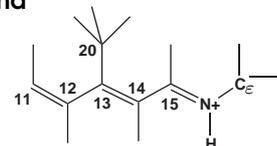
This representation of the membrane offers several advantages over the infinite low-dielectric slab model which is commonly used to model transmembrane proteins in continuum electrostatics. First, the ionic strength can be included in the calculation of the electrostatic potential of the system. Moreover, all protein cavities are filled with a high dielectric medium. In the infinite slab method, a cylinder is used to define the region which may contain protein cavities, *i.e.*, high-dielectric regions inside the membrane part of the protein. The regions in the cylinder volume which are not occupied by the protein are treated as high dielectric regions. Thus, if the cylinder is chosen too big, the high-dielectric region expands into the membrane. If it is chosen too small, a low dielectric medium fills the cavities lying outside the cylinder.

3.5 CALCULATION OF THE pK_{intr} VALUE AND THE INTERACTION ENERGY $W_{\mu\nu}$

All aspartate, arginine, cysteine, glutamate, histidine, lysine and tyrosine residues, the retinal Schiff base and the palmitic acid are considered as protonatable sites. The pK_{mod} values of the respective model compounds and the reference protonation form are listed in Table 3.2. The charges of the retinal Schiff base model compound are listed in Table 3.3. These charges are derived from a density functional calculation. Details are given in Ref. [64]. The charges of the retinal not listed in Table 3.3 are constant in the protonated and deprotonated form and are taken from Ref. [75]. Atomic radii are set to 1.7 Å for carbon atoms, 1.0 Å for hydrogen atoms, 1.5 Å for oxygen atoms, 1.55 Å for

Atom	Protonated	Deprotonated
C11	-0.03391	-0.23446
H11	0.20455	0.20970
C12	-0.27954	-0.27113
H12	0.21729	0.19015
C13	0.34170	0.36220
C20	-0.28589	-0.31866
H20A	0.12636	0.10206
H20B	0.12636	0.10206
H20C	0.12636	0.10206
C14	-0.50964	-0.62811
H14	0.24913	0.21799
C15	0.15469	0.34992
H15	0.19675	0.03840
NZ	-0.19133	-0.30054
HZ	0.33665	0.00000
CE	-0.15774	-0.19392
HE1	0.19361	0.14129
HE2	0.18459	0.13100

model compound



retinal Schiff base

Table 3.3. Model compound of the retinal Schiff base. The table lists the atomic partial charges for the protonated and deprotonated forms of the retinal Schiff base model compound. The partial charges have been determined by density functional calculations [64]. The protonated form of the retinal Schiff base and its model compound are depicted next to the table. For the model compound the hydrogen atoms are indicated. The nitrogen and the carbon atom C_ϵ are part of the lysine residue which, together with the retinal, forms the Schiff base. In the table, the hydrogen atoms are referred to with respect to the carbon or nitrogen atom they are bound to. The hydrogen atom bound to the nitrogen atom, *i.e.*, referred to as HZ and NZ in the table, is not present in the deprotonated form of the retinal Schiff base. The charges of the atoms not listed in the table are constant in the protonated and deprotonated form of the retinal Schiff base and taken from Ref. [75].

nitrogen atoms, 1.8 Å for sulfur atoms and 1.8 Å for the chloride ion. In general, these atomic radii correspond to Bondi radii [167].

The linearized Poisson-Boltzmann equation is implemented in the MEAD package (Macroscopic Electrostatics with Atomic Detail) [168]. MEAD uses a finite difference method for the numerical solution of the linearized Poisson-Boltzmann equation. The theory has been described in detail in Chapter 2, Subsection 2.3.3. The dielectric boundary, *i.e.*, the solvent accessible surface, between the protein and the solvent is derived using a probe sphere with a radius of 1.4 Å and the atomic radii as given above. The ion exclusion layer is derived by adding 2 Å to the value of all atomic radii. Inside the protein, the ionic strength is zero, since no internal cavity lies within the ion exclusion layer. The calculations are performed using a dielectric constant of 80 for the solvent and a dielectric constant of 4 for the protein and the membrane environment. In all calculations, the temperature is set to 300 K and the ionic strength of the solvent to 0.1 M.

The pK_{intr} (cf. Eq. (2.17)) and the interaction energies $W_{\mu\nu}$ (cf. Eq. (2.20)) between all pairs of protonatable sites are calculated using the MULTIFLEX program of MEAD. The finite-difference grids for solving the linearized Poisson-Boltzmann equation are focused in two steps. For the protein, a 100 Å cube with 1 Å spacing centered on the molecular center of the protein is used first, followed by a 20 Å cube with 0.25 Å spacing centered on each protonatable site. For the model compound, a 60 Å cube with 1 Å spacing centered on the molecular center of the protein is used first, followed by a 15 Å cube with 0.25 Å spacing centered on each model compound. These calculations are the basis for the calculations presented in the following chapters.

CHAPTER 4

THE INTER-PROTEIN SHIFT

Nichts Süßeres gibt es, als der Sonne Licht zu schaun.

Friedrich Schiller

Sun light is one of the determinant factors of life. The conversion of light into chemical energy during photosynthesis constitutes the main energy source of living organisms. Moreover, light is not only used as energy source but as an external signal. In phototaxis single-cell organisms use light signals to escape harmful and search for favorable light conditions. In higher organisms, photosensitive proteins provide the means for vision. Typically, chromophore-binding proteins tune the absorption of their chromophore to an exact maximum.

4.1 ABSORPTION TUNING IN RHODOPSIN PROTEINS

Absorption tuning of rhodopsins has been in the focus of research for decades [97, 98]. This protein family is of particular interest, since the visual pigments belong to this class. The comprehension of the tuning mechanism in rhodopsins will further the understanding of vision and in particular of the ability to sense color. In addition, rhodopsin proteins are of interest due to the extreme variation of absorption maxima in this protein family - ranging from ultra violet to the far red [105, 169]. However, despite considerable theoretical and experimental effort the molecular details of the tuning mechanism are still debated [99, 106–108, 110–112, 170–181].

As outlined in Chapter 1, rhodopsins are seven-transmembrane-helix proteins that bind their chromophore retinal via a Schiff base linkage to a lysine residue [8]. A protonated retinal Schiff base absorbs at 440 nm in polar solutions as indicated in Figure 4.1 [96]. The absorption shift between a protonated retinal Schiff base in solution and in the protein is termed opsin shift [99].

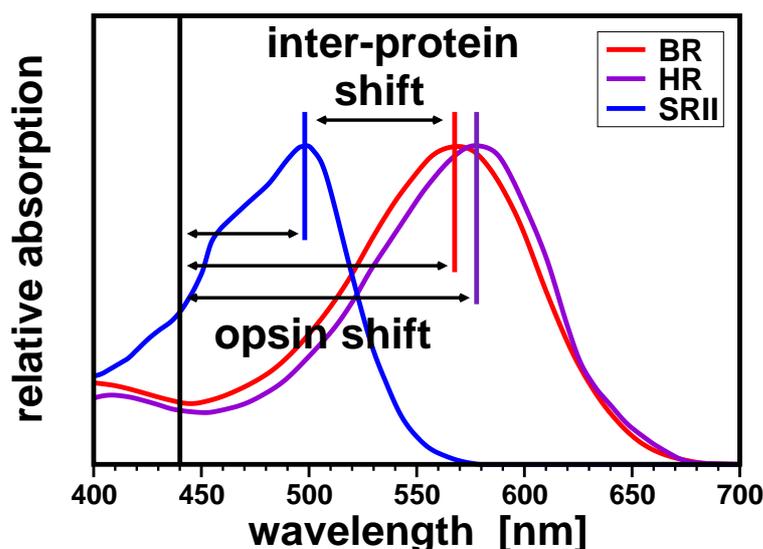


Figure 4.1. Inter-protein and opsin shift in archaeal rhodopsins. The experimental absorption spectra of BR [101], HR [102] and SRII [87] are depicted. The absorption maximum of a protonated retinal Schiff base is 440 nm in methanol solution as indicated by the black line [96]. The difference between opsin shift and inter-protein shift is indicated by the arrows. The opsin shift is defined as the shift in absorption between retinal in solution and retinal in a protein matrix. The inter-protein shift is defined as the shift between the retinal absorption maximum in different rhodopsin proteins.

This chapter does not focus on the opsin shift itself, but on the absorption shift between different rhodopsins. This shift is in the following called *inter-protein shift*. While the opsin shift is a measure for the influence of the protein on the retinal absorption, the inter-protein shift is a measure for the differences in influence the various rhodopsins exert on the absorption of retinal.

4.1.1 THE INTER-PROTEIN SHIFT IN ARCHAEOAL RHODOPSINS

Four different rhodopsins are described for halophilic archaea: the proton pump bacteriorhodopsin (BR), the chloride pump halorhodopsin (HR), and two signaling proteins, sensory rhodopsin I (SRI) and sensory rhodopsin II (SRII) [34, 35, 51, 52, 88, 89, 91]. In their ground state, these four rhodopsins bind a protonated retinal Schiff base in all-*trans* conformation (*cf.* Figure 1.2 in Chapter 1). High-resolution structures of the ground state are available for BR, HR and SRII [41, 42, 46, 48, 49, 182–184]. These structures are very similar, particularly with respect to the conformation of their chromophore retinal and the near retinal environment (*cf.* Figure 1.5 and 1.6 in Chapter 1). Despite the structural similarity, the spectral properties of the proteins differ significantly. While the absorption maxima of BR and HR are at about 570 nm, the absorption maximum of SRII is significantly blue-shifted to about 500 nm as depicted in Figure 4.1 [87, 101, 102]. Since models describing the inter-protein shift have to account for the differences between these proteins as well as for their similarities, archaeal rhodopsins provide an ideal test case for theoretical studies on absorption tuning.

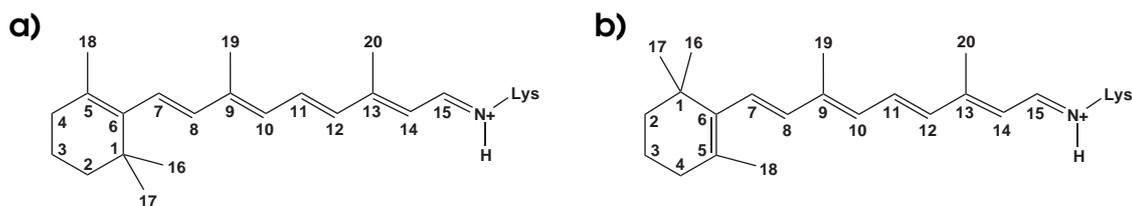


Figure 4.2. Protonated retinal Schiff base. **a)** Archaeal rhodopsins bind all-*trans* retinal via a protonated Schiff base linkage to a lysine residue in the 7th helix. The β -ionone ring is in 6-*s-trans* conformation. The conjugated π -system between C₅ and N _{ϵ} can absorb one photon whose energy is then utilized by the protein. **b)** Retinal Schiff base with 6-*s-cis* conformation of the β -ionone ring. The conjugated π -system extends from C₆ to the nitrogen atom. Thus, it is shorter than the π -system of a retinal Schiff base bound to an archaeal rhodopsin.

Two major contributions to the inter-protein shift have been discussed in the literature [108, 171, 179]. The proteins can differ either in their steric or in their electrostatic interaction with the chromophore. Steric interactions may alter the geometry of the chromophore, while electrostatic interactions influence its charge distribution. Several facts point towards a minor role of steric interactions for the inter-protein shift of archaeal rhodopsins:

(i) It is known that retinal adopts a 6-*s-trans* conformation in all archaeal rhodopsins, in contrast to the 6-*s-cis* conformation favored in solution (*cf.* Figure 4.2) [185–187]. Hence, the red shift of the absorption spectra upon the 6-*s-cis* to 6-*s-trans* isomerization, which is important for the opsin shift, does not contribute to the inter-protein shift.

(ii) The residues close to the retinal are mostly conserved among the archaeal rhodopsins, particularly the aromatic residues which define the shape of the binding pocket (*cf.* Figure 1.5 and 1.6 in Chapter 1). Accordingly, the crystal structures reveal a similar chromophore geometry for BR, HR, and SRII. In addition, the contribution of induced dipoles to the inter-protein shift that predominantly originate from the π -electrons of aromatic residues is small [111], since these aromatic residues are conserved among archaeal rhodopsins. This contribution is, however, of importance for the opsin shift as shown by previous calculations [172, 173].

(iii) FTIR spectroscopy [188] as well as theoretical calculations [111] show that the chromophore geometry is nearly identical in BR and SRII. Therefore, it seems likely that the major contribution to the inter-protein shift is due to a different distribution of charges and permanent dipoles [111, 112, 188].

4.1.2 QM/MM CALCULATIONS AND MUTATIONAL STUDIES

For the accurate calculation of excitation energies, high-level quantum chemical methods are required. However, present day limitations of computer resources allow only small molecules to be treated with these methods. The environment of the quantum-mechanically treated part may be included using a molecular mechanics force field. The investigation of the influence of this environment is, however, limited by the accuracy of

the force field representation. Apart from technical details such as the size of the quantum region and its linkage to the classical region, the treatment of the environment in quantum mechanics/molecular mechanics (QM/MM) remains challenging.

Studies using QM/MM methods were able to account for major parts of the observed inter-protein shift between BR and SRII [111, 112]. These studies each identified a single residue of the counterion complex to cause the major part of the inter-protein shift. However, these studies disagree on the nature of this residue, identifying either Arg72 in SRII that corresponds to Arg82 in BR or Asp201 in SRII that corresponds to Asp212 in BR as determinant for the inter-protein shift.

Mutational studies have shown that residues close to the retinal that do not belong to the counterion contribute by about 35% to the inter-protein shift between BR and HR [180]. This experimental finding was not reproduced by the QM/MM studies mentioned above. The structural elements responsible for the inter-protein shift remain, thus, highly debated.

4.1.3 OBJECTIVE

The calculations presented in this chapter aim at a better understanding of the electrostatic environment of retinal by calculating the electrostatic potentials created by the protein-membrane system at the retinal. The analysis focuses on the structural origin of the difference in electrostatic potential between three archaeal rhodopsins: BR, HR and SRII. HR which absorbs at a similar maximum as BR is included in the calculation, thereby allowing all results to be tested for similarities between BR and HR and differences between BR and SRII. The electrostatic potential is decomposed into contributions of individual residues which allows the identification of all residues that differ in their electrostatic potential between BR, HR and SRII. The electrostatic potential at the retinal is then related to the absorption maxima of BR, HR and SRII by a generalized model of a particle in a box that includes the electrostatic potential as a parameter.

4.2 COMPUTATIONAL DETAILS

The calculations were performed on the ground states of three archaeal rhodopsins: the BR structure with the PDB code 1qhj, the HR structure with the PDB code 1e12 and the SRII structure with the PDB code 1jgj (*cf.* Table 3.1) [41, 46, 48]. A prerequisite for the calculation of the electrostatic potential of a protein is the determination of its protonation state. Therefore, the lowest energy protonation state at pH = 7 was calculated for BR, HR and SRII. Aspartate, arginine, cysteine, glutamate, histidine, lysine and tyrosine residues, the retinal Schiff base and the palmitic acid of HR were considered as protonatable sites. The pK_{intr} (*cf.* Eq. (2.17)) and $W_{\mu\nu}$ (*cf.* Eq. (2.20)) were calculated as described in Chapter 2 and Chapter 3. The lowest energy protonation state can then be determined using dead-end elimination (DEE). The DEE method is explained in detail in Chapter 6.

For the calculation of the electrostatic potential, the lowest energy protonation state of BR, HR and SRII, respectively, is used. The retinal charges are set to zero, allowing for a

BR		HR		SRII		BR		HR		SRII	
		R24	1					C145	1		
		E25	0			K129	1				
R7	1									E122	0
E9	0					R134	1	R161	1	R123	1
K30	1							C169	1		
		R52	1	R27	1			D182	0		
				D28	0					E147	0
		R55	1			K159	1				
D36	0					E161	0				
D38	0	R58	1	E33	0					R152	1
K40	1	R60	1	R34	1	R164	1				
K41	1			R35	1	E166	0				
		E90	0							K157	1
		H95	0					E194	0		
E74	0	E100	0	E65	0	K172	1	D197	0		
				R66	1					R162	1
		R103	1			R175	1	R200	1	R164	1
R82	1	R108	1	R72	1	E194	0	E219	0		
D85	0			D75	0	E204	1			D193	0
D96	1					D212	0	D238	0	D201	0
D102	0	D128	0			R225	1	R251	1	D214	0
D104	0	D130	0	D94	0	R227	1				
				R96	1			E257	0		
				E97	0			R258	1		
D115	1	D141	1			E232	0				

Table 4.1. Lowest energy protonation state of BR, HR, and SRII. The states are calculated at pH=7; 0 indicates that the residue is deprotonated, 1 that the residue is protonated. Corresponding residues are aligned (*cf.* Figure 1.5 in Chapter 1). The retinal Schiff base is protonated in all three proteins. The palmitic acid of HR is protonated. The listed protonation states were used for the calculation of the electrostatic potential at the retinal.

comparison between the different protein environments to which the retinal is exposed. The electrostatic potential is calculated as described in Chapter 2 using the program POTENTIAL of MEAD [168]. Three levels of focusing were used for the finite-difference lattices: first, a coarse grid with a spacing of 1 Å, followed by a finer grid with a spacing of 0.5 Å and last a grid with a spacing of 0.25 Å. The grids were centered on the geometric center of the molecule. The temperature was set to 300 K and the ionic strength to 0.1 M. The resulting output is a field format that can be used to obtain the potential at any point in space, *e.g.*, at the molecular surface of retinal and also at the atomic centers of the retinal π -system.

4.3 PROTONATION STATE OF BR, HR AND SRII

The protonation states of lowest energy at pH=7 are listed in Table 4.1 for BR, HR and SRII. The protonation states are in excellent agreement with experimental data and previous calculations of protonation states [46, 166, 189–192]. The retinal Schiff base was

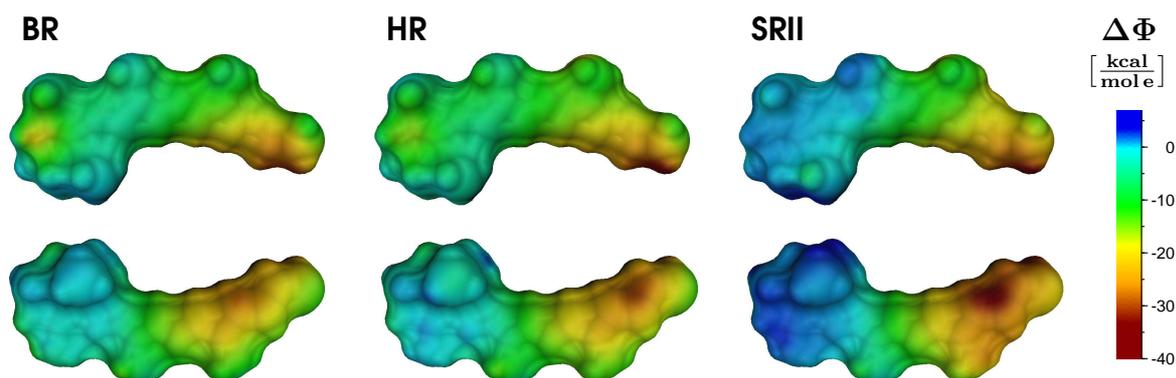


Figure 4.3. Electrostatic potential at the surface of retinal. The potential of the protein at the molecular surface of the retinal Schiff base is depicted for BR, HR and SRII. Both sides of the retinal are shown. The β -ionone ring is on the left and the Schiff base on the right. The magnitude of the potential is color coded. The molecular surface of retinal is calculated with MSMS using a probe sphere radius of 1.3Å [193].

found to be protonated in all three proteins. Asp115 in BR is protonated and also the corresponding Asp141 in HR which is in agreement with their location in a highly apolar region. Apart from Asp115 two other protonatable residues of BR are not in their reference state at pH=7: Asp96 and Glu204 were found to be protonated. Glu204 and another glutamate residue, Glu194, which are in close proximity and face each other, are part of the proton release group. In the ground state of BR, one proton is bound to the proton release group that is released to the extracellular site during the photocycle. However, the exact location of this proton is still under debate. Spectroscopic measurements as well as calculations indicate that this proton is localized within a network of internal water molecules which are stabilized by Glu194 and Glu204 [62, 63, 166]. However, since the distance between the proton release group and the retinal is large, the exact location of the proton within the proton release group has little influence on the electrostatic potential at the retinal Schiff base. The position of these residues is shown in Figure 1.7 in Chapter 1. In HR, the protonation of all protonatable residues, with the exception of Asp141, corresponds to the reference protonation form at pH=7. For SRII, all protonatable residues were found to be in their reference protonation form at pH=7.

4.4 ELECTROSTATIC POTENTIAL OF BR, HR AND SRII AT THE RETINAL

Figure 4.3 shows the electrostatic potential of BR, HR and SRII at the molecular surface of the retinal Schiff base. The general features of the potential are similar for all three proteins: the potential is close to neutral in the β -ionone ring region and negative in the Schiff base region. The difference between the potential in the β -ionone ring region and the potential in the Schiff base region is, however, more pronounced in SRII than in BR and HR. Thus, this representation indicates a greater similarity for BR and HR

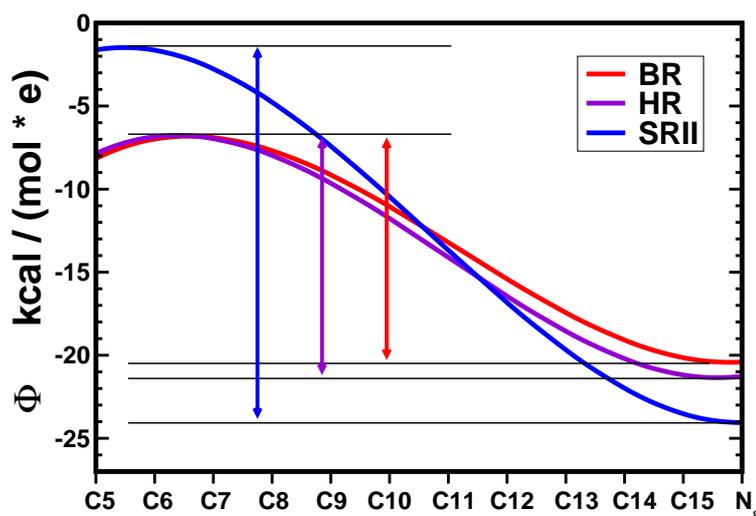


Figure 4.4. Electrostatic potential at the atomic centers of retinal. The electrostatic potential is plotted along the retinal π -system. The labels on the x-axis refer to the atoms of the π -system (*cf.* Figure 4.2). The β -ionone ring is on the left and the Schiff base on the right. Arrows indicate the magnitude of the difference in potential between Schiff base and β -ionone ring. A polynomial regression was used to display the potentials.

as compared to SRII. This observation is in agreement with the small inter-protein shift between BR and HR as compared to the large inter-protein shift between BR and SRII (*cf.* Figure 4.1).

To quantify the difference between the electrostatic potentials of the three archaeal rhodopsins, the electrostatic potentials were calculated at the atomic centers of the π -system of retinal. The resulting potentials are shown in Figure 4.4. As can be seen, the features of the potential at the retinal surface presented in Figure 4.3 are reproduced.

Two aspects of the electrostatic potential will be discussed: (i) the variation of the potential along the retinal π -system in one protein and (ii) the difference between the potentials of the archaeal rhodopsins. In all three proteins, the electrostatic potential changes from close to neutral at the β -ionone ring to negative at the Schiff base. The main dissimilarity between the archaeal rhodopsins is the magnitude of the potential difference between the β -ionone ring and the Schiff base. This difference is about $10 \text{ kcal mol}^{-1} e^{-1}$ greater in SRII than in BR and HR, as indicated in Figure 4.4.

RELATION BETWEEN ELECTROSTATIC POTENTIAL AND INTER-PROTEIN SHIFT

It is known from experiments that the dipole moment of a protonated retinal Schiff base changes upon excitation. The π -electrons move towards the Schiff base and the positive charge of the Schiff base delocalizes over the retinal [194]. Interactions that stabilize the excited state, *i.e.*, the charge delocalized state, lower the excitation energy, thereby, shifting the absorption maximum to the red. In contrast, interactions that stabilize the positive charge on the Schiff base lead to a blue shift of the absorption maximum.

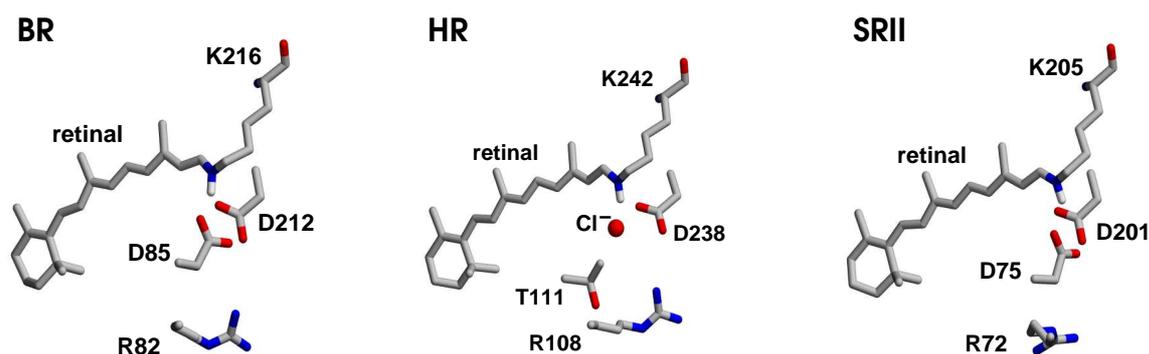


Figure 4.5. The Counterion complex of BR, HR and SRII. Two negative charges are located beneath the positively charged protonated retinal Schiff base. These are in the case of BR and SRII two aspartate residues, *i.e.*, Asp85 and Asp212 as well as Asp75 and Asp201, and in the case of HR one aspartate residue, *i.e.*, Asp238, and one chloride ion. A positively charged arginine residue is located beneath the negative charges in all three proteins, *i.e.*, Arg82 in BR, Arg108 in HR and Arg72 in SRII. Together, the counterion complex and the protonated retinal Schiff base are neutral.

Due to the more pronounced difference between the electrostatic potential at the β -ionone ring and the electrostatic potential at the Schiff base in SRII, the positive charge on the Schiff base is more stabilized in SRII as compared to BR and HR. Thus, in SRII, the transition energy between the ground state and the first excited state is increased compared to BR and HR. This implication is consistent with the measured absorption spectra of the three archaeal rhodopsins depicted in Figure 4.1.

4.5 DECOMPOSITION OF THE ELECTROSTATIC POTENTIAL

To understand the structural basis of the differences between BR, HR and SRII, the electrostatic potential is decomposed into individual components. In particular, the influence of the Schiff base counterion and the residues of the retinal binding pocket is analyzed. In addition, all residues that contribute significantly to the difference in electrostatic potential between the archaeal rhodopsins are identified.

4.5.1 THE COUNTERION COMPLEX

The counterion complex of BR, HR and SRII is depicted in Figure 4.5. In all three proteins, the counterion of the positively charged protonated Schiff base consists of an arrangement of one positive and two negative charges. The positive charge originates from an arginine in all three proteins. The negative charges originate either from two aspartate residues in the case of BR and SRII, or from an aspartate and a chloride ion in the case of HR. The chloride ion in HR takes the position of the missing aspartate carboxyl group. The aspartate itself is substituted by a threonine.

	Asp85 ^a			Asp212 ^a			Arg82 ^a			
	O _{δ1}	O _{δ2}	aver.	O _{δ1}	O _{δ2}	aver.	N _ε	N _{η1}	N _{η2}	aver.
BR	3.7	4.2	4.0	4.0	5.2	4.6	10.8	8.5	10.7	10.0
HR	3.8 ^b		3.8	3.5	4.3	3.9	9.1	7.2	9.4	8.6
SRII	4.2	4.0	4.1	3.8	4.9	4.4	9.3	10.9	11.2	10.5

^aThe residue numbers refer to BR. Asp85 in BR corresponds to Thr111 and a chloride ion in HR and to Asp75 in SRII. Asp212 in BR corresponds to Asp238 in HR and to Asp201 in SRII. Arg82 in BR corresponds to Arg108 in HR and Arg72 in SRII.

^bThe distance between the chloride ion and the Schiff base nitrogen atom in HR.

Table 4.2. Distance between the Schiff base and the counterion complex. The table lists the distances between the nitrogen atom of the retinal Schiff base and selected atoms of the counterion residues of BR, HR and SRII. The atoms were chosen to represent the ionic charge of the respective amino acid. All distances are given in Å.

THE COUNTERION EFFECT

It is known from model compounds that changes in the distance between the Schiff base and the counterion complex influence the absorption maximum [179, 195–197]. A shorter distance, *i.e.*, a stronger interaction, between the Schiff base and its counterion causes a blue-shift of the absorption maximum. However, an analysis of the X-ray structures of the archaeal rhodopsins does not reveal significant differences between the proteins. The distances between charged atoms of the counterion complex and the Schiff base are given in Table 4.2. The main difference between SRII and BR is the orientation of the arginine side chain (*cf.* Figure 4.5). In a QM/MM calculation, it was found that upon structure optimization Asp201 of SRII moves towards the retinal Schiff base [112]. As shown in Hayashi *et al.* [112] such a displacement would imply that Asp201 dominates the inter-protein shift between SRII and BR. Similar calculations by Ren *et al.* [111] showed that without this displacement Asp201 loses its dominant role. In contrast to Hayashi *et al.*, Ren *et al.* suggested that the orientation of Arg72 is the main cause for the inter-protein shift between SRII and BR.

ELECTROSTATIC POTENTIAL OF THE COUNTERION

To quantify the interaction between the counterion complex and the retinal, the electrostatic potential of the counterion at the retinal was calculated for BR, HR and SRII. All backbone and side chain atoms of the counterion residues as well as the chloride ion were included in these calculations. As can be seen in Figure 4.6, the electrostatic potential caused by the counterion is very similar in all three proteins. A comparison between Figure 4.5 and 4.6 reveals that the main contribution to the total electrostatic potential at the retinal originates from the counterion residues. This observation is not surprising, since the only charged residues close to the retinal are the counterion residues. However, the counterion complex is not the origin of the observed difference between the electrostatic potentials of BR and HR on the one hand and SRII on the other hand. A comparison of the potentials of each individual counterion residue leads to the same

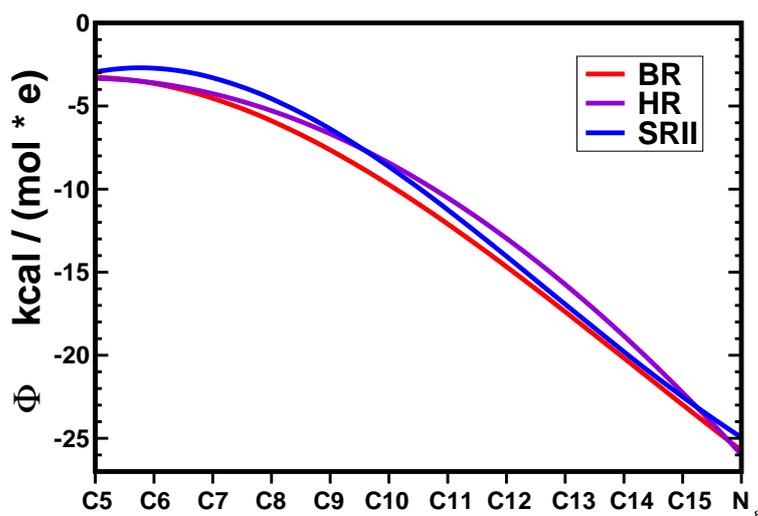


Figure 4.6. Electrostatic potential of the counterion. The electrostatic potential is plotted along the retinal π -system. The labels on the x-axis refer to the atoms of the π -system (cf. Figure 4.2). The β -ionone ring is on the left and the Schiff base on the right. A polynomial regression was used to display the potentials.

conclusion. The dominant role of a single counterion residue found in QM/MM calculations cannot be confirmed by these calculations of the electrostatic potential resting on the known crystal structures [111, 112].

4.5.2 THE RETINAL BINDING POCKET

In the next step, the influence of the residues in the near retinal environment on the retinal was analyzed. Residues located within 5 Å of the heavy atom centers of the retinal Schiff base, *i.e.*, atoms C1 to C20 and the Schiff base nitrogen atom (cf. Figure 4.2), are considered to form the retinal binding pocket (cf. Figure 1.6 in Chapter 1).

The residues of the binding pocket are highly conserved among the archaeal rhodopsins (see highlighted residues in Figure 1.5 in Chapter 1). Three differences between SRII and the other two proteins, BR and HR, are, however, notable:

1. Thr204 in SRII which is located close to the Schiff base is an alanine residue in BR and in HR, *i.e.*, Ala215 and Ala241, respectively.
2. Gly130 in SRII which is located close to the β -ionone ring is a serine residue in BR and in HR, *i.e.*, Ser141 and Ser168, respectively.
3. Ala131 in SRII which is also located close to the β -ionone ring is a threonine residue in BR and a cysteine in HR, *i.e.*, Thr142 and Cys169, respectively.

Thus, in SRII, the β -ionone ring region is more apolar and the Schiff base region more polar than in BR and HR. The increased polarity of the ring region found for BR and HR might facilitate charge dislocation upon excitation, thereby lowering their excitation energy compared to SRII [109, 112].

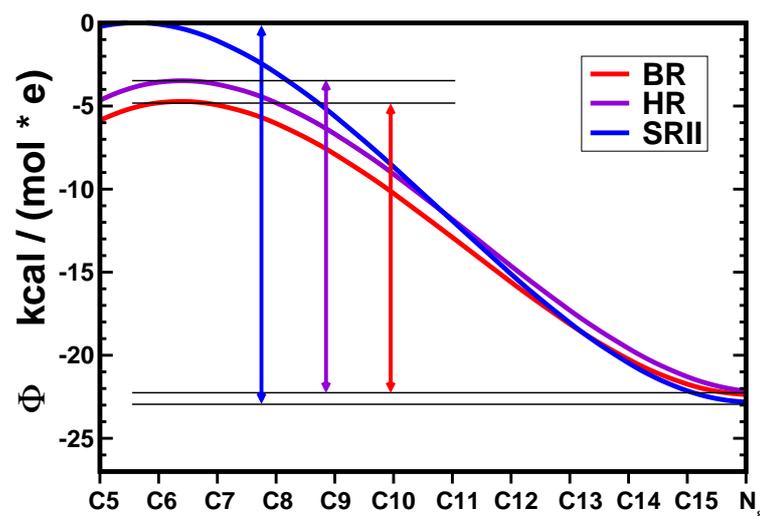


Figure 4.7. Electrostatic potential of the retinal binding pocket. The electrostatic potential is plotted along the retinal π -system. The labels on the x-axis refer to the atoms of the π -system (cf. Figure 4.2). The β -ionone ring is on the left and the Schiff base on the right. Arrows indicate the magnitude of the difference in potential between Schiff base and β -ionone ring. A polynomial regression was used to display the potentials.

The electrostatic potential caused by the residues of the binding pocket is shown in Figure 4.7. The potential of Arg82, Arg108 and Arg72 is added to the potential of the binding pocket residues of BR, HR and SRII, respectively, although this arginine is located more than 8 Å away from the retinal (see Table 4.2) and is, thus, not part of the binding pocket. However, this arginine contributes largely to the potential at the retinal but not to the difference between BR, HR and SRII as established in the previous subsection. This residue is included to allow a quantitative comparison between the graphs shown in Figure 4.4 and 4.7.

As depicted in Figure 4.7, the difference between the electrostatic potential at the β -ionone ring and the Schiff base is about $5 \text{ kcal mol}^{-1} e^{-1}$ greater in SRII than in BR and HR due to the binding pocket residues. If the complete protein is considered, this difference is about $10 \text{ kcal mol}^{-1} e^{-1}$ greater in SRII than in BR and HR (cf. Figure 4.4). Thus, the retinal binding pocket contributes about 50% to the dissimilarity in electrostatic potential between the archaeal rhodopsins. Shimono *et al.* [180] showed that the mutation of the retinal binding pocket in SRII to the wild type binding pocket of BR leads to a 25 nm red-shift of the absorption maximum, thereby accounting for about 35% of the inter-protein shift between BR and SRII. In agreement with these experiments, the presented result suggests that the residues of the binding pocket do play an important role for the differences between BR, HR and SRII. Furthermore, the experiments and these calculation indicate that residues outside the binding pocket significantly contribute to the different absorption maxima of BR, HR and SRII.

4.5.3 POTENTIAL OF INDIVIDUAL RESIDUES

To identify all residues that contribute significantly to the dissimilarity between the electrostatic potentials of the proteins, the contribution of each residue individually was analyzed. The structures of BR, HR and SRII were superposed to align corresponding residues (*cf.* Figure 1.5 in Chapter 1). The loop regions of the archaeal rhodopsins differ in structure and number of residues. Therefore, not all loop residues could be aligned. The contribution of these residues to the electrostatic potential at the retinal is, however, negligible. First, the residues of the binding pocket, which are expected to have the strongest influence because of their close proximity to the retinal, are discussed. In the following all other residues are analyzed.

RESIDUES OF THE RETINAL BINDING POCKET

Two residues were found to contribute significantly to the difference between the potentials of BR, HR and SRII. These two residues are Ser141 in BR that corresponds to Ser168 in HR and Gly130 in SRII and Ala215 in BR corresponding to Ala241 in HR and Thr204 in SRII. A smaller contribution was found for Thr142 in BR, which corresponds to Cys169 in HR and Ala131 in SRII. All three residues were discussed in the previous section to cause a different polarity in the binding pockets of BR, HR and SRII.

As can be seen in Figure 4.8 a, Thr204 in SRII generates a negative potential close to the Schiff base, whereas Ala215 in BR and Ala241 in HR generate a slightly positive potential close to the Schiff base. A minor contribution of Thr204 to the inter-protein shift has also been found in QM/MM studies [111, 112]. As shown in Figure 4.8 b, Gly130 in SRII does not contribute significantly to the potential along the retinal π -system, whereas Ser141 in BR and Ser168 in HR generate a negative potential close to the β -ionone ring. As depicted in Figure 4.8 c, Ala131 in SRII and Cys169 in HR do not contribute significantly to the potential along the retinal π -system, whereas Thr142 in BR generates a negative potential close to the β -ionone ring. Thus, for the corresponding residues Thr142 in BR, Cys169 in HR and Ala131 in SRII, a different potential is observed for BR and SRII, while the potential of HR is similar to the potential of SRII and not BR. However, the contribution to the difference between the proteins is small compared for example to the difference between Ser141 in BR and Ala215 in BR and the corresponding residues in HR and SRII (*cf.* Figure 4.8 a and b). None of the other binding pocket residues shows significant differences between the proteins.

The order of the binding pocket residues by magnitude of contribution to the difference between BR, HR and SRII is:

1. Ala215 in BR, Ala241 in HR and Thr204 in SRII
2. Ser141 in BR, Ser168 in HR and Gly130 in SRII
3. Thr142 in BR, Cys169 in HR and Ala131 in SRII

Shimono *et al.* [109, 180] measured the absorption of several mutants of SRII, where single or multiple binding pocket residues of SRII are mutated to the corresponding residues of BR. The order of the SRII single mutants by magnitude of observed red-shift is: 1. Thr204Ala, 2. Gly130Thr, and 3. Ala131Thr [180]. Thus, the differences observed

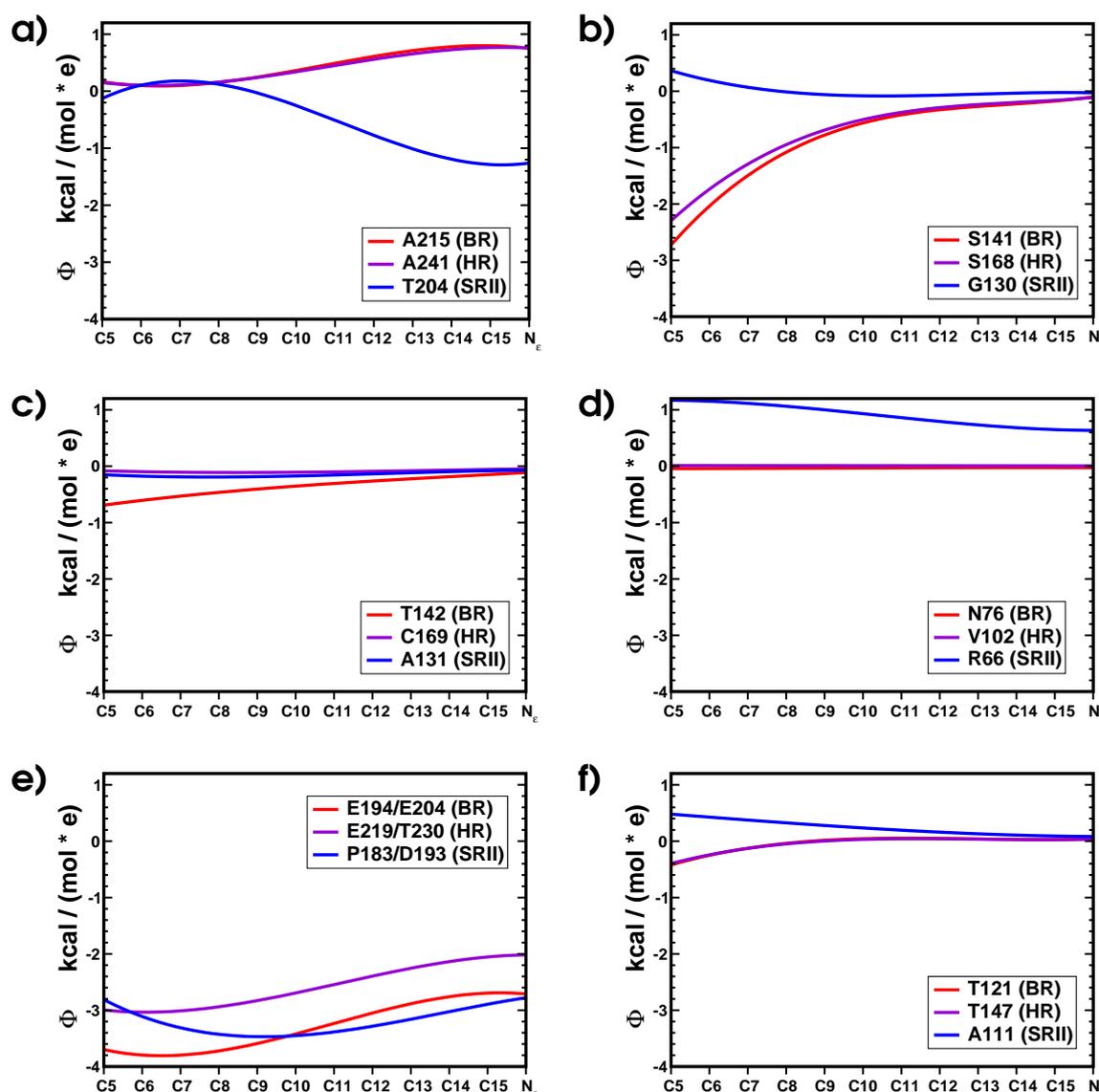


Figure 4.8. Electrostatic potential of individual residues. The electrostatic potential is plotted along the retinal π -system for **a)** Ala215, Ala241 and Thr204, **b)** Ser141, Ser168 and Gly130, **c)** Thr142, Cys169 and Ala131, **d)** Asn76, Val102 and Arg66, **e)** the sum of the potentials of Glu194 and Glu204, Glu219 and Thr230 and Pro183 and Asp193 and **f)** for Thr121, T147 and Ala111. The labels on the x-axis refer to the atoms of the π -system (*cf.* Figure 4.2). The β -ionone ring is on the left and the Schiff base on the right. A polynomial regression was used to display the potentials.

between BR, HR and SRII and the order by magnitude of contribution is in excellent agreement with the mutational studies. Shimono *et al.* [180] found two other mutations that red-shift the spectrum of SRII by as much as the Ala131Thr mutant: Ile43Val and Val108Met. For these residues significant differences in electrostatic potential between BR, HR and SRII could be observed. However, since isoleucine and methionine have large side chains, whereas valine has not, it cannot be excluded that these mutations significantly affect the structure of SRII. Furthermore, Shimono *et al.* [109] generated the dou-

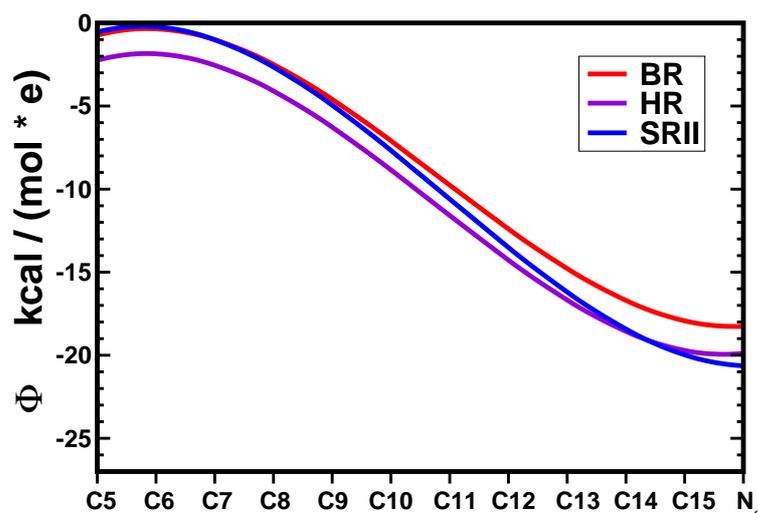


Figure 4.9. Electrostatic potential of the protein omitting seven residues. The residues that were found to cause the difference between the electrostatic potential at the retinal are subtracted from the potential of the protein (*cf.* Figures 4.4 and 4.8). The electrostatic potential is plotted along the retinal π -system. The labels on the x-axis refer to the atoms of the π -system (*cf.* Figure 4.2). The β -ionone ring is on the left and the Schiff base on the right. A polynomial regression was used to display the potentials.

ble mutant Gly130Ser/Thr204Ala of SRII. The absorption maximum of the resulting SRII mutant is 512 nm. Thus, about 20% of the inter-protein shift are compensated [109]. The results presented here underlines the important role of Gly130 and Thr204 in SRII for the inter-protein shift, whereas Ala131 in SRII seems less important.

RESIDUES OUTSIDE OF THE BINDING POCKET

In the next step, those residues were identified for which the potentials differ by more than $1 \text{ kcal mol}^{-1} e^{-1}$ anywhere along the π -system of retinal. A group of eleven such residues was found. Ser141 in BR corresponding to Ser168 in HR and Gly130 in SRII and also Ala215 in BR corresponding to Ala241 in HR and Thr204 in SRII belong to this group. These are discussed in the previous subsection (*cf.* Figure 4.8 a and 4.8 b).

Three more residues that contribute to the dissimilarity between the archaeal rhodopsins could be identified. These residues are Asn76 in BR corresponding to Val102 in HR and Arg66 in SRII and Glu194 and Glu204 in BR corresponding to Glu219 and Thr230 in HR and to Pro183 and Asp193 in SRII. These three residues are located more than 8 \AA away from the retinal.

As shown in Figure 4.8 d, the electrostatic potential of Arg66 in SRII is positive along the π -system of retinal with a maximum close to the β -ionone ring, whereas the potential of the corresponding residues in BR and HR, *i.e.*, Asn76 and Val102, respectively, is neutral along the retinal π -system.

Glu194 and Glu204 in BR which are located close and adverse to each other are part of the proton release group. Here, the proton resides on Glu204 (*cf.* Table 4.1). Thus,

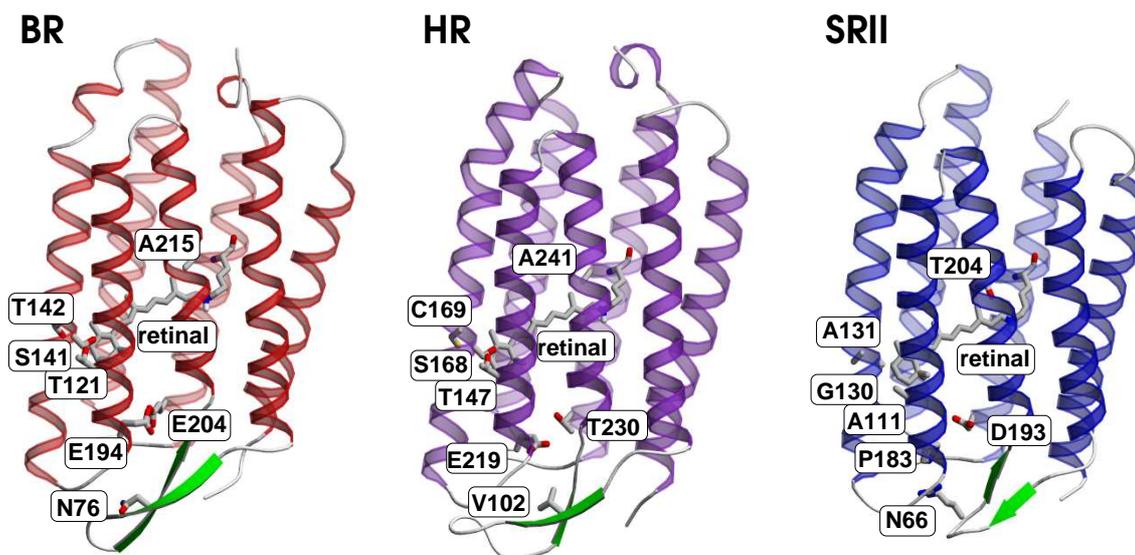


Figure 4.10. Structure of BR, HR and SRII. The retinal, the lysine residue that binds the retinal covalently and the residues that contribute to the different potential along the retinal π -system are depicted explicitly.

Glu204 is neutral and Glu194 is negatively charged in this BR structure. Glu219 in HR, corresponding to Glu194 in BR is also charged, whereas SRII has a proline residue, *i.e.*, Pro183, at this position. However, Asp193 in SRII is negatively charged, whereas the corresponding Glu204 in BR is protonated, and Thr230 in HR is uncharged. Since all three proteins carry one negative charge in the region of the proton release group of BR, these two residues are regarded as one group. Therefore, the sum of the potentials of Glu194 and Glu204 in BR is compared to the sum of the potentials of Glu219 and Thr230 in HR and to the sum of the potentials of Pro183 and Asp193 in SRII. As shown in Figure 4.8 e, the potential caused by these two residues is more positive along the retinal π -system in HR than in BR, whereas the potential in SRII is similar to the potential in BR in the Schiff base region and significantly more positive at the β -ionone ring. The other residues in this group of eleven residues does not exhibit significant differences between the three proteins.

Additionally, residues that differ by more than $0.5 \text{ kcal mol}^{-1} e^{-1}$ anywhere along the retinal π -system were identified. 26 such residues were found, among them the eleven residues that differ by more than $1 \text{ kcal mol}^{-1} e^{-1}$. The comparison of the individual residues revealed one additional residue that contributes significantly to the dissimilarity between the electrostatic potentials of the archaeal rhodopsins. As depicted in Figure 4.8 f, the electrostatic potential of Ala111 in SRII is nearly $1 \text{ kcal mol}^{-1} e^{-1}$ more positive close to the β -ionone ring than the potential of the corresponding residues in BR and HR, *i.e.*, Thr121 and Thr147. Ala111 is located close to the β -ionone ring just outside the retinal binding pocket.

The potentials of the seven residues that show significant differences between BR, HR and SRII are subtracted from the total potential of the proteins. As shown in Figure 4.9,

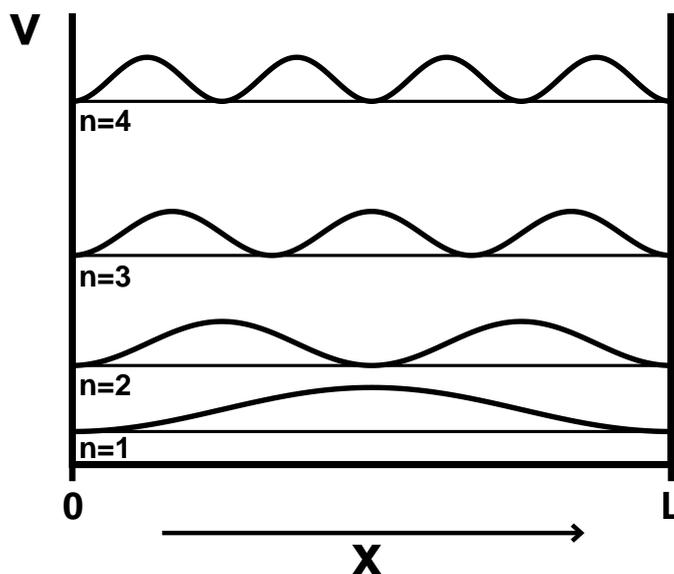


Figure 4.11. Standard model of a particle in a box. The walls of the box at $x=0$ and $x=L$, where L is the length of the box, are impermeable. V is the potential energy. The allowed energy levels at $n=1$ to 4 are shown together with the respective wave function.

virtually the complete dissimilarity between the potentials of the complete proteins is accounted for by these residues. The difference between the electrostatic potential of BR, HR and SRII is a combined effect of these seven residues. The position of these residues in the structure of BR, HR and SRII is shown in Figure 4.10.

4.6 QUALITATIVE MODEL OF THE INTER-PROTEIN SHIFT

Light absorption and emission by atoms or molecules are quantum mechanical processes. The allowed energy states of the system are described by Schrödinger's equation:

$$H\Psi = E\Psi, \quad (4.1)$$

where E is the energy. The wave function Ψ can in principle be understood as a representative of the location of the particle. H is the Hamilton operator which is the observable corresponding to the total energy of the system. In general, no simple solutions of Schrödinger's equation exist. It is, therefore, far from trivial to derive the absorption spectrum of a molecule from its molecular structure and to calculate consistently the effect of the environment on the spectrum [198–204].

For polyenes, however, the absorption maximum can be qualitatively understood from a model of a quantum mechanical particle located in a box [104, 205]. Since the π -electrons of polyenes are located in a diffuse cloud around the conjugated system, they can be assumed to perform a one-dimensional translational motion along the π -system.

4.6.1 MODEL OF A PARTICLE IN A BOX

Figure 4.11 shows a one-dimensional box with impenetrable borders at $x=0$ and $x=L$. The potential energy of the particle is infinitely high beyond the borders. The probability of the particle to be beyond these borders is zero. Thus, the wave function Ψ must be equal to zero for $x \leq 0$ and $x \geq L$. In the interval $[0, L]$, the wave function takes the form:

$$\Psi = A \sin(kx) , \quad (4.2)$$

where k is the wave number. At $x=0$, Eq. (4.2) is obviously zero, at $x=L$ it becomes zero if $\sin(kL)=0$, which holds true if kL is a whole-numbered multiple of π . It follows that the allowed values for k are those for which $kL=n\pi$ holds with $n=1, 2, \dots$. The energy is consequently quantized. The connection between k and the energy E is given by:

$$E = \frac{k^2 \hbar^2}{2m} , \quad (4.3)$$

where m is the mass of the particle and \hbar denotes Planck's constant h divided by 2π . Substituting k by $n\pi/L$, the allowed energy values of the particle are obtained:

$$E = \frac{n^2 \hbar^2 \pi^2}{2mL^2} . \quad (4.4)$$

In the standard model of a particle in a box, the allowed energy levels of a particle depend on its mass m and on the length of the box L . The larger the value of L , the smaller the difference between the allowed energy levels.

For polyenes, the box extension L corresponds to the length of the π -system and the mass of the particle is the electron mass m_e . The excitation energy is given by the energy difference between the first two energy levels that correspond to the ground state S_0 and to the first excited state S_1 , respectively. According to this model, the excitation energy of a polyene can be calculated in dependence on the extension of the π -system. In analogy, a generalized model of a particle in a box is formulated that qualitatively relates the electrostatic potential of the archaeal rhodopsins to their absorption maxima.

4.6.2 MODEL OF A PARTICLE IN A BOX WITH STEP POTENTIAL

In comparison to the standard model of a particle in a box, the environment, here the protein potential, is included as an additional parameter. The electrostatic contribution of the environment is approximated by a step potential. The allowed energy levels of a particle depend on two additional parameters, the height of the potential step and its position. A schematic picture of the generalized model is shown in Figure 4.12. The box length L represents the extension of the π -system. The potential step is located at position a , where $0 < a < L$. The height of the potential step is given by $\Delta\Phi$. For a particle of charge q located in this potential, the stationary Schrödinger equation is given by:

$$E \Psi(x) = -\frac{\hbar^2}{2m_e} \frac{\partial^2}{\partial x^2} \Psi(x) + V(x) \Psi(x) , \quad (4.5)$$

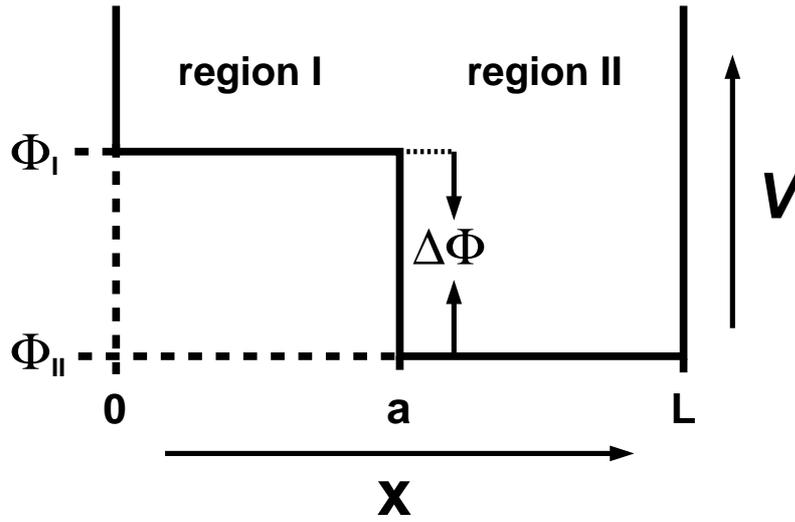


Figure 4.12. Generalized model of a particle in a box with potential step. L denotes the length of the box, a the position of the step potential, Φ_I and Φ_{II} are the additional potentials in region I and region II, respectively, and $\Delta\Phi$ denotes the height of the potential step.

where E is the energy of the particle, Ψ its wave function, \hbar denotes Planck's constant divided by 2π and m_e is the mass of the particle, *i.e.*, here the electron mass. The potential energy V is given by:

$$V(x) = \begin{cases} q\Phi_I & 0 < x \leq a \\ q\Phi_{II} & a < x < L \\ \infty & \text{elsewhere} \end{cases}, \quad (4.6)$$

where Φ_I and Φ_{II} denote the potential of the environment in region I and region II, respectively. The boundary conditions $\Psi(0) = \Psi(L) = 0$ together with the continuity of the wave function and its derivative at $x = a$ lead to a quantization condition for the allowed energy levels:

$$\frac{1}{k_1} \frac{\exp(ik_1 a) - \exp(-ik_1 a)}{\exp(ik_1 a) + \exp(-ik_1 a)} = \frac{1}{k_2} \frac{\exp(ik_2(a-L)) - \exp(-ik_2(a-L))}{\exp(ik_2(a-L)) + \exp(-ik_2(a-L))}, \quad (4.7)$$

where k_1 and k_2 are the wave numbers in region I and region II, respectively, given by:

$$k_j = \sqrt{2m_e/\hbar^2(E - q\Phi_j)}. \quad (4.8)$$

The term $(E - q\Phi_j)$ can become negative for one of the two regions. In that case, the corresponding wave number k_j is a complex number: $k_j = i\rho$, with $\rho = \sqrt{2m_e/\hbar^2|E - q\Phi_j|}$. The corresponding solution decays exponentially, reflecting the fact that the region is classically forbidden.

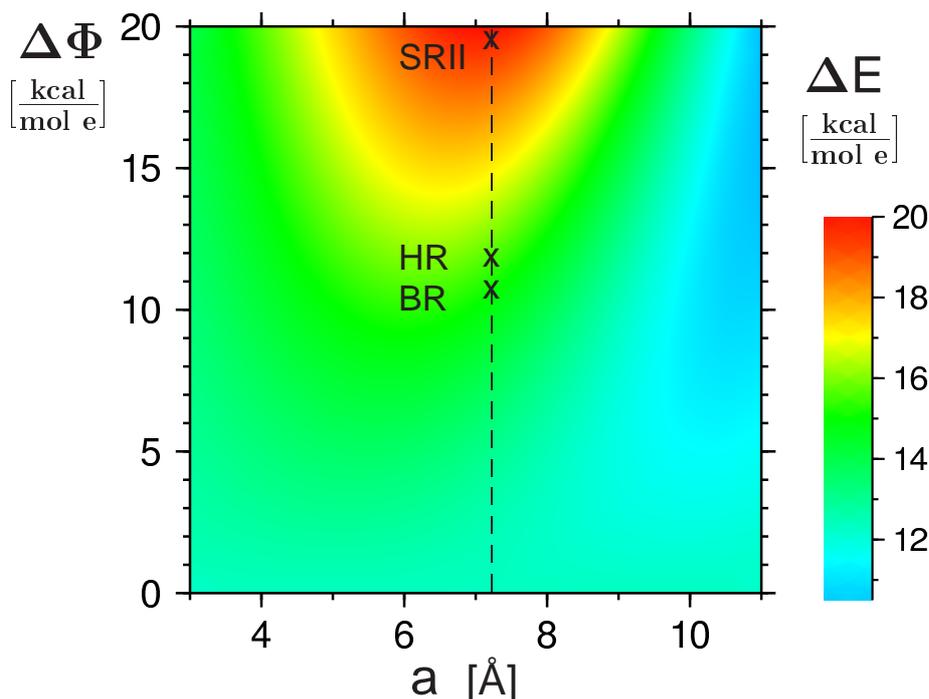


Figure 4.13. Dependence of the excitation energy ΔE on the position of the step a and the potential difference $\Delta\Phi$. The length of the box L is set to 14.5 \AA . The excitation energy ΔE , given in $\text{kcal mol}^{-1} \text{ e}^{-1}$, is indicated by the color code. The dashed line marks the position of the step in the archaeal rhodopsins at $a=7.25\text{ \AA}$. The crosses indicate the height of the potential step $\Delta\Phi$ for BR,HR and SRII.

Eq. (4.7) can be solved numerically leading to a discrete energy spectrum of the system. Two limiting cases can be distinguished:

- (i) When the step is located close to one of the potential walls of the box, *i.e.*, a is close to 0 or L , the model approaches the standard model of a particle in a box. Consequently, the dependence of ΔE , *i.e.*, the energy difference between two allowed energy levels, on a and $\Delta\Phi$ is low in this limit.
- (ii) When the height of the potential step $\Delta\Phi$ approaches zero, a similar effect is seen and the behavior of Eq. (4.7) resembles again the standard model of a particle in a box.

Figure 4.13 shows the excitation energy ΔE for a positively-charged particle. The positive charge corresponds to the charge of the retinal Schiff base, which delocalizes over the π -system upon excitation. The box length L was set to 14.5 \AA , representing the extension of the retinal π -system in the archaeal rhodopsins. The step position a is varied from 3 to 11 \AA and the height of the potential step $\Delta\Phi$ varies from 0 to $20\text{ kcal mol}^{-1} \text{ e}^{-1}$. The qualitative picture offered by this model reveals an interesting aspect of spectral tuning. As can be seen, spectral tuning is most effective, if the controlling potential changes close to the center of the polyene π -system: then, even small changes in the height of the potential step $\Delta\Phi$ have a pronounced influence on the excitation energy ΔE .

APPLICATION TO BR, HR AND SRII

For the electrostatic potential of the three archaeal rhodopsins, two plateaus can be distinguished (*cf.* Figure 4.4): one in the β -ionone ring region and another one in the Schiff base region. The separation between these two plateaus is located approximately in the middle of the retinal π -system. The position of the step for the archaeal rhodopsins is indicated by the dashed line in Figure 4.13. The height of the potential step $\Delta\Phi$ was defined as the difference between the potential at the β -ionone ring and the potential at the Schiff base and is indicated by the crosses for BR, HR and SRII. The box length L was set to 14.5 Å, corresponding to the extension of the retinal π -system.

As can be seen in Figure 4.13, the excitation energy ΔE is significantly higher for SRII than for BR and HR, which have similar excitation energies. Using the generalized model of a particle in a box with step potential the observed differences in electrostatic potential in BR, HR and SRII are related to differences in the absorption maxima. The general trend of the experimentally measured absorption maxima could be reproduced by grouping BR and HR together and showing a significantly higher excitation energy for SRII (*cf.* Figure 4.1).

4.7 CONCLUDING REMARKS

In this chapter, the electrostatic potential at the retinal π -system was analyzed for three archaeal rhodopsins. As discussed in Section 4.1, electrostatic interactions of the retinal with the protein are likely to be the main reason for the different absorption behavior of these proteins. First, the electrostatic potential of the proteins at the retinal π -system was calculated. The potentials differ significantly between the archaeal rhodopsins. The potential of SRII shows a more pronounced difference between the β -ionone ring and the Schiff base than the potentials of BR and HR. Compared to the experimental absorption maxima of BR and HR at about 570 nm, the experimental absorption maximum of SRII is considerably blue-shifted to about 500 nm, *i.e.*, shifted to light of higher energy. Thus, the significantly higher potential difference between β -ionone ring and the Schiff base observed in SRII as compared to BR and HR is in accordance with their experimental absorption maxima. However, this conclusion is intuitive. Therefore, a mathematical model was introduced to qualitatively calculate absorption maxima based on the electrostatic potential. To derive absorption maxima from the calculated electrostatic potential, the model of a quantum mechanical particle in a box was generalized by including a step potential. This *model of a particle in a box with step potential* allowed to relate the observed differences in the electrostatic potential of the archaeal rhodopsins to their different absorption maxima.

To identify the origin of the difference between the archaeal rhodopsins, the potential was decomposed into the contribution of individual residues. The presented data showed that the counterion cannot explain the difference between the electrostatic potential of BR, HR and SRII. In agreement with mutational experiments, it could be shown that the retinal binding pocket contributes significantly to the difference between the electrostatic potentials at the retinal of BR, HR and SRII. Altogether, seven residues were identified

that account for the difference between the electrostatic potential of the proteins. Three of these residues are located in the retinal binding pocket. Another residue is located close to the β -ionone ring, but outside of the binding pocket. Three residues are, however, located more than 8 Å away from the retinal. The four residues that are far from the retinal have not been discussed before as contributing to the absorption shift between BR, HR and SRII. The electrostatic potential of the three archaeal rhodopsins at the retinal omitting these seven residues is virtually identical.

One approximation of continuum electrostatics is the representation of atoms as point charges, neglecting the electron distribution. The behavior of electrons, therefore, cannot be explicitly analyzed by classical methods. Instead, quantum mechanical methods are required for an explicit description of the electron distribution and in general for excited state processes. These methods are, however, limited to relatively few atoms. Therefore, those residues that are located far from the retinal, but, nevertheless, significantly interact with the chromophore electrostatically, would not have been identified using a quantum mechanical approach. Classical methods in general and continuum electrostatics in particular offer, thus, meaningful insights also for excited state processes. The analysis of an observed phenomenon can be greatly advanced by applying different methods and the interplay between these investigations may subsequently lead to a true understanding of this phenomenon.

CHAPTER 5

PROTONATION PROBABILITIES AND CORRELATIONS IN BACTERIORHODOPSIN INTERMEDIATE STRUCTURES

I didn't discover curves, I only uncovered them.

Mae West

Proton gradients across membranes are the most important potential energy source in living organisms. Proton gradients are generated by transmembrane proteins, so-called proton pumps, which translocate protons across the membrane. The energy for proton pumping is, for example, provided by light energy. The resulting gradient is a potential energy form that can subsequently be utilized, for example, for solute transport across the membrane. In bacteria, proton gradients may also be used to drive flagellar movement. Most importantly, ATP synthases can convert the potential energy of a proton gradient into chemical energy in the form of ATP.

A proton moves through a protein according to a mechanism first outlined by Theodor von Grotthuß in 1806 [206]. As showcased in Figure 5.1 proton transfer occurs along a hydrogen-bonded chain of proton donors and acceptors. Effectively, a single positive charge is transported along this chain through a rearrangement of covalent and hydrogen bonds. However, no proton undergoes substantial movement. Due to this phenomenon, proton transfer is significantly faster than the diffusion of an ion of the same size. In proteins, the proton may be transferred along a hydrogen-bonded chain comprising internal water molecules and protein side-chains. Proton transfer is accompanied by a successive change of protonation states of the involved groups. Understanding the protonation of proton pumps is, thus, a fundamental precondition for an understanding of their proton transfer mechanism.

Bacteriorhodopsin (BR), the smallest proton pump known to date, has become the paradigm in the investigation of proton transfer across membranes. BR utilizes light energy to translocate a proton across the plasma membrane. The resulting proton gradient is converted into chemical energy by an ATP synthase. Together, BR and the ATP syn-

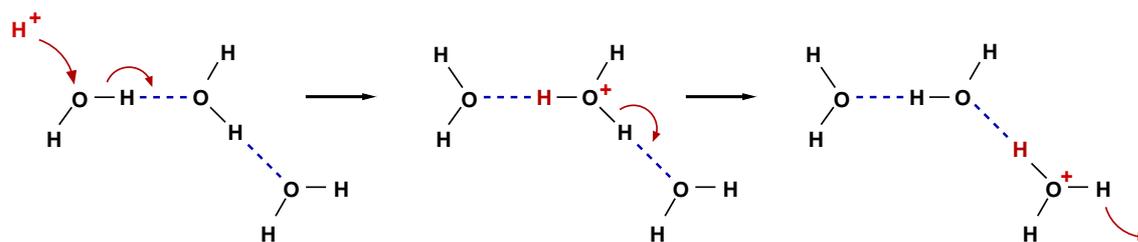


Figure 5.1. The Grotthuß mechanism. Three water molecules are depicted that are connected through hydrogen bonds as indicated by the blue dotted lines. Grotthuß envisioned proton transfer as a chain reaction: each oxygen atom consecutively receives a proton and passes another proton to the next oxygen atom. The proton transfer events are indicated by red arrows. The hydrogen atoms that change their position and the positive charge that is transported along the chain of water molecules are shown in red.

these constitute a photosynthetic pathway in halophilic archaea. In contrast to the photosynthesis in bacteria and plants, no carbon fixation occurs in the archaeal photosynthesis. Furthermore, a retinal Schiff base and not chlorophyll is the light absorbing chromophore. Retinal-based photosynthesis occurs only in archaea. It is, moreover, the only photosynthetic process known to occur in the archaeal kingdom.

This chapter presents the results of Metropolis Monte Carlo calculations to determine the protonation behavior of multiple BR structures. The computational details of the performed calculations are outlined in Section 5.2. The protonation states relevant for the proton transfer in BR are examined and the results are presented in Section 5.3. In Section 5.4, the individual protonation probabilities of functionally important residues are analyzed. Thereafter, the pair-correlation of the protonation behavior is investigated in Section 5.5. The chapter concludes with a general discussion of the results (Section 5.6). To begin with, the main events of the proton transfer of BR are summarized in the following section. A more detailed discussion of the functional mechanism can be found in Chapter 1.

5.1 PROTON TRANSFER OF BACTERIORHODOPSIN

The sequence of events triggered by light absorption that subsequently leads to proton transfer is termed *photocycle*. As shown in Figure 5.2 a, the photocycle consists of several intermediate states. The intermediate states differ from each other with respect to absorption maximum, structure and protonation state. In the last decade, crystallographers attempted to solve structures representing each intermediate state. These structures are categorized according to their proposed intermediate state as bR, K, L, M, N and O structures. The M structures are not further subdivided into M_1 and M_2 structures, since the structural data, in general, does not allow a specific allocation to either M_1 or M_2 .

The proton is transported through BR along a series of protein residues. These key residues of proton transfer are explicitly shown in the structure of BR depicted in Chapter 1, Figure 1.7. Furthermore in Figure 5.2 b, the general position of the key residues

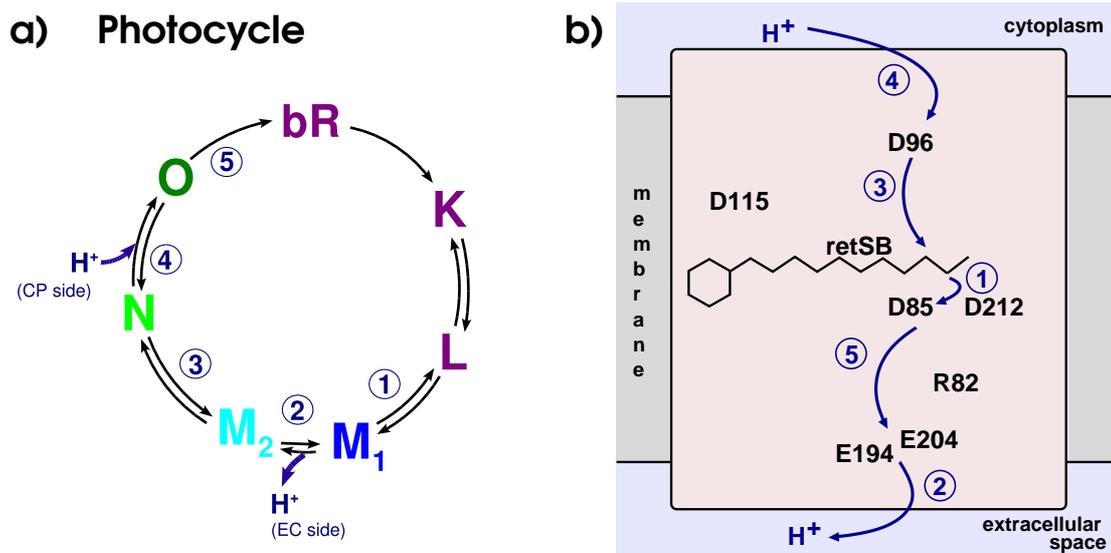


Figure 5.2. Proton Transfer of bacteriorhodopsin. **a)** When light is absorbed by the retinal Schiff base while BR is in the bR state, a sequence of events is triggered termed the photocycle. Consecutively, BR adopts a serial of intermediate states: bR, K, L, M₁, M₂, N and O. During the photocycle, a proton is transferred from the cytoplasm to the extracellular space. **b)** In the schematic of BR, the key residues and the proton transfer steps from the cytoplasmic to the extracellular side of the cell are indicated. The encircled numbers refer to the sequence of the five proton transfer steps and their occurrence during the photocycle. The protonation states of the photocycle intermediate states are given in Table 5.1.

and the five proton transfer steps are indicated in a schematic of BR. The protonation of these residues has been identified experimentally for the different photocycle intermediates [55, 57, 63, 207, 208]. In the transition from the bR to the K and from the K to the L intermediate state, no proton transfer takes place. The bR, K and L intermediate state differ from each other in small structural rearrangements, but not in the protonation of the key residues. The first change in protonation occurs during the transition from the L to the M₁ intermediate state, when the proton is transferred from the retinal Schiff base to Asp85. During the M₁ to M₂ transition a proton is released to the extracellular space from the proton release group Glu194/Glu204. In the transition from the M₂ to the N intermediate state, the retinal Schiff base receives a proton from Asp96. Asp96 is then protonated from the cytoplasm during the transition from the N to the O intermediate state. During the last step, the transition from the O state back to the bR state, the proton is transferred from Asp85 to the proton release group Glu194/Glu204. Although they do not change their protonation during the photocycle, Arg82, Asp115 and Asp212 are functionally important residues. Asp115 and Arg82 were experimentally shown to remain protonated, while Asp212 was shown to remain deprotonated during the photocycle [189, 208].

Other protonatable residues of BR are speculated to function as proton antennas. Their involvement in the proton transfer is, however, not well established. The focus of the present work lies, therefore, on the key residues of the BR proton transfer described

	[bR]/[K]/[L]	[M ₁]	[M ₂]	[N]	[O]
Asp96	1	1	1	0	1
Asp115 ^a	1	1	1	1	1
retinal Schiff base	1	0	0	1	1
Asp85	0	1	1	1	1
Asp212 ^b	0	0	0	0	0
Arg82 ^a	1	1	1	1	1
Glu194/Glu204	1	1	0	0	0

^a Asp115 and Arg82 remain protonated during the physiological photocycle.

^b Asp212 remains deprotonated during the physiological photocycle.

Table 5.1. Protonation state subsets of bacteriorhodopsin photocycle intermediates. The key residues of proton transfer are listed in sequential order starting from the cytoplasmic side of BR (*cf.* Figure 5.2 b). [bR], [K], [L], [M₁], [M₂], [N] and [O] refer to the protonation state of the respective photocycle intermediate (*cf.* Figure 5.2 a). In principle, the table defines protonation state subsets, since residues not listed in the table may be protonated or deprotonated. The bR, K and L intermediate adopt the same protonation state. In the following, this protonation state will be referred to as [bR] state.

above. In sequential order from the cytoplasmic to the extracellular side these are: Asp96, Asp115, the retinal Schiff base, Asp85, Asp212, Arg82, Glu194 and Glu204. The protonation behavior of these key residues is analyzed in detail in the results part of this chapter.

5.2 COMPUTATIONAL DETAILS

The Metropolis Monte Carlo calculations presented in this chapter are performed on all 20 structures of BR listed in Chapter 3, Table 3.1. The structures are grouped according to their intermediate state as bR, K, L, M and N structures. The M structures are not further subdivided into M₁ and M₂ structures, since an unambiguous classification of the structural data into M₁ and M₂ state is, in general, not feasible. No structure is available for the O intermediate of the BR photocycle. Instead, calculations are performed on the Asp85Ser mutant structure. This structure is proposed to resemble the O intermediate and is, therefore, termed O-like intermediate structure. To simplify the differentiation between the numerous structures, the individual structures will be referred to by their intermediate state and their PDB code as given in Table 3.1: *e.g.*, bR:1c3w or M:1f4z.

The proton release group from which the proton is released to the extracellular space is rather complex. It is the only proton donor/acceptor of BR that includes several groups. Specifically, the proton release consists of Glu194, Glu204 and a water cluster located in between the two glutamate residues. The proton is delocalized over the proton release group [63]. For the calculation performed in this work, water molecules are not modeled explicitly but are represented by a continuum. The proton release group, thus, comprises

the two glutamate residues Glu194 and Glu204. In the following this group will be referred to as Glu194/Glu204. Since a proton can reside on both Glu194 and Glu204, they are considered here as one group which can bind up to two protons.

The pK_{intr} (cf. Eq. (2.17)) and $W_{\mu\nu}$ (cf. Eq. (2.20)) are calculated as described in Chapter 2 and Chapter 3. From the pK_{intr} and $W_{\mu\nu}$, low energy protonation states are sampled using the Metropolis Monte Carlo approach as described in Chapter 2, Subsection 2.5.5. The Metropolis Monte Carlo algorithm is implemented in the program MCT (Monte Carlo Titration) by G. Matthias Ullmann. The Monte Carlo sampling is performed in a pH range from 0 to 14 at pH intervals of 0.1 pH units. The temperature is set to 300 K. The coupling threshold for double moves is 2 pH units and for triple moves 3 pH units. For a double or a triple move, the protonation form of two and three sites, respectively, is changed simultaneously during a Monte Carlo step. At each pH value, 30 000 Monte Carlo scans are performed. A Monte Carlo scan consists of one Monte Carlo step per site, pair and triplet, respectively. The result of a Metropolis Monte Carlo calculation is a large sample of protonation states biased towards the Boltzmann distribution.

Protonation probabilities can be derived from the Boltzmann distributed states sampled with MCT using the equations given in Chapter 2, Subsection 2.5.5. The equations are implemented in the program TRAJANALYSIS by G. Matthias Ullmann. The probability of protonation substates is derived from the solution of Eq. (2.34). The the probabilities of the functionally meaningful protonation state subsets of BR are presented in Section 5.3. The protonation probability of an individual protonatable site in a protein environment is obtained by solving Eq. (2.32). The protonation probabilities of the key residues are discussed in Section 5.4. The correlation between the protonation of each pair of protonatable sites is calculated by solving Eq. (2.35). The correlation of the protonation behavior of the key residues are presented in Section 5.5.

5.3 PROTONATION STATE SUBSETS

In this section, the probability of functionally important protonation states is investigated for the BR structures. For the examined states, the protonation of the functionally important residues Asp96, Asp115, the retinal Schiff base, Asp85, Asp212, Arg82 and Glu194/Glu204 is considered (Figure 5.2 b). The protonation states listed in Table 5.1 define the protonation of the key residues of the proton transfer in BR. These protonation states actually characterize a subset of protonation states, since all other protonatable residues of BR, *e.g.*, surface residues, may be protonated or deprotonated. The protonation states listed in Table 5.1 correspond to the protonation states of the photocycle intermediates (cf. Figure 5.2 a). To differentiate between an intermediate state, which is characterized by structure, absorption maximum and protonation state, or an intermediate structure, *i.e.*, atomic coordinates, and their corresponding protonation state, the latter are written in squared brackets, *e.g.*, L intermediate state, L structure and [L] protonation state.

As can be seen in Table 5.1, Arg82, Asp115 and Asp212 do not change their protonation during the photocycle. They are, however, functionally important and all physiological protonation states mentioned in this thesis explicitly include these three residues. In

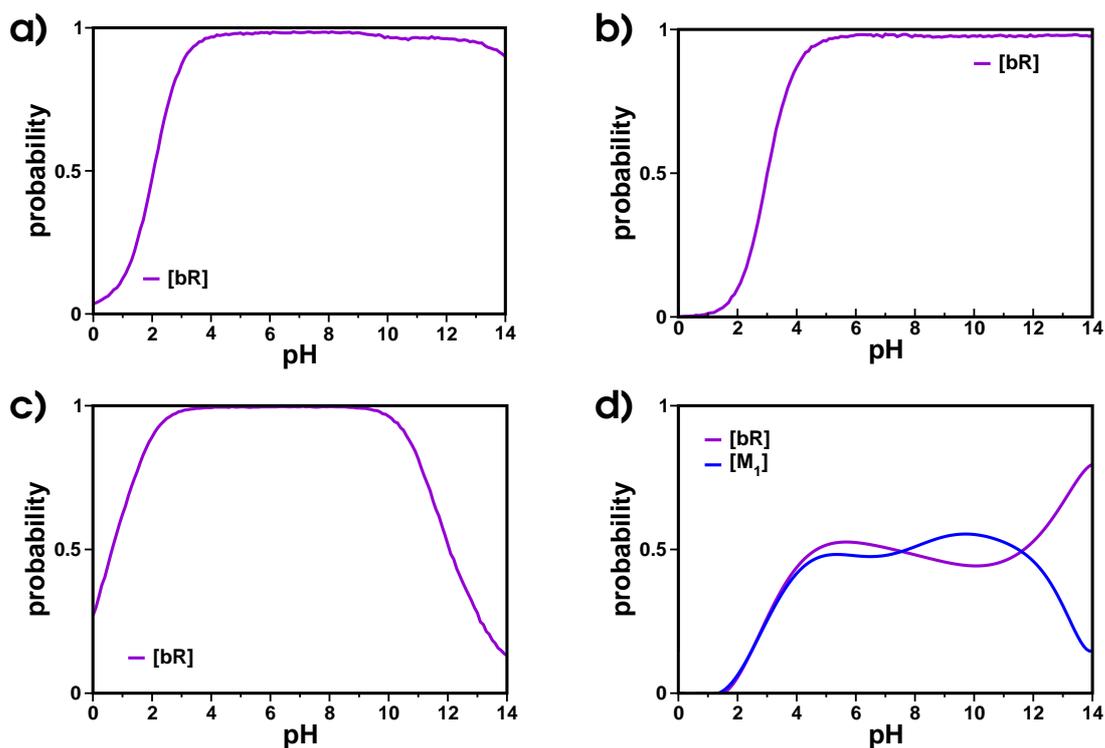


Figure 5.3. Protonation state subsets of the bR structures. The probability of the protonation states is plotted against pH for: **a)** bR:1c3w, **b)** bR:1qhj, **c)** bR:1f50 and **d)** bR:1iw6. The [bR] and the [M₁] protonation state correspond to those listed in Table 5.1. The colors of the curves correspond to the photocycle intermediate states depicted in Figure 5.2. The structures are named by their intermediate state and PDB code (*cf.* Table 3.1).

addition to the protonation states occurring during the photocycle as listed in Table 5.1, multiple non-physiological states are considered, *e.g.*, states where Arg82 is deprotonated or Asp212 is protonated. However, the complete combinatorial space of $2^8 = 256$ protonation states that the considered key residues can, in principle, adopt is not examined. The total probability of the state subsets may, therefore, be less than 1.

5.3.1 BR STRUCTURES

In total, about 20 high-resolution structures of the bR state, *i.e.*, ground state, are resolved. The four structures analyzed here, bR:1c3w, bR:1qhj, bR:1f50 and bR:1iw6, are chosen to give an adequate overview of the behavior of the bR structures. Figure 5.3 shows the protonation state probability for the four bR state structures. As can be seen in Figure 5.3 a, b and c, the only significant protonation state for bR:1c3w, bR:1qhj and bR:1f50, respectively, is the [bR] state. In the physiological pH range, the probability of the [bR] state is close to 1 for all three structures. This is true for the majority of bR structures (data not shown). Thus, the high-resolution structures of the bR state capture one of the main features associated with this state, namely the protonation of the functionally important residues, correctly.

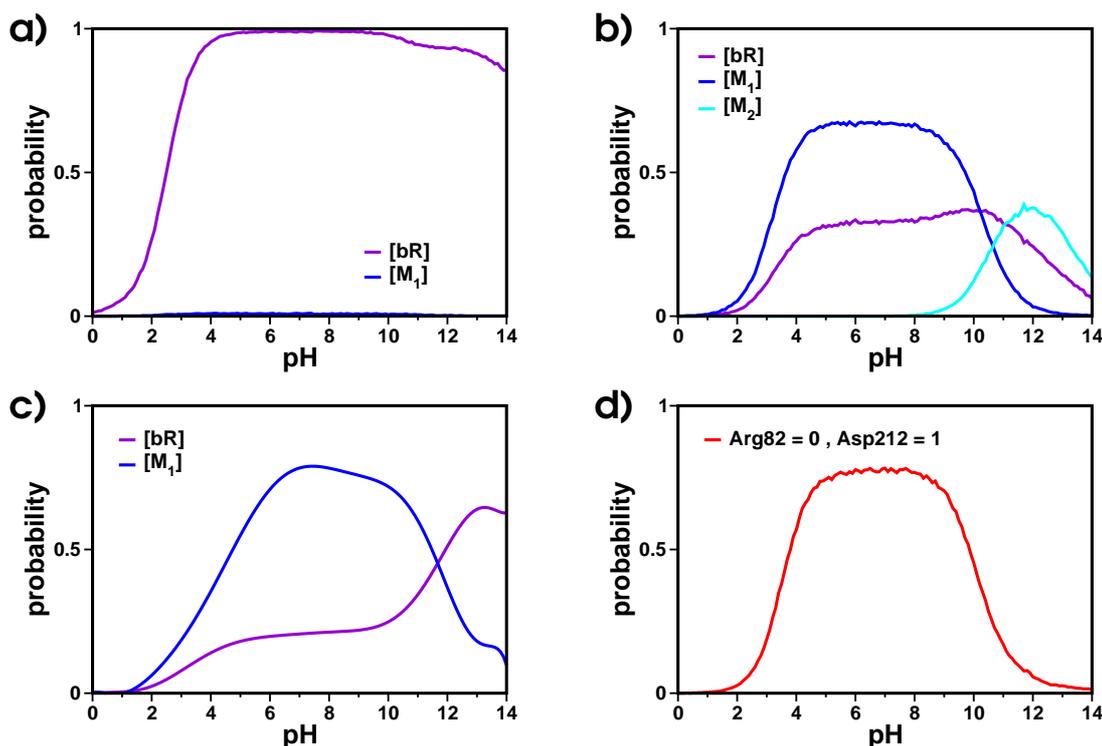


Figure 5.4. Protonation state subsets of the K structures. The probability of the protonation states is plotted against pH for: **a)** K:1m0k, **b)** K:1qko, **c)** K:1ixf and **d)** K:1qkp. The [bR], [M₁] and [M₂] protonation state correspond to those listed in Table 5.1. The red curve describes a non-physiological protonation state with deprotonated Arg82 and protonated Asp212 as indicated in the graph. The colors of all other curves correspond to the photocycle intermediate states depicted in Figure 5.2. The structures are named by their intermediate state and PDB code (*cf.* Table 3.1).

Figure 5.3 d shows the probability of states for bR:1iw6. The X-ray data of this structure was sampled using a protocol that aimed at preventing ‘*specific damage induced by X-ray radiation*’ to the crystal and the protein structure [151]. More precisely, the X-ray data was collected relatively slowly under relatively low radiation to avoid damage due to heat and radiation. As can be seen, the [bR] and the [M₁] state have a similar probability in this structure indicating that the crystal contained not a pure bR ensemble during the data collection.

5.3.2 K STRUCTURES

The protonation state of the K intermediate is identical to the [bR] state (Table 5.1). Four K structures are analyzed here: K:1m0k, K:1qko, K:1ixf and K:1qkp. Figure 5.4 a shows the protonation state probability for K:1m0k. Similar to bR:1c3w, bR:1qhj and bR:1f50, in K:1m0k the [bR] state is the only significant state. Furthermore, at physiological pH the probability of the [bR] state is close to 1. Additionally, a small probability for the [M₁] state can be observed for K:1m0k.

As shown in Figure 5.4 b, the behavior of K:1qko differs from that of K:1m0k and the bR structures. The state with the highest probability at physiological pH is the [M₁] state. However, the [bR] state has also a significant probability. Additionally at higher pH values, K:1qko has a probability for the [M₂] state indicating that the proton release group Glu194/Glu204 interacts with the aqueous phase. Thus, Glu194/Glu204 deprotonates when the pH value of the solvent is high. For K:1qko, this implies that the [M₁] state and at higher pH values the [M₂] state are energetically favorable compared to the [bR] state, though the [M₁] or the [M₂] state may not be accessible to the structure. Since the presented data are obtained from equilibrium calculations no conclusion with respect to the energy barrier between the [M₁] and the [bR] state can be drawn. However, it can be concluded that K:1qko has already adopted certain features of the M intermediate state.

For K:1ixf, the state with the highest probability is the [M₁] state, while the probability of the [bR] state is approximately 0.2 in the physiological pH range as shown in Figure 5.4 c. The behavior observed for K:1qko (*cf.* Figure 5.4 b) is even more pronounced for K:1ixf. The data indicates that K:1ixf has adopted certain features of the M₁ intermediate state.

For K:1qkp, the main state is characterized by a deprotonated Arg82 and a protonated Asp212 (*cf.* Figure 5.4 d). Since Arg82 is protonated and Asp212 is deprotonated throughout the photocycle, this state does not constitute a functionally meaningful protonation state. Except for the protonation of Arg82 and Asp212 this state is in accordance with the [M₁] state, *i.e.*, the retinal Schiff base is deprotonated and Asp85 is protonated. Since the protonated Asp85 and Asp212 are neutral as well as the deprotonated Arg82 and the retinal Schiff base, the active center of this structure is completely uncharged. For K:1qkp, Glu194/Glu204, which binds one proton, carries the only internal charge.

5.3.3 L STRUCTURES

As for the K intermediate, the protonation state of the L intermediate is identical to the [bR] state (Table 5.1). The first proton transfer event takes place during the transition from the L to the M₁ intermediate. Therefore, a tendency towards the [M₁] protonation state can be expected for the L structures.

For the L structures L:1e0a and L:1vjm.a the main state is the [bR] state as shown in Figure 5.5 a and b, respectively. Only a very low probability for the [M₁] state can be observed. The protonation behavior of L:1e0a and L:1vjm.a is, thus, very similar to the bR structures (*cf.* Figure 5.3) and the K structure K:1m0k (*cf.* Figure 5.4 a).

In contrast, the L structure L:1vjm.b has no probability for the [bR] state. Instead, L:1vjm.b has a significant probability for the [M₂] as well as for the [O] state as shown in Figure 5.5 c. While the [O] state has a significant probability in the pH range from 2 to 6, the [M₂] state has a significant probability at physiological pH values.

As shown in Figure 5.5 d, L:1e0p shows a probability for a functionally meaningful protonation state, namely the [M₁] state, only at high pH values. The two other significant states are characterized by a protonated Asp212. Furthermore, the retinal Schiff base is deprotonated and Asp85 protonated, while Glu194/Glu204 are protonated (red curve) or deprotonated (orange curve). Except for Asp212, the protonation behavior is in accordance with the [M₁] and [M₂] protonation state, respectively. Since Asp212 is depro-

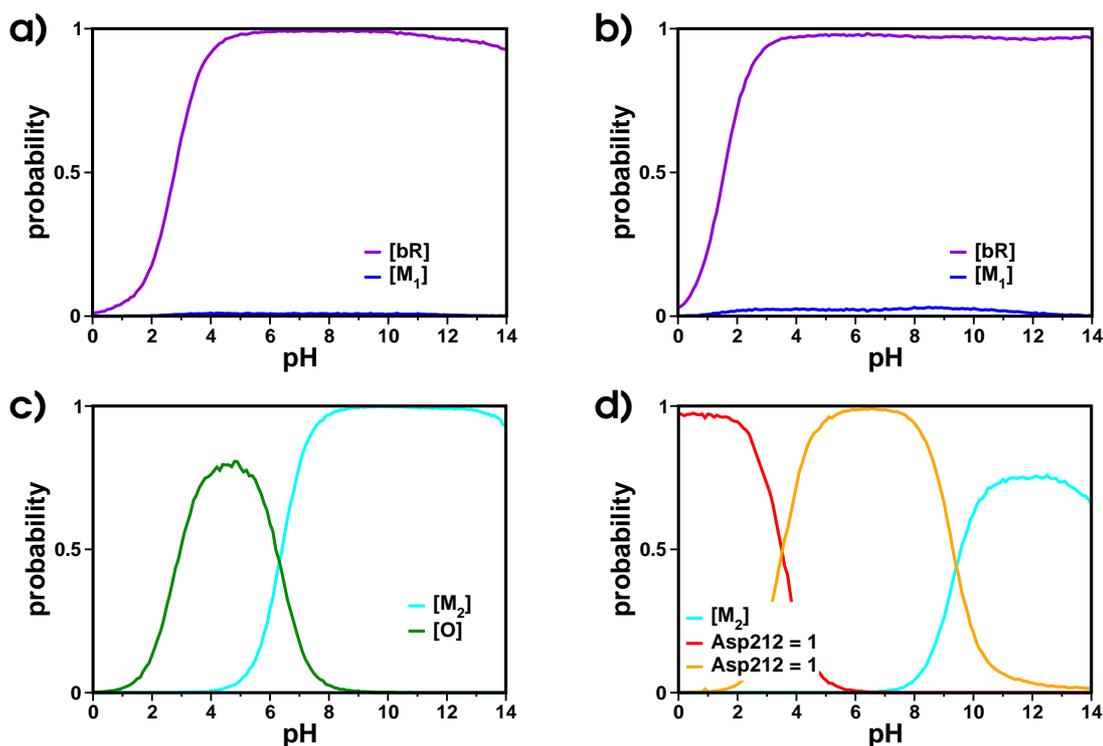


Figure 5.5. Protonation state subsets of the L structures. The probability of the protonation states is plotted against pH for: **a)** L:1o0a, **b)** L:1vjm.a, **c)** L:1vjm.b and **d)** L:1e0p. The [bR], [M₁], [M₂] and [O] protonation state correspond to those listed in Table 5.1, if not denoted otherwise. The orange and the red curve describe non-physiological protonation states with a protonated Asp212 is protonated as indicated in the graph. These two states differ in the protonation of Glu194/Glu204 which is either protonated (red curve) or deprotonated (orange curve). The colors of all other curves correspond to the photocycle intermediate states depicted in Figure 5.2. The structures are named by their intermediate state and PDB code (*cf.* Table 3.1).

tonated throughout the photocycle, these states are not populated under physiological conditions. Both Asp85 and Asp212 are protonated and the retinal Schiff base is deprotonated, therefore, the active center of L:1e0p is neutral. A similar behavior is described in the previous subsection for the K structure K:1qkp which additionally has a neutral Arg82.

5.3.4 M STRUCTURES

The M intermediate is characterized by a deprotonated retinal Schiff base and a protonated Asp85. The neutral Schiff base is the cause for the significantly blue shifted absorption maximum of the M intermediate state. The M intermediate can be subdivided into two states, the M₁ and the M₂ state, that differ in the protonation of the release group Glu194/Glu204. In the [M₁] protonation state, Glu194/Glu204 is protonated and in the [M₂] protonation state, its proton has been released to the extracellular space and Glu194/Glu204 is, therefore, deprotonated.

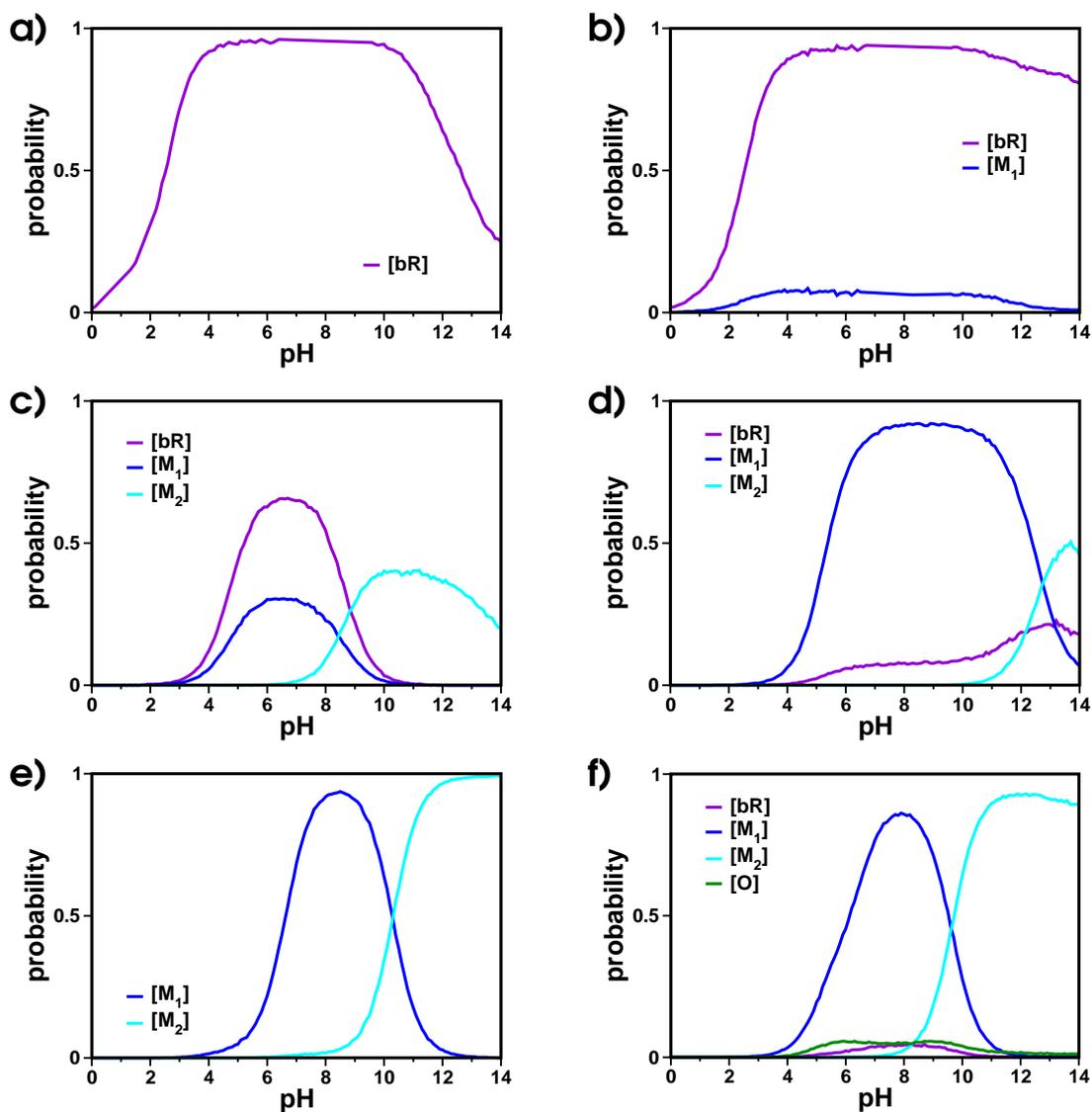


Figure 5.6. Protonation state subsets of the M structures. The probability of the protonation states is plotted against pH for: **a)** M:1kg9, **b)** M:1p8h, **c)** M:1m0m, **d)** M:1f4z, **e)** M:1cwq and **f)** M:1kg8. The [bR], [M₁], [M₂] and [O] protonation state correspond to those listed in Table 5.1. The colors of the curves correspond to the photocycle intermediate states depicted in Figure 5.2. The structures are named by their intermediate state and PDB code (*cf.* Table 3.1).

Figure 5.6 shows that the [bR] state is the main state for the M structures M:1kg9, M:1p8h and M:1m0m, the protonation state preceding the [M₁] and [M₂] state. As depicted in Figure 5.6 a, the [bR] state is the only state for which M:1kg9 has a probability. M:1p8h has a probability of more than 0.9 for the [bR] state and a probability of less than 0.1 for the [M₁] state (Figure 5.6 b). As shown in Figure 5.6 c, the behavior of M:1m0m is more complex. While the [bR] state has the highest probability at physiological pH, the [M₁] state has a significant probability, too. Furthermore, M:1m0m has a significant probability for the [M₂] state at higher pH values. Although M:1p8h and M:1m0m have a

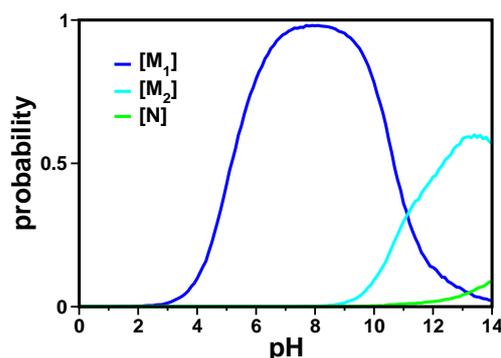


Figure 5.7. Protonation state subsets of the N structure. The probability of the protonation states is plotted against pH for N:1p8u. The [M₁], [M₂] and [N] protonation state correspond to those listed in Table 5.1. The colors of the curves correspond to the photocycle intermediate states depicted in Figure 5.2. The structure is named by its intermediate state and PDB code: N:1p8u (*cf.* Table 3.1).

probability for the [M₁] state at physiological pH, their protonation behavior more closely resembles that of the bR intermediate.

For the M structures M:1f4z, M:1cwq and M:1kg8, the main state at physiological pH is the [M₁] state as depicted in Figure 5.6 d, e and f, respectively. In addition, all three M structures have a probability for the [M₂] at higher pH values. Furthermore, a minor probability of the [bR] state can be observed for M:1f4z. M:1kg8 shows a small probability for the [bR] as well as the for [O] state.

The three M structures M:1kg9, M:1p8h and M:1m0m do not represent the M intermediate of the photocycle, but resemble instead the bR intermediate as shown in Figure 5.6 a, b and c, respectively. The three M structures M:1f4z, M:1cwq and M:1kg8 represent the M₁ intermediate state as shown in Figure 5.6 d, 5.6 e and 5.6 f, respectively. None of the M structures, however, shows a behavior in accordance with the M₂ intermediate at physiological pH. A probability for the [M₂] protonation state can only be seen at high pH values.

5.3.5 N STRUCTURE

In the [N] state, Asp96 is deprotonated, while the retinal Schiff base is again protonated. Only one structure is available that is proposed to represent the N photocycle intermediate: N:1p8u. As depicted in Figure 5.7, N:1p8u has a significant probability for the [M₁] protonation state at physiological pH. At higher pH values, a significant probability for the [M₂] state, but only a minor probability for the [N] state can be observed. With respect to the probability of protonation states N:1p8u may, therefore, essentially represent an M structure. However, N:1p8u is the only structures that shows probability for the [N] state.

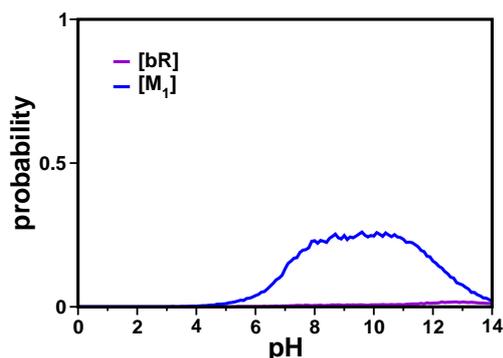


Figure 5.8. Protonation state subsets of the O-like intermediate structure. The probability of the protonation states is plotted against pH for O:1jv7. The [bR] and [M₁] protonation state correspond to those listed in Table 5.1. The colors of the protonation state probability curves correspond to the photocycle intermediate states depicted in Figure 5.2. The structure is named by its intermediate state and PDB code: O:1jv7 (*cf.* Table 3.1).

5.3.6 O-LIKE STRUCTURE

No structure of the O intermediate could be resolved so far. However, the Asp85Ser mutant of BR exhibits an absorption maximum similar to that of the O intermediate. At low pH values, the Asp85Ser mutant is proposed to have a protonated Asp85, Asp96 and retinal Schiff base. Furthermore, it could be shown that the retinal assumes an all-*trans* conformation. These features are characteristic for the O intermediate. A high-resolution structure of the Asp85Ser mutant is available: O:1jv7. This structure is termed O-like intermediate structure.

As shown in Figure 5.8, O:1jv7 has a probability of up to 0.25 for the [M₁] protonation state in the pH range from 6 to 14. A very low probability for the [bR] state can also be seen. However, O:1jv7 shows no probability for the [O] state.

5.3.7 SYNOPSIS OF SECTION 5.3

In the previous subsections, the probability of functionally important protonation states are presented for multiple photocycle intermediate structures of BR. The functionally important protonation states, namely [bR], [M₁], [M₂], [N] and [O], are listed in Table 5.1. This subsection gives a short summary of the presented results.

[BR] STATE

It could be shown that the [bR] protonation state is well represented by a number of BR structures. Three bR structures are presented that show a probability for the [bR] state of close to 1 over a significant pH range (*cf.* Figure 5.3 a, 5.3 b and 5.3 c). Also one K structure, K:1m0k, and two L structures, L:1o0a and L:1vjm.a, show a probability for the [bR] state, *i.e.*, their physiological protonation state, of nearly 1 over a large pH range (*cf.* Figure 5.4 and 5.5). Furthermore, two of the presented M structures, M:1kg9 and

M:1p8h, have a probability to be in the [bR] state close to 1 over a significant pH range (*cf.* Figure 5.6 a and 5.6 b). In particular M:1kg9 which shows no probability for any protonation state apart from the [bR] state can be classified as a ground state structure. Another M state structure, M:1m0m, has a principal probability of about 0.6 for the [bR] state at physiological pH values (*cf.* Figure 5.6 b).

[M₁] STATE

One bR structure, bR:1iw6, has an similar probability of about 0.5 for both the [bR] and the [M₁] state over a large pH range (*cf.* Figure 5.3 d). At physiological pH values, two K structures, K:1qko and K:1ixf, show a significant probability for the [M₁] protonation state (*cf.* Figure 5.4). Three M structures, M:1f4z, M:1cwq and M:1kg8, have a high probability for the [M₁] state as shown in Figure 5.6. Additionally, M:1m0m shows a minor probability of about 0.25 for the [M₁] state. Furthermore, the N structure has a probability for the [M₁] state close to 1 at physiological pH values. At high pH values, the O-like structure shows a small probability for the [M₁]

[M₂] STATE

The L structure L:1vjm.b has a probability of about 1 for the [M₂] protonation state over a large pH range (Figure 5.5 c). Furthermore at high pH values, several structures show a significant probability for the [M₂] state: the K structure K:1qko (*cf.* Figure 5.4 b), the L structure L:1e0p (*cf.* Figure 5.5 d), four M structures, M:1m0m, M:1f4z, M:1cwq and M:1kg8 (*cf.* Figure 5.6), and the N structure (*cf.* Figure 5.7).

[N] STATE

The N structure N:1p8u shows a minor probability for the [N] state as shown in Figure 5.7.

[O] STATE

In the pH range from 2 to 6, the L structure L:1vjm.b shows a significant probability for the [O] protonation state as shown in Figure 5.5 c. Furthermore, the M structure M:1kg8 has a small probability of approximately 0.05 for the [O] state in the pH range from 4 to 12 (*cf.* Figure 5.6 f).

5.4 PROTONATION PROBABILITIES OF THE KEY RESIDUES OF PROTON TRANSFER

The key residues of the BR proton transfer and the main events of the photocycle have been detailed in Chapter 1 and a short summary is given at the beginning of this chapter. The functionally meaningful protonation states of the proton transfer are listed in Table 5.1. During proton transfer, the protonation of several residues changes: Asp96, the retinal Schiff base, Asp85 and Glu195/Glu204. The protonation of other residues remains constant throughout the physiological photocycle: Asp115, Asp212 and Arg82.

In the previous section, the probability of subsets of protonation states is discussed. This section analyzes the individual protonation probabilities of the key residues. The protonation probability of the residues is discussed in sequential order starting from the cytoplasmic side of BR towards the extracellular side: Asp96, Asp115, retinal Schiff base, Asp85, Asp212, Arg82 and Glu194/Glu204 (*cf.* Figure 5.2). The individual protonation behavior of the residues is compared for the different intermediate structures. The individual structures are named by their PDB code and grouped with respect to the photocycle intermediate state they represent. Squared brackets are used to emphasize the protonation states corresponding to the photocycle intermediate states as listed in Table 5.1: [bR], [M₁], [M₂], [N] and [O] (protonation) state.

5.4.1 PROTONATION PROBABILITY OF ASP96

Asp96 is the proton donor for the retinal Schiff base. It is protonated in the [bR], [M₁] and [M₂] state and deprotonated in the [N] state, when the retinal Schiff base is protonated. In the [O] state, Asp96 has received a proton from the cytoplasm and is again protonated. Hence, Asp96 is protonated in all intermediate states but the N state.

The protonation probability of Asp96 is depicted in Figure 5.9 for all BR structures: bR, K, L, M, N and O-like. As can be seen, Asp96 has a protonation probability close to 1 in the pH range between 0 and 10 in all BR structures. Although Asp96 is deprotonated in the [N] state, it is mostly protonated in the available N structure (Figure 5.9 e). Only at high pH values, Asp96 deprotonates slightly. The observed protonation behavior is in accordance with the fact that the N structure shows only a small probability for the [N] protonation state at high pH values (*cf.* Figure 5.7).

5.4.2 PROTONATION PROBABILITY OF ASP115

Asp115 is located close to the β -ionone ring of the retinal Schiff base relatively far from the proton transfer channel. This aspartate was shown to remain protonated throughout the physiological photocycle. Deprotonation of Asp115 is, however, implicated in the so-called back pressure effect: if the proton gradient or the membrane potential is too large, the BR proton pumping ceases. This effect may be caused by the deprotonation of Asp115 that can be observed, if the pH gradient or the membrane potential are large [64, 65]. The deprotonated Asp115 interacts with Asp85 and may induce its protonation. Then, no acceptor is available for the retinal Schiff base proton.

The calculations presented here are performed without considering a transmembrane pH gradient or a membrane potential. Under these conditions, Asp115 is shown to be protonated in all intermediate structures in the physiological pH range in all BR structures (Figure 5.10). However, at pH > 10, Asp115 partly deprotonates in all structures. In most structures its protonation probability decreases to less than 0.3 at pH = 14.

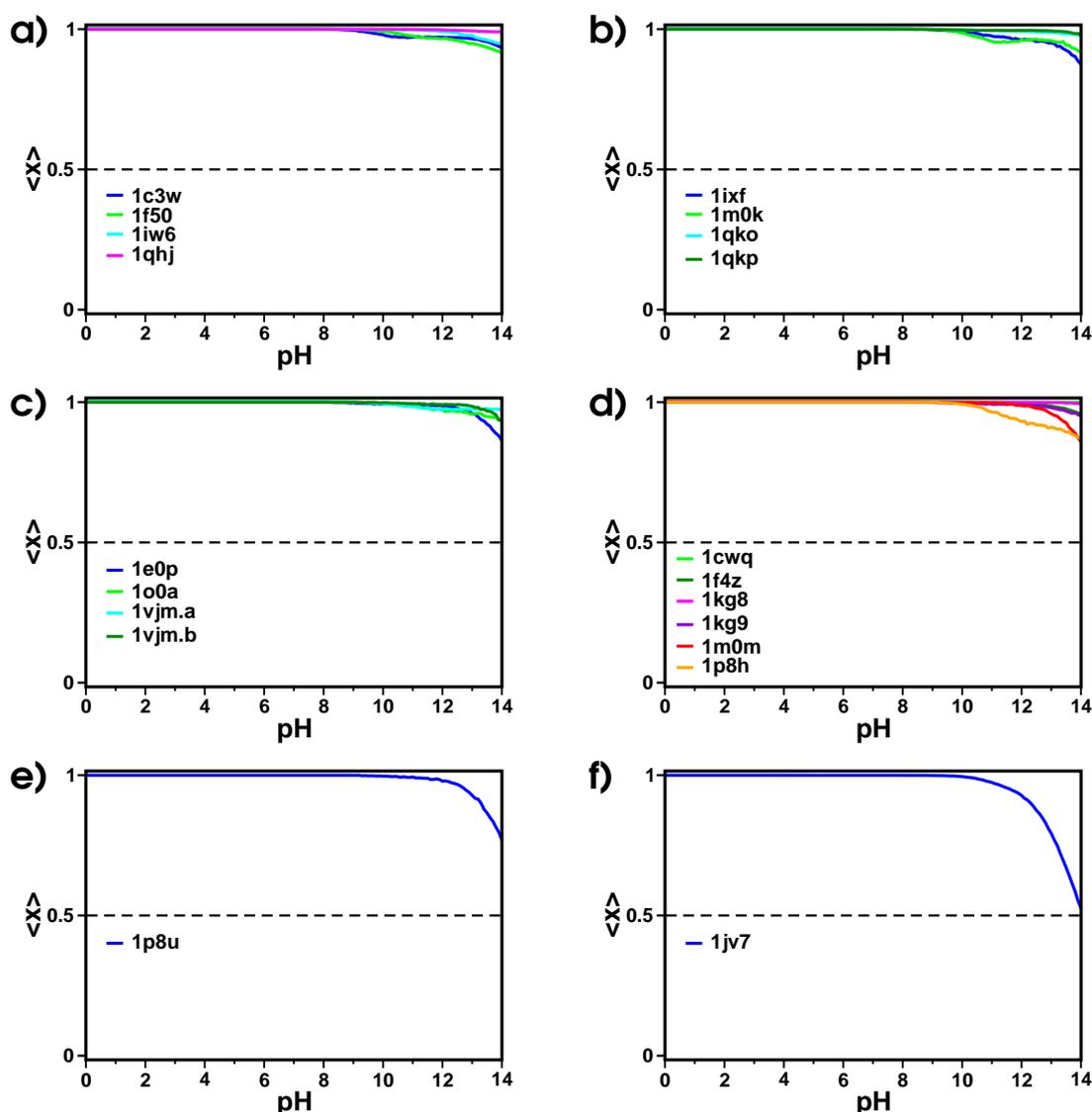


Figure 5.9. Protonation probability of Asp96. The protonation probability of Asp96 is plotted against pH for the different intermediate structures: **a)** bR, **b)** K, **c)** L, and **d)** M structures, **e)** N structure, and **f)** O-like intermediate structure. The curves are assigned to the individual structures by their respective PDB code (*cf.* Table 3.1).

5.4.3 PROTONATION PROBABILITY OF THE RETINAL SCHIFF BASE

The retinal Schiff base is protonated in the [bR] state. During the L to M₁ transition the proton is translocated to Asp85 and the retinal Schiff base is deprotonated in the [M₁] and [M₂] state. In the N intermediate, the retinal Schiff becomes protonated again and is, thus, protonated in the [N] and [O] state.

In accordance with the experimental data, the bR structures bR:1c3w, bR:1f50 and bR:1qhj have a protonated Schiff base (Figure 5.11 a). In the bR structure bR:1iw6, the Schiff base is partially deprotonated which suggests a tendency towards the M interme-

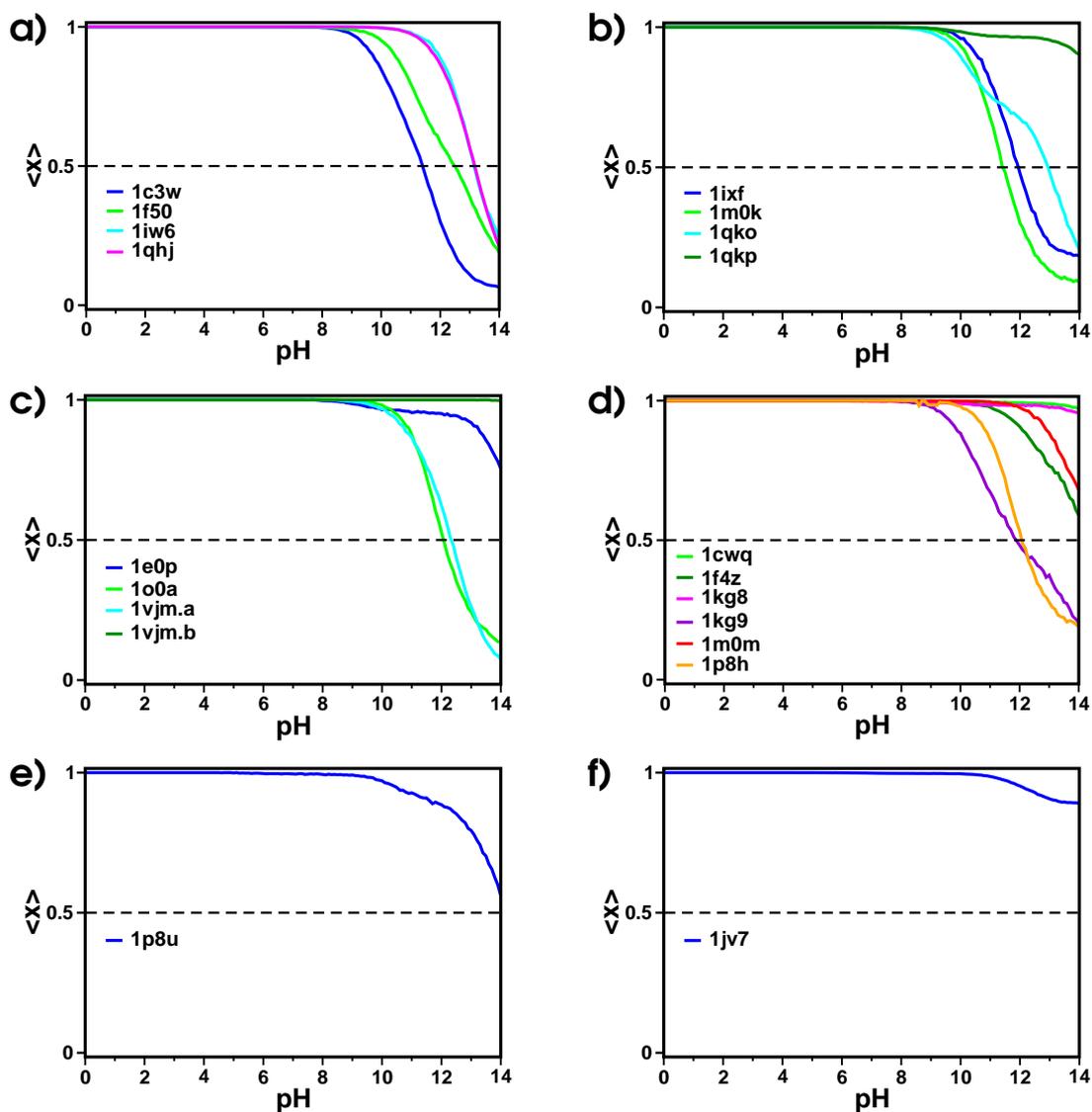


Figure 5.10. Protonation probability of Asp115. The protonation probability of Asp115 is plotted against pH for the different intermediate structures: **a)** bR, **b)** K, **c)** L, and **d)** M structures, **e)** N structure, and **f)** O-like intermediate structure. The curves are assigned to the individual structures by their respective intermediate state and PDB code (*cf.* Table 3.1).

diate. Indeed, bR:1iw6 has a similar probability for the [bR] and the $[M_1]$ protonation state to a similar degree at physiological pH (*cf.* Figure 5.3 d).

One of the K structures, K:1m0k, has a protonated retinal Schiff base, while the Schiff base is mostly deprotonated in the other K structures K:1ixf, K:1qko and K:1qkp (Figure 5.11 b). As shown in Figure 5.4, K:1m0k has a probability for the [bR] state with a probability close to 1. The K structures K:1qko and K:1ixf have a significant probability for both the [bR] and the $[M_1]$ state and K:1qkp has a probability for a non-physiological protonation state, where the Schiff base is deprotonated as well.

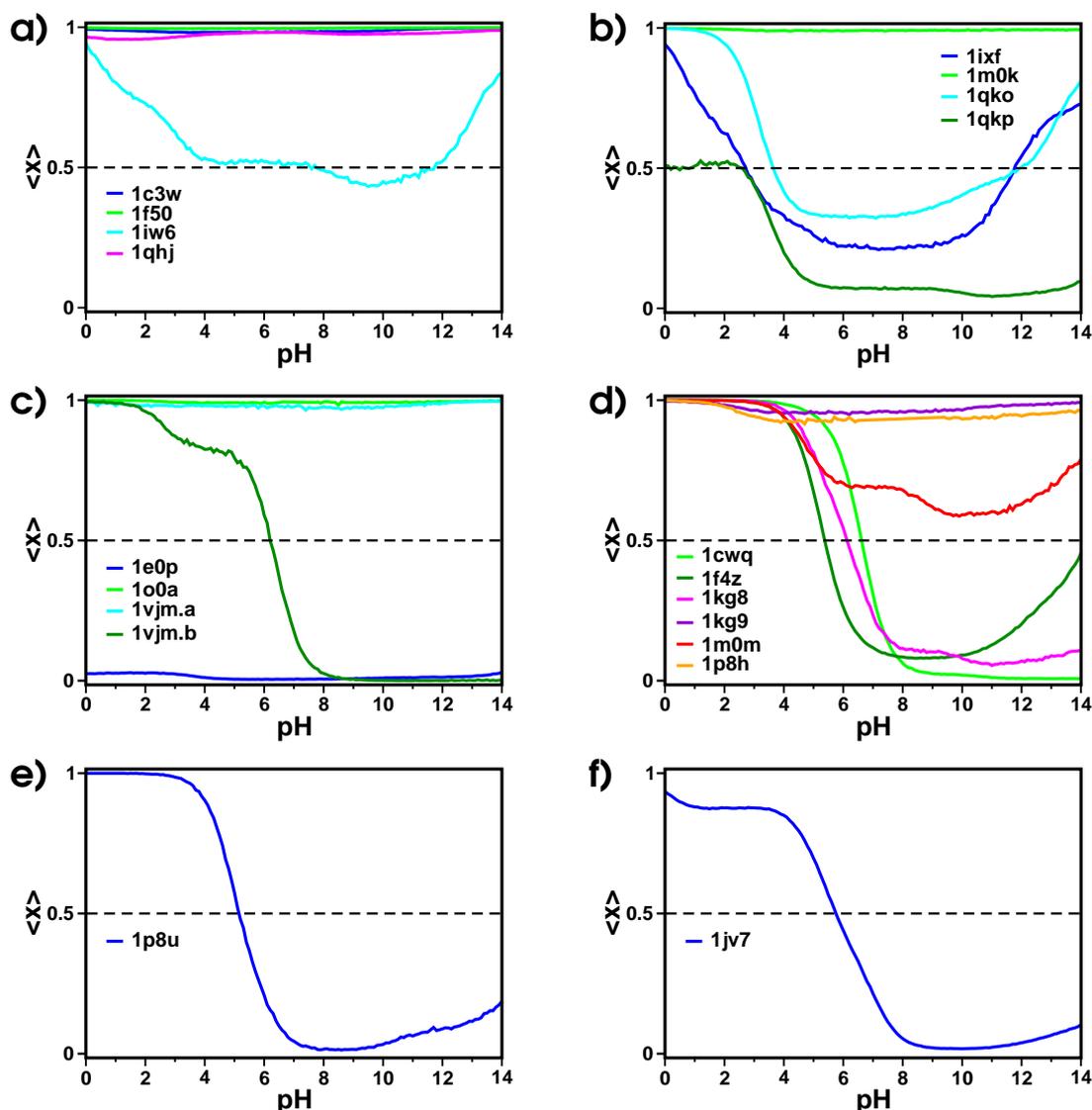


Figure 5.11. Protonation probability of the retinal Schiff base. The protonation probability of the retinal Schiff base is plotted against pH for the different intermediate structures: **a)** bR, **b)** K, **c)** L, and **d)** M structures, **e)** N structure, and **f)** O-like intermediate structure. The curves are assigned to the individual structures by their respective intermediate state and PDB code (*cf.* Table 3.1).

Two L structures, L:1o0a and L:1vjm.a, have a protonated Schiff base (Figure 5.11 c). These structures have a probability of nearly 1 for the [bR] state (*cf.* Figure 5.5 a and 5.5 b). In the L structure L:1vjm.b, the Schiff base is mainly protonated at low and deprotonated at higher pH values. This protonation behavior is in accordance with the probability of states shown in Figure 5.5 c. In the pH range from 2 to 6, L:1vjm.b a probability for the [O] state, where the Schiff base is protonated, can be observed and for the [M²] state which has a deprotonated Schiff base at higher pH values. The Schiff base is deprotonated in L:1e0p. As shown in Figure 5.5 d, L:1e0p has no probability for a functionally meaningful protonation state at physiological pH values. At high pH

values, L:1e0p has a significant probability for the [M₂] state. All these states have a deprotonated retinal Schiff base.

As shown in Figure 5.11 d for the M structures, the protonation probability of the Schiff base varies at physiological pH values over a large range from completely deprotonated to nearly protonated. In accordance with this behavior also the probability of states varies at physiological pH between [bR] and [M₁] as shown in Figure 5.6.

The protonation behavior of the retinal Schiff base is very similar in the N and the O-like intermediate structure as shown in Figure 5.11 e and 5.11 f. The Schiff base deprotonates at pH=4 and is deprotonated at physiological pH values. Both N:1p8u and O:1jv7 have a probability for the [M₁] state which has an deprotonated retinal Schiff base (Figure 5.7 and 5.8, respectively). Furthermore, the N structure has a probability for the [M₂] state at higher pH values. In the [N] state, the Schiff base is protonated. Although the Schiff base is protonated at low pH values, N:1p8u shows no probability for the [N] state.

5.4.4 PROTONATION PROBABILITY OF ASP85

Asp85 is the proton acceptor for the retinal Schiff base proton. Its protonation behavior is, thus, directly opposite to that of the retinal Schiff base during the early photocycle. Asp85 is deprotonated in the [bR] state. After receiving the proton from the retinal Schiff base during the L to M₁ transition, it is protonated in the [M₁], [M₂], [N] and [O] state. Asp85 remains protonated until it deprotonates during the final O to bR transition. While a proton is bound to either the retinal Schiff base or Asp85 in the [bR], [M₁] and [M₂] state, in the [N] and [O] state both groups are protonated.

In agreement with experimental data, Asp85 is deprotonated in the bR structures bR:1c3w, bR:1f50 and bR:1qhj as shown in Figure 5.12 a. The bR structure bR:1iw6 which has a significant probability for both the [bR] and the [M₁] state (Figure 5.3 d) shows a protonation probability of about 0.5 for Asp85. Asp85 is deprotonated in the K structure K:1m0k as shown in Figure 5.12 b. In the other three K structures K:1ixf, K:1qkp and K:1qko its protonation probability is above 0.7. Two L structures, L:1o0a and L:1vjm.a have a deprotonated Asp85 while it is protonated in the other two, L:1e0p and L:1vjm.b (Figure 5.12 c). For the M structures, the protonation behavior of Asp85 varies over a large range from nearly deprotonated to completely protonated (Figure 5.12 d). In accordance with this behavior, the probability of states is shown to vary between [bR] and [M₁] (Figure 5.6). Asp85 is protonated in the N and the O-like structure (Figure 5.12 e and 5.12 f).

A comparison of the protonation behavior of the retinal Schiff base and Asp85, demonstrates that the protonation probability of these two groups in general adds up to 1 at physiological pH values. Thus, in the Schiff base region of most structures, one proton is located that resides either on the Schiff base or on Asp85 or is shared between these two groups. A marked exception is the L structure L:1vjm.b. This structure has a protonated Asp85, while the retinal Schiff base deprotonates in the physiological pH range. Thus, for L:1vjm.b, a significant probability for the [O] state, where both residues are protonated, can be observed (*cf.* Figure 5.5 c). This L structure is the only BR structure with a significant probability for the [O] state.

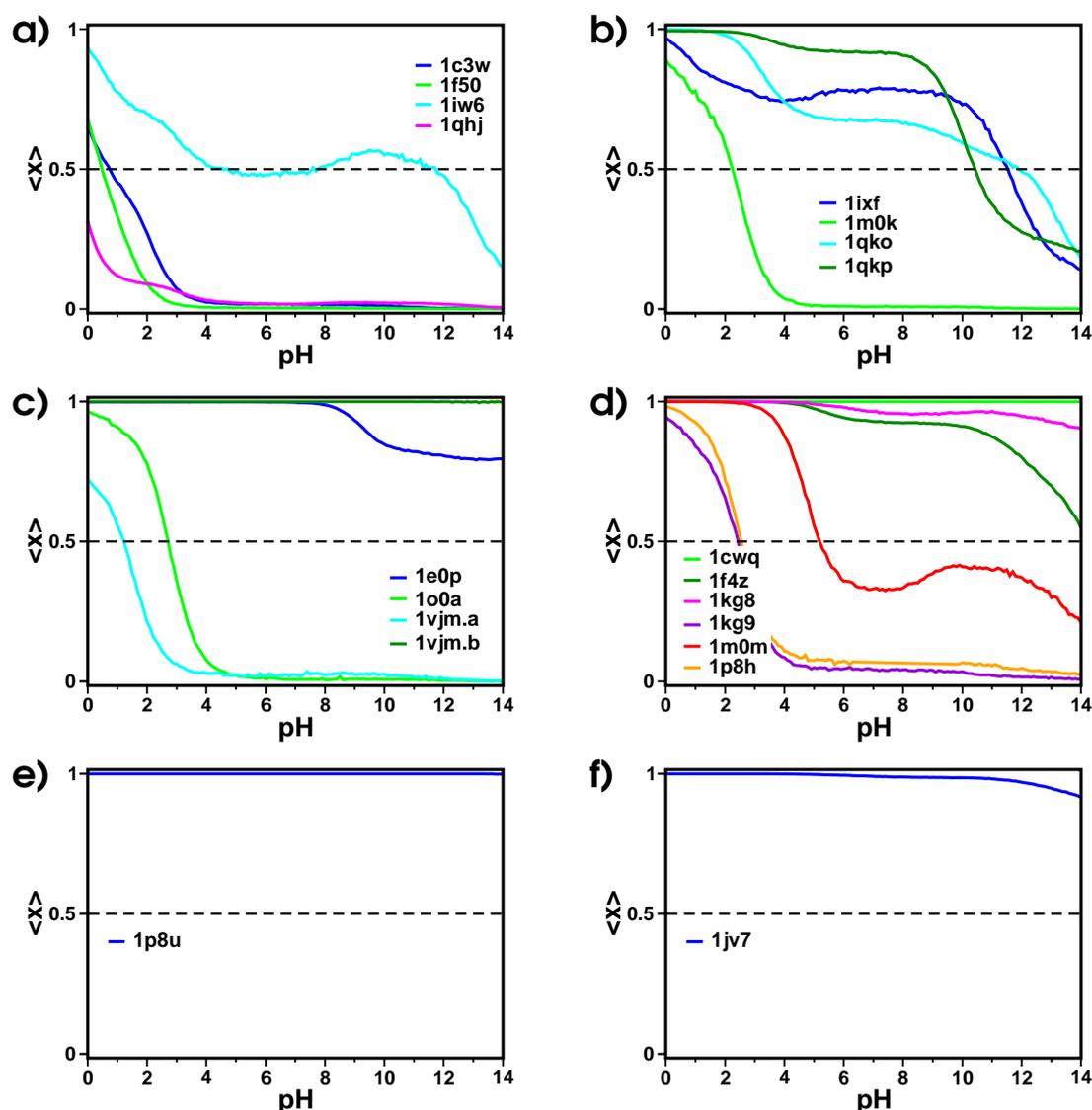


Figure 5.12. Protonation probability of Asp85. The protonation probability of Asp85 is plotted against pH for the different intermediate structures: **a)** bR, **b)** K, **c)** L, and **d)** M structures, **e)** N structure, and **f)** O-like intermediate structure. The curves are assigned to the individual structures by their respective intermediate state and PDB code (*cf.* Table 3.1).

5.4.5 PROTONATION PROBABILITY OF ASP212

Asp212 is located opposite Asp85 on the extracellular side of the retinal Schiff base. In contrast to Asp85 which was identified as the proton acceptor for the retinal Schiff base, Asp212 remains deprotonated during the photocycle. In accordance with experimental data, Asp212 is deprotonated in most BR structures as shown in Figure 5.13.

However, some structures that have shown unusual behavior with respect to other key residues exhibit a protonation probability for this aspartate. In the K structure K:1qkp, Asp212 has a protonation probability of about 0.75 (Figure 5.13b). The protonation be-

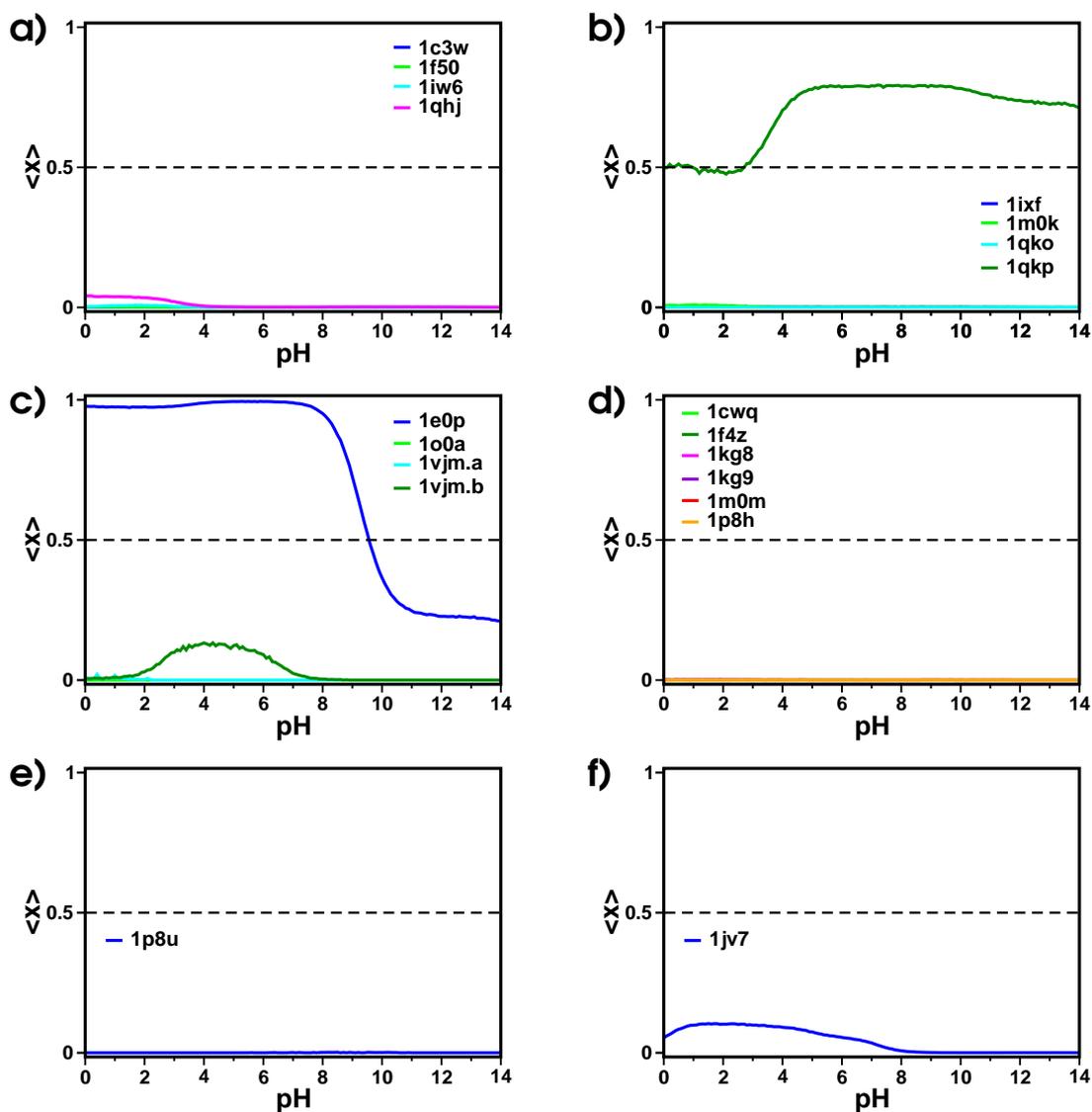


Figure 5.13. Protonation probability of Asp212. The protonation probability of Asp212 is plotted against pH for the different intermediate structures: **a)** bR, **b)** K, **c)** L, and **d)** M structures, **e)** N structure, and **f)** O-like intermediate structure. The curves are assigned to the individual structures by their respective intermediate state and PDB code (cf. Table 3.1).

havior of Asp212 is one of the reasons that K:1qkp has no probability for any functionally meaningful protonation state as shown in Figure 5.4 d. At physiological pH values, Asp212 is virtually protonated in the L structure L:1e0p (Figure 5.13 c). This protonation behavior is the main reason for L:1e0p shows no probability for a functionally meaningful protonation state at this pH as shown in Figure 5.5 d. Additionally, the L structure L:1vjm.b and the O-like structure O:1jv7 show a small protonation probability for Asp212.

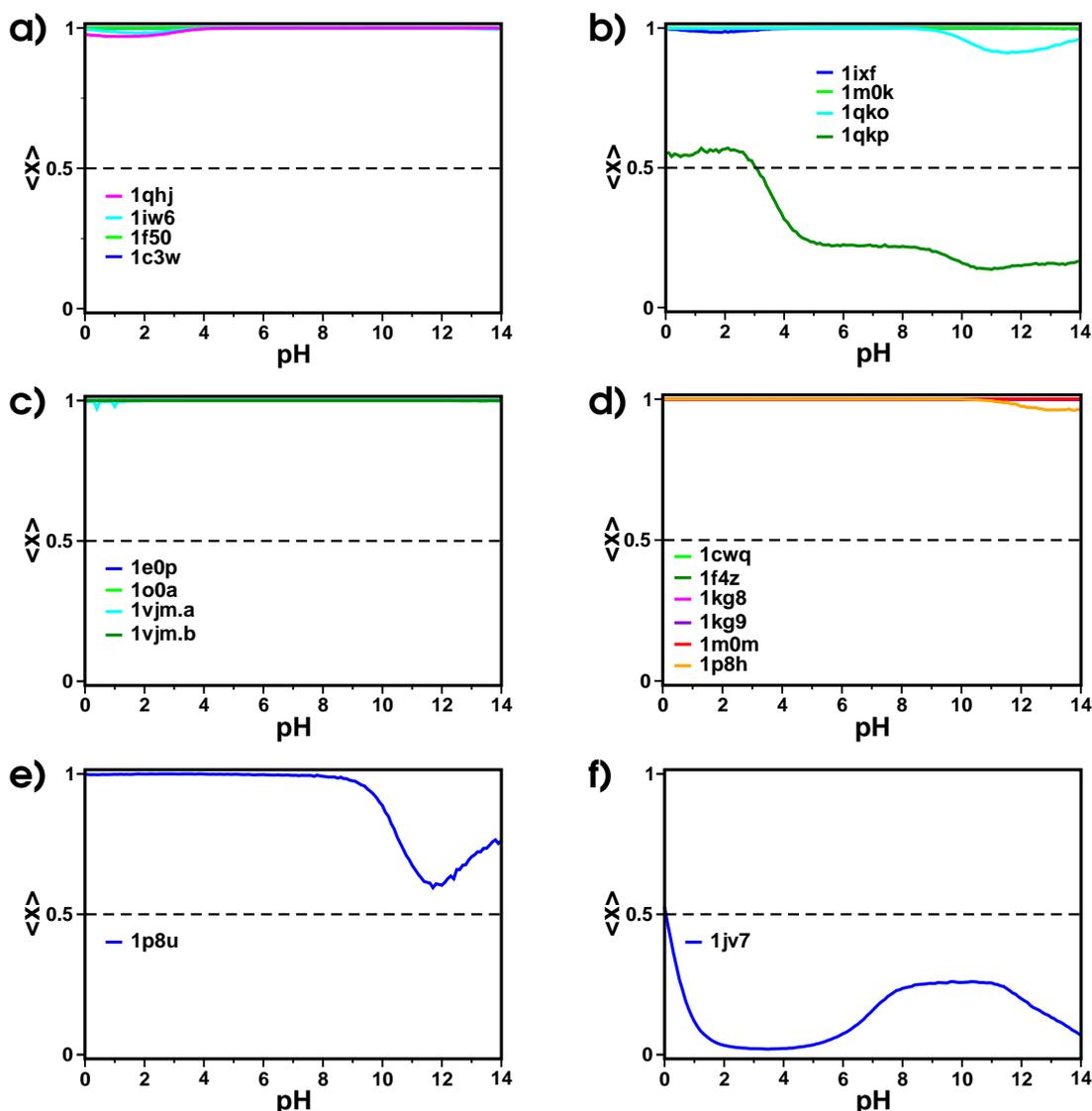


Figure 5.14. Protonation probability of Arg82. The protonation probability of Arg82 is plotted against pH for the different intermediate structures: **a)** bR, **b)** K, **c)** L, and **d)** M structures, **e)** N structure, and **f)** O-like intermediate structure. The curves are assigned to the individual structures by their respective PDB code (*cf.* Table 3.1).

5.4.6 PROTONATION PROBABILITY OF ARG82

Arg82 is located between Asp85 and the proton release group Glu194/Glu204. Since this arginine is connected to the proton transfer path between Asp85 and Glu104/Glu204, it may serve as a transient proton donor and acceptor during the O to bR transition, when Asp85 deprotonates and Glu194/Glu204 protonates. In general, however, Arg82 is assumed to be protonated in all photocycle intermediate protonation states. Accordingly, for most BR structures a protonated Arg82 can be observed as shown in Figure 5.14. The only exceptions are the K structure K:1qkp and the O-like structure (Figure 5.14 b and 5.14 f, respectively). The deprotonated and, therefore, neutral Arg82 is one of the reasons

that the O-like structure does not show a significant probability for a functionally meaningful protonation state (*cf.* Figure 5.8). The K structure K:1qkp shows non-physiological protonation behavior also for Asp212 (*cf.* Figure 5.13 b). The protonation behavior of Arg82 and Asp212 together cause the high probability for the non-physiological protonation state in K:1qkp (*cf.* Figure 5.4).

5.4.7 PROTONATION PROBABILITY OF THE PROTON RELEASE GROUP

The proton release group consists of Glu194 and Glu204. The water molecules located in between these two glutamate residues are represented by a continuum with a high permittivity. The two glutamate residues are considered as one group, Glu194/Glu204, that can bind two protons. The protonation probability of these two glutamate residues is considered together. So far the protonation probability of single residues is considered. For residues that can bind one proton, the protonation probability equals the expected value for the number of protons bound. Glu194/Glu204 can bind up to two protons. Therefore, the expected value for the number of protons bound to the group takes values between 0 and 2. This expected value is given by the sum of the individual protonation probabilities of Glu194 and Glu204. For instance, an expected value of 2 indicates that both Glu194 and Glu204 are protonated.

In the [bR] and [M₁] state, Glu194/Glu204 bind one proton that is released to the extracellular side during the M₁ to M₂ transition. Glu194/Glu204 are deprotonated in the [M₂], [N] and [O] protonation state. The protonation of Glu194/Glu204 takes place during the last step of the photocycle after which the protein has returned to the ground state.

The bR and K structures have a doubly protonated Glu194/Glu204 at low pH values as shown in Figure 5.15 a and b, respectively. At physiological pH values, exactly one proton is bound to Glu194/Glu204. At high pH values, in most bR and K structures still one proton is bound to Glu194/Glu204. Only the bR structure bR:1f50 and the K structure K:1qko show a tendency for a completely deprotonated proton release group at higher pH values.

Two L structures, L:1o0a and L:1vjm.a, bind one proton to Glu194/Glu204 over the whole pH range as shown in Figure 5.15 c. The other two L structures, L:1e0p and L:1vjm.b, are the only BR structures with a completely deprotonated Glu194/Glu204 at physiological pH values. A deprotonated Glu194/Glu204 is characteristic for the [M₂], the [N] and the [O] protonation state. For L:1vjm.b, a high probability for both the [M₂] and the [O] state can indeed be seen (*cf.* Figure 5.5 c). In contrast, L:1e0p shows unusual protonation behavior for Asp212 and has, thus, no probability for any of the later photocycle intermediate protonation states at physiological pH values (*cf.* Figure 5.13 c). At higher pH values, however, a probability for the [M₂] state can be observed reflecting the fact that Asp212 has deprotonated and Glu194/Glu204 are also deprotonated (*cf.* Figure 5.5 d).

In the M structures, Glu194/Glu204 binds in general one proton in the pH range between 0 and 8 as can be seen in Figure 5.15 d. A deprotonation of Glu194/Glu204 can only be seen at higher pH values. Thus, for these M structures, a probability for the [M₂] can only be observed at high pH values (*cf.* Figure 5.6). As discussed above at physiological

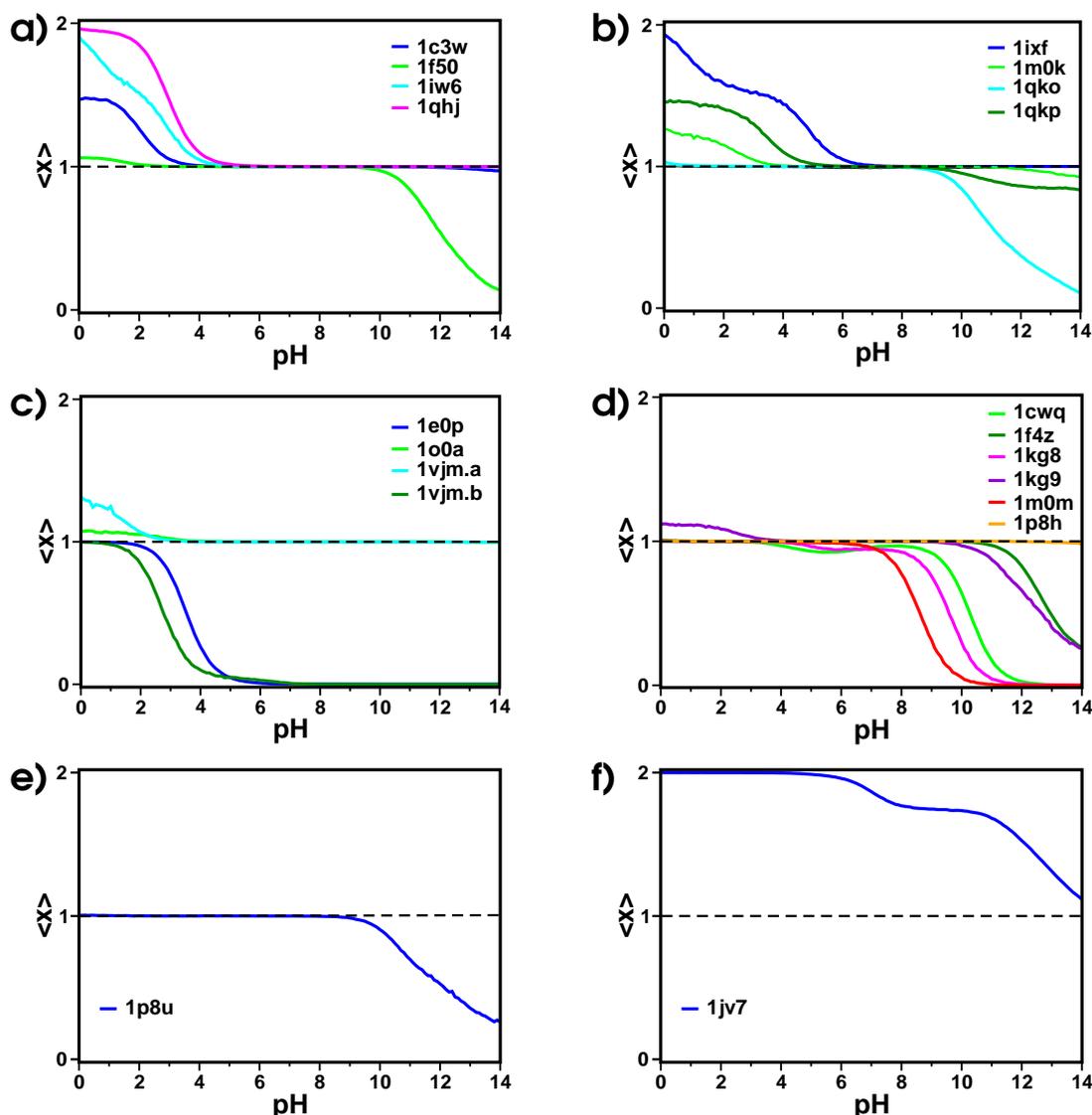


Figure 5.15. Sum of protonation probabilities (expected value) of Glu194/Glu204.

The two glutamate residues of the proton release group, Glu194 and Glu204, are considered as one group that can bind two protons. The sum of protonation probabilities ranges, therefore, from 0 to 2. It is plotted against pH for the different intermediate structures: **a)** bR, **b)** K, **c)** L, and **d)** M structures, **e)** N structure, and **f)** O-like intermediate structure. The curves are assigned to the individual structures by their respective intermediate state and PDB code (*cf.* Table 3.1).

pH values, a probability for the $[M_2]$ state can only be seen for the L structure L:1vjm.b. With respect to the protonation of Glu194/Glu204, the N structure behaves like the M structures as can be seen by comparing Figure 5.15 e and 5.15 d. Since Glu194/ Glu204 are deprotonated in the later photocycle intermediate protonation states, *i.e.*, the $[M_2]$, $[N]$ and $[O]$ state, the N structure shows no probability for these states at physiological pH. At high pH values, however, this structure has a low probability for both the $[M_2]$ and the $[N]$ state (*cf.* Figure 5.7).

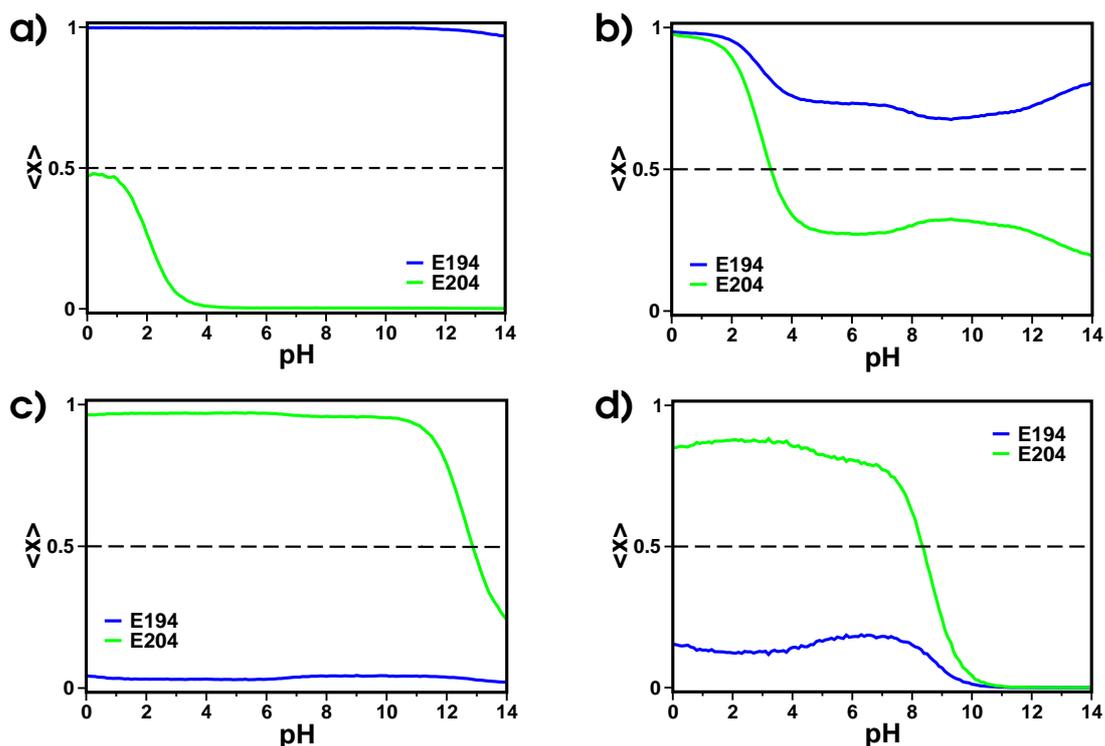


Figure 5.16. Individual protonation probabilities of Glu194 and Glu204. The protonation probabilities of Glu194 and Glu204 are plotted against pH for: **a)** bR:1c3w, **b)** bR:1qhj, **c)** M:1f4z and **d)** M:1m0m. The examples are chosen to give a general impression of the individual behavior of Glu194 and Glu204. The structures are named by their intermediate state and PDB code (*cf.* Table 3.1).

In the O-like structure, Glu194/Glu204 bind two protons at low pH values. With increasing pH, the protonation probability decreases slowly, but remains above 1 as can be seen in Figure 5.15f. As discussed above, Arg82 is deprotonated in this structure (Figure 5.14f). Since this structure was resolved from the Asp85Ser mutant of BR, non-physiological protonation behavior may have been expected for the mutated Asp85 and the protonatable residues closest to it, *i.e.*, the retinal Schiff base and Asp212. Instead, the protonatable residue that bridge the distance between Asp85 and the extracellular space, *i.e.*, Arg82 and Glu194/Glu204, show a tendency to assume their neutral state.

Figure 5.16 depicts individual protonation probabilities for Glu194 and Glu204. Exemplary four structures were chosen to give an overview of the protonation behavior of Glu194 and Glu204. As can be seen, the protonation probabilities of Glu194 and Glu204 add up to 1 at physiological pH values. Interestingly, the distribution of the proton varies significantly between the structures. For bR:1c3w, in general Glu194 is protonated while Glu204 is deprotonated. In contrast for M:1f4z, Glu194 is deprotonated while Glu204 is protonated. For bR:1qhj and M:1m0m, both Glu194 and Glu204 are partly protonated.

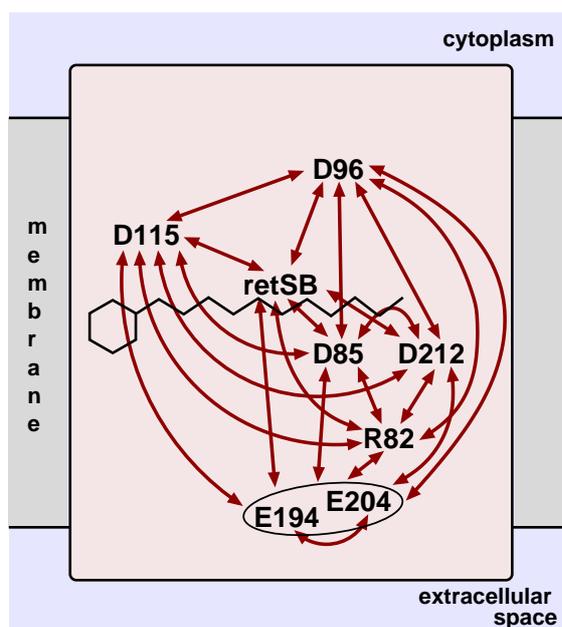


Figure 5.17. Correlation between protonatable residues of bacteriorhodopsin. The eight key residues of proton transfer are shown. 28 pair-correlations exist between these eight residues. In this schematic, the two residues that constitute the proton release group, Glu194 and Glu204, are grouped together. The remaining 21 pair-correlations are indicated by the red arrows. Additionally, the correlation between Glu194 and Glu204 is indicated. For the actual calculation Glu194 and Glu204 are considered individually.

5.5 CORRELATION OF THE PROTONATION BEHAVIOR

This section provides an analysis of the pair-correlation of the protonation behavior of BR's protonatable sites. As detailed in Subsection 2.5.2, the Pearson product-moment correlation coefficient has been used in this thesis. This coefficient is obtained by dividing the covariance of two variables by the product of their standard deviations [131]. The correlation coefficient indicates the strength of a linear relationship between two variables. A strong correlation between two variables is, however, not proof of a causal connection between these variables. Correlation between two variables merely suggests that knowledge of the behavior of one variable can help to predict the behavior of the other.

To verify the significance of a correlation, the behavior of the two variables should, therefore, be individually analyzed. With respect to the correlation between the protonation of two residues this can be achieved by examining probabilities for protonation substates and the protonation probabilities of the residues. Furthermore, the interaction energy may be of interest. In general, correlations between protonatable sites that are far from each other are unlikely to reflect a causal connection. Such claims can be verified by checking the interaction energy of two protonatable sites with each other. In the case that the interaction energy is zero, the sites do not interact.

As has been shown in the derivation of the pair-correlation in Subsection 2.5.2, the cor-

relation is a normalization of the covariance of two random variables. Thus, the behavior of two sites can only be correlated, when there is variance in their behavior, *i.e.*, when the variance is unequal to 0.

The correlation between all 39 protonatable residues has been calculated for all BR structures, *i.e.*, 741 pair-correlations for each of the 20 BR structures. Results from a representative subset of the structures have been chosen to provide an adequate overview of the correlation behavior in BR. The following six structures are discussed in this section: two bR structures, bR:1c3w and bR:1f50, one L structure, L:1vjm.b, two M structures, M:1f4z and M:1kg8 and the N structure N:1p8u. Although the K and the L intermediate state do not differ in their protonation state from the bR state, the L structure L:1vjm.b is included here, since it has a high probability for both the [M₂] or the [O] protonation state (*cf.* Table 5.1 and Figure 5.5 c).

Due to the large amount of data, the focus of the following discussion will - as has been the case in the previous sections of this chapter - lie on the functionally relevant groups of BR. The eight key residues of proton transfer in BR are: Asp96, Asp115, retinal Schiff base, Asp85, Asp212, Arg82 and the two residues of the proton release group, Glu194 and Glu204. For these residues 28 pair-correlations exist. Figure 5.17 shows a schematic of BR indicating the pair-correlations between the key residues by the connecting arrows. The graphs in Figure 5.18 to 5.22 depict 9 of the 28 pair-correlations of the key residues. Presented are correlations for the six BR structures mentioned above, *i.e.*, in total 66 correlations are shown that will be analyzed in the following subsections.

5.5.1 CORRELATION BETWEEN ASP85 AND THE RETINAL SCHIFF BASE

The retinal Schiff base is located in the center of BR and Asp85 close to the Schiff base on its extracellular side. In the [bR] protonation state, the retinal Schiff base is protonated and Asp85 is deprotonated. In the [M₁] protonation state, the proton has been transferred from the Schiff base to Asp85. In both the [M₁] and [M₂] protonation state, the retinal Schiff base is deprotonated. In the [N] and [O] protonation state, the Schiff base is again protonated. Asp85 remains protonated, too, and only deprotonates when BR returns to the [bR] state. The photocycle and its proton transfer steps are depicted in Figure 5.2 and the intermediate protonation states are listed in Table 5.1.

In Figure 5.18, the protonation behavior of Asp85 and the retinal Schiff base is depicted for six BR structures. As can be seen, the protonation behavior of Asp85 and the retinal Schiff base is strongly negatively correlated in bR:1c3w, bR:1f50, M:1f4z and M:1kg8. In L:1vjm.b, the behavior of Asp85 and the retinal Schiff base is uncorrelated in the pH range from 0 to 11. At pH > 11, the correlation coefficient fluctuates between 0 and approximately -1, *i.e.*, perfect negative correlation. In N:1p8u, the behavior of Asp85 and the retinal Schiff base is uncorrelated in the pH range from 0 to 13.5. At pH > 13.5, a very slight negative correlation can be observed. In the following subsections, the correlation behavior is discussed in detail for the individual structures.

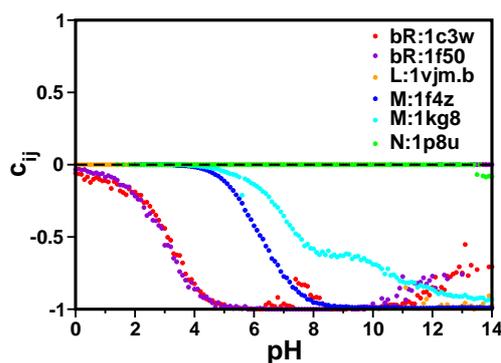


Figure 5.18. Correlation between Asp85 and the retinal Schiff base. The coefficient of the correlation between Asp85 and the retinal Schiff base is plotted against pH for six BR structures.

BR STRUCTURES

For bR:1c3w and bR:1f50, the correlation coefficient reaches its minimum value of -1 at pH=5 and the correlation remains close to -1 in the pH range from 5 to 12. The protonation probability of the retinal Schiff base is close to 1 over the complete pH range from 0 to 14 for both bR:1c3w and bR:1f50 (*cf.* Figure 5.11 a). At pH > 4, the protonation probability of Asp85 is close to 0 for both bR:1c3w and bR:1f50 (*cf.* Figure 5.12 a). This protonation behavior and the negative correlation indicate that Asp85 is deprotonated in all protonation states, where the retinal Schiff base is protonated and *vice versa*. As specified in Table 5.1, the [bR] protonation state, for which bR:1c3w and bR:1f50 have an almost exclusive probability as shown in Figure 5.3 a and c, respectively, has a protonated Schiff base and a deprotonated Asp85.

L STRUCTURE

For L:1vjm.b the correlation coefficient is 0 in the pH range from 0 to 11. However, at pH > 13.5, a significant negative correlation can be observed at several pH values. As shown in Figure 5.11 c, at pH < 2 the protonation probability of the retinal Schiff base is close to 1. In the pH range from 2 to 8, the Schiff base deprotonates. At pH > 8, the protonation probability is close to 0. The protonation probability of Asp85 is 1 in the pH range from 0 to 11 (*cf.* Figure 5.12 c). At pH > 11, probability values of 0.99 are observed several times: at pH=11.1, 11.7, 11.8, 12.3, 12.5 to 12.7, 12.9 to 13.5, 13.7, 13.9 and 14. At each of these 17 pH values, the correlation is significantly negative, *i.e.*, the coefficient takes values between -0.85 and -1.

This case illustrates one important characteristic of the correlation coefficient: when one of the variables does not show any variance, the covariance between this variable and another equals 0. Consequently, the correlation coefficient will be 0, too. That is, without variance the occurrence of positively or negatively correlated behavior cannot be assessed. In the case of L:1vjm.b, Asp85 does not show any variance, *i.e.*, is fully protonated, at all pH values except at the 17 values mentioned above. As a result, the covariance and the correlation between Asp85 and the retinal Schiff base (or any other

protonatable group) are 0 at all pH values except those 17 values listed above. Thus, although the difference between the behavior of the two residues for example at pH = 12.2 and pH = 12.3 is very small, in terms of the correlation coefficient it is amplified to a difference between uncorrelated and perfectly negatively correlated behavior.

M STRUCTURES

For M:1f4z, the correlation coefficient is close to -1 in the pH range from 8 to 14. For M:1kg8, the correlation decreases from 0 to -1 in the pH range from 5 to 14. The negative correlation between Asp85 and the retinal Schiff base is more pronounced in M:1f4z than in M:1kg8. The protonation curves of Asp85 and the retinal Schiff base are, however, similar in M:1f4z and M:1kg8. As depicted in Figure 5.11 d, at pH < 4 the protonation probability of the retinal Schiff base is close to 1. In the pH range from 4 to 8, the Schiff base deprotonates. In M:1f4z, the minimum protonation probability is approximately 0.15. At pH = 10, the protonation probability increases up to approximately 0.5 at pH = 14. In M:1kg8, the protonation probability is close to 0 in the pH range from 8 to 14. As shown in Figure 5.12 d, at pH < 3.5 the protonation probability of Asp85 is 1. At pH > 3.5, Asp85 deprotonates. In M:1kg8, the protonation probability remains above 0.8, while in M:1f4z, the protonation probability decreases to about 0.5 at pH = 14. In the [M₁] or [M₂] protonation state Asp85 is protonated and the retinal Schiff base is deprotonated as characterized in Table 5.1. As depicted in Figure 5.6, for M:1f4z and M:1kg8 the [M₁] and [M₂] protonation state are the primary states. Additionally at high pH values, M:1f4z has a probability for the [bR] protonation state, where Asp85 is deprotonated and the Schiff base deprotonated. This is in accordance with the protonation behavior of Asp85 and the Schiff base in M:1f4z and also with the perfect negative correlation. In M:1f4z, when Asp85 is protonated, the Schiff base is deprotonated and *vice versa*. In the first case either a probability for the [M₁] or the [M₂] protonation state is observed and in the latter for the [bR] state. In M:1kg8, the negative correlation between Asp85 and the retinal Schiff base is not as strong. This indicates that M:1kg8 has a slight probability for protonation states, where Asp85 and the retinal Schiff base are both either protonated or deprotonated. In general, however, their behavior is converse.

N STRUCTURE

For N:1p8u the correlation coefficient is 0 in the pH range from 0 to 13.4. In the pH range from 13.5 to 14 a small negative correlation can be observed. As shown in Figure 5.11 e, at pH < 3 the retinal Schiff base is protonated. In the pH range from 3 to 8, the Schiff base deprotonates. At pH = 8, the protonation probability is close to 0. In the pH range from 8 to 14, the probability increases slightly to approximately 0.2. In the pH range from 0 to 13.4, Asp85 is fully protonated (*cf.* Figure 5.12 e). Since Asp85 does not show any variance, the covariance and the correlation between Asp85 and the Schiff base are 0. In the pH range from 13.5 to 14, probability values of 0.99 are observed. Only at these pH values can the covariance and correlation be assessed, which is negative but not very strong. As can be seen, a correlation coefficient of 0 does not necessarily indicate that two variable are not correlated. It can also indicate that one or both of the variable does not show any variance.

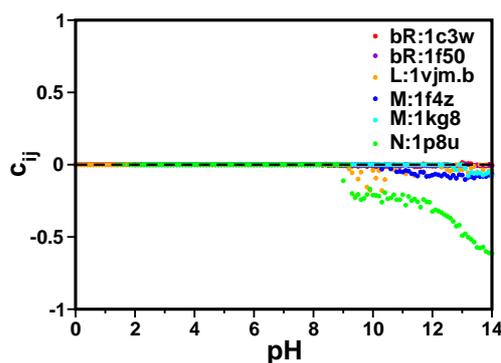


Figure 5.19. Correlation between Asp96 and the retinal Schiff base. The coefficient of the correlation between Asp96 and the retinal Schiff base is plotted against pH for six BR structures.

5.5.2 CORRELATION BETWEEN ASP96 AND THE RETINAL SCHIFF BASE

The retinal Schiff base is located in the center of BR and Asp96 is located in between the cytoplasmic side of BR and the Schiff base. In the [bR] protonation state, Asp96 and the retinal Schiff base are protonated. In both the [M₁] and [M₂] protonation state, the retinal Schiff base is deprotonated, while Asp96 remains protonated. In the [N] protonation state, the proton is transferred from Asp96 to the Schiff base, *i.e.*, Asp96 is deprotonated and the retinal Schiff base is protonated. The photocycle and its proton transfer steps are depicted in Figure 5.2 and the intermediate protonation states are listed in Table 5.1.

The correlation between Asp96 and the retinal Schiff base is shown in Figure 5.19 for six BR structures. As depicted, the protonation behavior of Asp96 and the retinal Schiff base is mostly uncorrelated. Only in N:1p8u, the behavior of Asp96 and the retinal Schiff base is significantly negatively correlated at pH > 10. In the following subsections, the correlation behavior is discussed in detail for the individual structures.

BR STRUCTURES

For bR:1c3w and bR:1f50, the correlation coefficient is 0 or very close to 0 in the pH range from 0 to 14. The protonation probability of the Schiff base is about 0.99 for both bR structures in the pH range from 0 to 14 (*cf.* Figure 5.11 a). As shown in Figure 5.9 a, Asp96 is fully protonated in the pH range from 0 to 8. Due to the missing variance of the fully protonated Asp96, the covariance and the correlation coefficient are 0 at pH < 8. At pH > 8, the protonation probability decreases slightly, but remains above 0.9. The correlation coefficient deviates, however, only minimally from 0. Thus, in bR:1c3w and bR:1f50, the protonation behavior of Asp96 and the retinal Schiff base is uncorrelated.

L STRUCTURE

For L:1vjm.b, the correlation coefficient of Asp96 and the retinal Schiff base is 0 in the pH range from 0 to 9. At pH > 9, the correlation coefficient fluctuates between 0 and -0.2. In L:1vjm.b, the protonation probability of the Schiff base decreases from nearly 1 at low

pH values to nearly 0 at pH=8 (*cf.* Figure 5.11 c). At pH > 10 at several pH values, the retinal Schiff base has a protonation probability of 0. At pH < 9, Asp96 is fully protonated. At higher pH values, the protonation probability is still high but decreases to 0.93. At pH < 9, the correlation is 0 due to the fully protonated Asp96 (*cf.* Figure 5.9 c). At pH > 10, the correlation is 0 at those pH values where the retinal Schiff base is fully deprotonated. When both Asp96 and the Schiff base show a slight variance, the protonation of Asp96 and the retinal Schiff base is slightly negatively correlated.

M STRUCTURES

For M:1f4z, the correlation coefficient is 0 at pH < 10. At higher pH values the protonation of Asp96 and the retinal Schiff base is slightly negatively correlated with a correlation coefficient of about -0.1. For M:1kg8, the correlation coefficient is 0 at pH < 12.5. Thereafter, a slight negative correlation of less than -0.1 can be observed. The protonation probabilities for Asp96 and the Schiff base are shown in Figure 5.9 d and Figure 5.11 d, respectively. At pH < 2 in both M:1f4z and M:1kg8, Asp96 as well as the retinal Schiff base are fully protonated. Thereafter, the retinal Schiff base deprotonates. In M:1f4z, Asp96 is fully protonated up to pH=10.2. In M:1kg8, Asp96 is fully protonated up to pH=12.7. In both M:1f4z and M:1kg8, the protonation probability of Asp96 remains above 0.95. At pH < 2, none of the sites shows any variance. Accordingly, the correlation coefficient is 0. Due to the absent variance in Asp96, the correlation coefficient is 0 in M:1f4z at pH ≤ 10.2 and in M:1kg8 at pH ≤ 12.7.

N STRUCTURE

For N:1p8u, the correlation between Asp96 and the retinal Schiff base is 0 in the pH range from 0 to 9. At pH > 9, the correlation coefficient decreases to -0.6. At very low pH values, *i.e.*, pH < 1.5, the retinal Schiff base is fully protonated (*cf.* Figure 5.11 e). Thereafter, the protonation probability decreases. At pH=7, the Schiff base is virtually deprotonated. The protonation probability of Asp96 is 1 in the pH range from 0 to 9 (*cf.* Figure 5.9 e). At pH > 9, Asp96 deprotonates slightly and at pH=8, its protonation probability is 0.8. As can be seen in Figure 5.7, in this pH range N:1p8u shows a probability for the [M₂] state, where Asp96 is protonated and the Schiff base is deprotonated, and the [N] state, where Asp96 is deprotonated and the Schiff base is protonated. In accordance with the high protonation probability of Asp96, the probability for the [N] state is low while the probability for the [M₂] state is significantly higher in this pH range.

At very low pH values, neither Asp96 nor the Schiff base show any variance. Therefore, the correlation coefficient is 0. In the pH range from 1.5 to 9, the correlation is 0 due to the absent variance of Asp96. At pH > 9, when both Asp96 and the retinal Schiff base show variance, the correlation is increasingly negative. The negative correlation is in accordance with the protonation states [M₂] and [N], where the protonation behavior of Asp96 and the retinal Schiff base is contrary to each other. However, the negative correlation coefficient of -0.6 indicates that the behavior is not contrary in all states. Indeed, the sum of the probabilities for the [M₂] and [N] state is less than 1, *i.e.*, approximately 0.7. Thus, N:1p8u has a probability for additional non-physiological states, where both Asp96

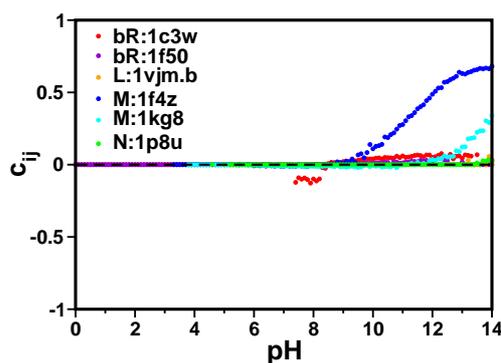


Figure 5.20. Correlation between Asp85 and Asp115. The coefficient of the correlation between Asp85 and Asp115 is plotted against pH for six BR structures.

and the Schiff base show the same behavior. The protonation probabilities of Asp96 and the Schiff base indicate that they are both deprotonated in these states.

5.5.3 CORRELATION BETWEEN ASP85 AND ASP115

Asp85 is located on the extracellular side of the retinal Schiff base in the center of BR. Asp115 is located close to the β -ionone ring of the retinal on its cytoplasmic side. Asp115 is not a proton donor or acceptor during the physiological proton transfer in BR. However, it is implicated in regulating the proton transfer. Asp115 is protonated in all intermediate protonation states of BR. Asp85 is deprotonated in the [bR] state and protonated in all other protonation states. The photocycle and its proton transfer steps are depicted in Figure 5.2 and the intermediate protonation states are listed in Table 5.1.

In Figure 5.20, the protonation behavior of Asp85 and Asp115 is depicted for six BR structures. In general, the protonation behavior of the two residues is uncorrelated. Only in M:1f4z and at $\text{pH} > 10$, the behavior of Asp85 and Asp115 is significantly correlated. Interestingly, the correlation is positive. In the following subsections, the correlation behavior is discussed in detail for the individual structures.

BR STRUCTURES

For bR:1c3w and bR:1f50, the correlation coefficient is 0 or very close to 0 in the pH range from 0 to 14. The protonation probability of Asp85 is slightly larger than 0.5 at $\text{pH} = 0$, but decreases rapidly with increasing pH as shown in Figure 5.12 a. At $\text{pH} > 3$, the protonation probability is close to 0 for both structures. As depicted in Figure 5.10 a, Asp115 is fully protonated in the pH range from 0 to 7.5 in both structures. Thereafter, Asp115 deprotonates. At $\text{pH} = 14$, the protonation probability is less than 0.1 in bR:1c3w and about 0.2 in bR:1f50. Due to the fully protonated Asp115, the covariance and the correlation coefficient are 0 at $\text{pH} < 7.5$. At higher pH values, the correlation coefficient is still close to 0. Between pH 7.5 and 8.5 a slight negative correlation can be observed for bR:1c3w. At $\text{pH} > 10$, the small correlation takes positive values.

L STRUCTURE

For L:lvjm.b, the correlation between Asp85 and Asp115 is 0. Only at $\text{pH} > 13$, a minimal positive correlation can be observed at four pH values. As shown in Figure 5.10 c, Asp115 is fully protonated. Only at $\text{pH} > 13$, a protonation probability slightly less than 1 can be observed at five pH values. In the pH range from 0 to 11, Asp85 is fully protonated (*cf.* Figure 5.12 c). At $\text{pH} > 11$, its protonation probability is slightly smaller than 1 at 17 pH values. Since the protonation probability of Asp85 is 1 at one of the five pH values the probability of Asp115 is unequal to 1, a small correlation is observed only at four pH values.

M STRUCTURES

For M:1f4z, the correlation coefficient is 0 in the pH range from 0 to 9.5. At higher pH values, the correlation increases and Asp85 and Asp115 are significantly positively correlated. For M:1kg8, the correlation coefficient is 0 in the pH range from 0 to 12.5. Thereafter, the correlation increases to 0.3 at $\text{pH} = 14$. As shown in Figure 5.12 d, at $\text{pH} < 3.5$ the protonation probability of Asp85 is 1. At $\text{pH} > 3.5$, Asp85 deprotonates. In M:1kg8, the protonation probability remains above 0.8, while in M:1f4z, the protonation probability decreases to about 0.5 at $\text{pH} = 14$. In the pH range from 0 to 6, Asp115 is fully protonated in M:1f4z and M:1kg8 (*cf.* Figure 5.10 d). In the pH range from 6 to 10, the protonation probability is close to 1 for both structures. In the pH range from 10 to 14, the probability remains high in M:1kg8, while in M:1f4z the probability decreases to 0.6. Thus, a significant positive correlation is observed, when Asp115 partly deprotonates. Since the protonation probability of Asp115 decreases further in M:1f4z than in M:1kg8, the positive correlation is more pronounced in M:1f4z.

N STRUCTURE

For L:lvjm.b, the correlation between Asp85 and Asp115 is 0 in the pH range from 0 to 13.4. At $\text{pH} \geq 13.5$ at five pH values, a negligible positive correlation can be seen. As shown in Figure 5.10 e, Asp115 is fully protonated up to $\text{pH} = 5$. Up to $\text{pH} = 9$, the protonation probability is 0.99. Thereafter, Asp115 slowly deprotonates. At $\text{pH} = 14$, its protonation probability is approximately 0.6. In the pH range from 0 to 13.4, the protonation probability of Asp85 is 1 (*cf.* Figure 5.12 e). At $\text{pH} \geq 13.5$ at five pH values, a probability of 0.99 is observed. When Asp85 or Asp115 do not show any variance, the covariance and the correlation between Asp85 and Asp115 are 0. Although the data shows that the behavior of Asp85 and Asp115 is not correlated in N:1p8u, no meaningful conclusion can be drawn due to the low variance of Asp85.

5.5.4 CORRELATION BETWEEN ASP115 AND THE RETINAL SCHIFF BASE

The retinal Schiff base is located in the center of BR. Asp115 is located close to the β -ionone ring of the retinal on its cytoplasmic side. As stated in the previous subsection, Asp115 is not a proton donor or acceptor during the proton transfer in BR. It is, however, implicated in regulating the proton transfer. Asp115 is protonated in all intermediate

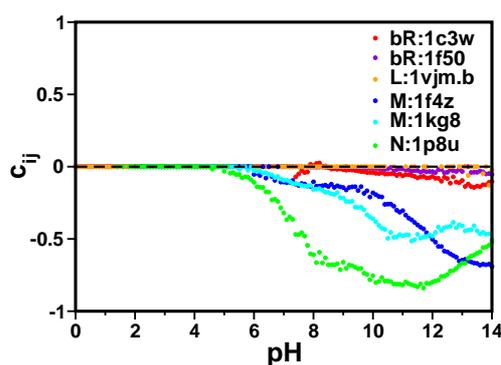


Figure 5.21. Correlation between Asp115 and the retinal Schiff base. The coefficient of the correlation between Asp115 and the retinal Schiff base is plotted against pH for six BR structures.

protonation states of BR. The retinal Schiff base is protonated in the [bR], [N] and [O] protonation state and deprotonated in the [M₁] and [M₂] state. The photocycle and its proton transfer steps are depicted in Figure 5.2 and the intermediate protonation states are listed in Table 5.1.

The correlation between Asp115 and the retinal Schiff base is shown in Figure 5.21 for six BR structures. No significant correlation can be observed for bR:1c3w, bR:1f50 or L:1vjm.b. In the two M structures, M:1f4z and M:1kg8, and the N structure N:1p8u, however, the protonation behavior of Asp115 and the retinal Schiff base is negatively correlated. In the following subsections, the correlation behavior is discussed in detail for the individual structures.

BR STRUCTURES

For both bR:1c3w and bR:1f50, the correlation coefficient is 0 at pH < 7.5. In the pH range from 7.5 to 14, the correlation coefficient takes values between 0 and -0.1. As shown in Figure 5.10 a, Asp115 is fully protonated in the pH range from 0 to 7.5 in both structures. Thereafter, Asp115 deprotonates. At pH = 14, the protonation probability is less than 0.1 in bR:1c3w and about 0.2 in bR:1f50. The protonation probability of the retinal Schiff base is close to 1 over the complete pH range from 0 to 14 for both bR:1c3w and bR:1f50 (*cf.* Figure 5.11 a). In the pH range from 0 to 7.5, the covariance and the correlation are 0 for both bR state structures due to the absent variance in Asp115. At higher pH values, when the correlation can be assessed the protonation behavior of Asp115 and the retinal Schiff base is uncorrelated.

L STRUCTURE

For L:1vjm.b, the correlation coefficient is 0. Only at pH > 13 at four pH values, an insignificant negative correlation can be observed. As shown in Figure 5.10 c, Asp115 is fully protonated. Only at pH > 13, a protonation probability slightly less than 1 can be observed at five pH values. As shown in Figure 5.11 c, at pH < 2 the protonation

probability of the retinal Schiff base is close to 1. In the pH range from 2 to 8, the Schiff base deprotonates. At $\text{pH} > 8$, the protonation probability is close to 0 or equal to 0. In general, the correlation coefficient is 0 due to the absent variance of Asp115. However, at one of the five pH values, where Asp115 shows variance, the retinal Schiff base is fully deprotonated. Thus, a correlation coefficient unequal to 0 can only be observed at four pH values. The correlation coefficient at the four pH values where both residues show variance indicates that the protonation behavior of Asp115 and the retinal Schiff base is uncorrelated in L:lvjm.b

M STRUCTURES

For M:1f4z and M:1kg8, the correlation of the protonation behavior of Asp115 and the retinal Schiff base is 0 in the pH range from 0 to 6. At $\text{pH} > 6$, the correlation decreases. In the pH range from 10 to 14, Asp115 and the Schiff base show significant negative correlation in both structures. In M:1f4z, the correlation decreases considerably to about -0.7 at $\text{pH} = 14$. In M:1kg8, the correlation coefficient fluctuates between -0.4 and -0.5. In the pH range from 0 to 6, Asp115 is fully protonated in both M:1f4z and M:1kg8 (*cf.* Figure 5.10 d). In the pH range from 6 to 10, the protonation probability is close to 1 for both structures. In the pH range from 10 to 14, the probability remains high in M:1kg8, while in M:1f4z the probability decreases to 0.6. As shown in Figure 5.11 d, at $\text{pH} < 4$ the protonation probability of the retinal Schiff base is close to 1. In the pH range from 4 to 8, the Schiff base deprotonates. In M:1f4z, the minimum protonation probability is approximately 0.15. At $\text{pH} = 10$, the protonation probability increases up to approximately 0.5 at $\text{pH} = 14$. In M:1kg8, the protonation probability is close to 0 in the pH range from 8 to 14.

In the pH range from 0 to 6, the covariance and the correlation of Asp115 and the retinal Schiff base are 0 due to the absent variance of Asp115 in both structures. In the pH range from 6 to 10, when the protonation probability of Asp115 is 0.99 in both structures, *i.e.*, a small variance is observed, the correlation is slightly negative. In the pH range from 10 to 14, Asp115 is partly deprotonated and the Schiff base is partly protonated in M:1f4z. In contrast in M:1kg8, the protonation probability of Asp115 remains high and that of the Schiff base remains low. The different behavior of Asp115 and the Schiff base in M:1f4z and M:1kg8 results in a more pronounced negative correlation in M:1f4z.

N STRUCTURE

For N:1p8u, the correlation coefficient is 0 in the pH range from 0 to 5. In the pH range from 5 to 11.5, the correlation between Asp115 and the retinal Schiff base decreases to about -0.8. Thereafter, the coefficient increases to -0.5 at $\text{pH} = 14$. As shown in Figure 5.10 e, Asp115 is fully protonated up to $\text{pH} = 5$. Up to $\text{pH} = 9$, the protonation probability is 0.99. Thereafter, Asp115 slowly deprotonates. At $\text{pH} = 14$, its protonation probability is approximately 0.6. As shown in Figure 5.11 e, at $\text{pH} < 3$ the retinal Schiff base is protonated. In the pH range from 3 to 8, the Schiff base deprotonates. At $\text{pH} = 8$, the protonation probability is close to 0. In the pH range from 8 to 14, the probability increases slightly to approximately 0.2. In the pH range from 7 to 14, the protonation behavior of Asp115 and the retinal Schiff base is negatively correlated. Thus, in most

protonation states where Asp115 is protonated, the Schiff base is deprotonated and *vice versa*.

5.5.5 CORRELATION BETWEEN GLU194 AND GLU204

The proton release group consists of the two glutamate residues Glu194 and Glu204. Glu194 and Glu204 are located adverse and close to each other on the extracellular side of BR. In the [bR] protonation state, the proton release group binds one proton. That is, either Glu194 or Glu204 are protonated while the other is deprotonated. In the [M₂] protonation state, the proton is released to the extracellular side. Both glutamate residues are deprotonated in the [M₂], [N] and [O] protonation state. The photocycle and its proton transfer steps are depicted in Figure 5.2 and the intermediate protonation states are listed in Table 5.1.

The correlation between Glu194 and Glu204 is depicted in Figure 5.22 for six BR structures. As shown, the protonation behavior of Glu194 and Glu204 is in general strongly negatively correlated. However, in L:lvjm.b and M:1kg8, the behavior of Glu194 and Glu204 is uncorrelated. In the following subsections, the correlation behavior is discussed in detail for the individual structures.

BR STRUCTURES

For bR:1c3w, a correlation coefficient of approximately 0 is observed at pH=0. In the pH range from 1 to 5 the correlation decreases to nearly -1. In the pH range from 5 to 10, the behavior of Glu194 and Glu204 is perfectly negatively correlated. Thereafter, the correlation increases to -0.2 at pH=14. For bR:1f50, the correlation coefficient is -0.2 at pH=0. Then the correlation decreases. In the pH range from 4 to 8, the behavior of Glu194 and Glu204 is nearly perfectly negatively correlated. Thereafter, the correlation coefficient increases and at pH=14, Glu194 and Glu204 are uncorrelated.

In bR:1c3w, the protonation probability of Glu194 is 0.99 at pH < 12.5. Thereafter, it decreases slightly to 0.97 (*cf.* Figure 5.16 a). The protonation probability of Glu204 is 0.5 from pH=0 to 1, decreases to 0 at pH=4 and remains close to 0 up to pH=14. In bR:1f50, the protonation probability of Glu194 is close to 0 over the whole pH range. And the protonation probability of Glu204 is close to 1 in the pH range from 0 to 9.5. Then, the probability decreases to 0.15 at pH=14. For bR:1f50, the individual probability curves of Glu194 and Glu204 are not shown. The sum of their protonation probabilities is depicted in Figure 5.15 a. In short in bR:1c3w, Glu194 is protonated and Glu204 is deprotonated and in bR:1f50 the protonation is reversed.

In bR:1c3w, the protonation behavior of Glu194 and Glu204 is perfectly negatively correlated when Glu204 is deprotonated. However, at high pH values Glu194 and Glu204 are increasingly uncorrelated even though the protonation behavior of Glu194 and Glu204 remains virtually unchanged. In bR:1f50, the protonation behavior of Glu194 and Glu204 appears contrary to each other in the pH range from 0 to 9.5. Nevertheless, at low pH values the negative correlation is not pronounced. However, in the pH range from 4 to 8, Glu194 and Glu204 are perfectly negatively correlated. Then, the deprotonation of

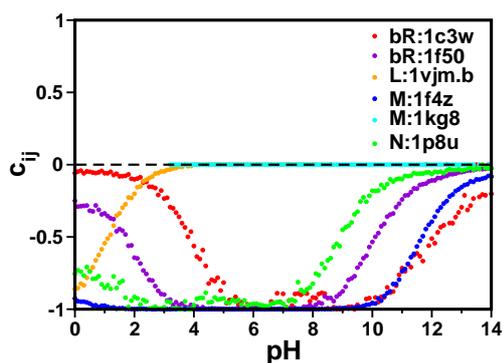


Figure 5.22. Correlation between Glu194 and Glu204. The coefficient of the correlation between the residues of the proton release group, Glu194 and Glu204, is plotted against pH for six BR structures.

Glu204 at pH=9.5 results in a correlation coefficient of 0. Although in bR:1f50, Glu194 and Glu204 are mostly deprotonated at pH = 14, their behavior is uncorrelated.

L STRUCTURE

At pH=0, the correlation between Glu194 and Glu204 is -0.85 for L:1vjm.b. In the pH range from 0 to 4, the correlation coefficient increases to 0. At pH > 4, the protonation behavior between Glu194 and Glu204 is uncorrelated in L:1vjm.b. In the pH range from 0 to 4, Glu194 is nearly deprotonated. At pH > 4, its protonation probability is exactly 0. At pH=0, the protonation probability of Glu204 is close to 1. Then, the protonation probability decreases to approximately 0.1 at pH=4 and approximately 0 at pH=5. In the pH range from 8.5 to 10, Glu204 is completely deprotonated. The individual probability curves of Glu194 and Glu204 are not shown. The sum of their protonation probabilities is depicted in Figure 5.15 c.

In the pH range from 0 to 4, the variance of Glu194 decreases. Additionally, the protonation probability of Glu204 decreases. In this pH range, the correlation between Glu194 and Glu204 vanishes although it is significantly negative at first. This may be either due to the decreasing variance of Glu194 or the fact that Glu204 deprotonates. Glu194 is virtually deprotonated and when Glu204 deprotonates, their behavior resembles each other. Nevertheless, no positive correlation is observed and the negative correlation is insignificant. At pH > 4, the correlation coefficient is 0 due to the missing variance of Glu194 which is fully deprotonated.

M STRUCTURES

For M:1f4z, the correlation between Glu194 and Glu204 is close to -1 at pH < 10. In the pH range from 10 to 14, the coefficient increases to -0.1. For M:1kg8, the correlation coefficient is 0 over the whole pH range. In M:1f4z, the protonation probability of Glu194 is close to 0 over the whole pH range (*cf.* Figure 5.16 c). The protonation probability of Glu204 is close to 1 in the pH range from 0 to 11. Then, Glu204 deprotonates and at

pH = 14, the protonation probability is 0.25. In M:1kg8, Glu194 is fully protonated at pH < 3,. In the pH range from 3 to 7.5, the protonation probability decreases slightly to 0.95. Thereafter, Glu194 quickly deprotonates. In the pH range from 11.5 to 14, the probability is close to 0. Glu204 is fully deprotonated over the whole pH range. For M:1kg8, the individual probability curves of Glu194 and Glu204 are not shown. The sum of their protonation probabilities is depicted in Figure 5.15 d.

In M:1f4z at pH < 11, Glu194 and Glu204 behave contrary to each other, the first being deprotonated and the latter protonated. However, both show some variance in their behavior, *i.e.*, protonation states exist, where Glu194 is protonated and Glu204 deprotonated. The perfect negative correlation indicates that the protonation behavior of Glu194 and Glu204 is antithetic in all states. At pH > 11, when Glu204 deprotonates, the negative correlation is less pronounced. In M:1kg8, the behavior of Glu194 and Glu204 is opposed to each other in the pH range from 0 to 7.5. Nevertheless, the correlation coefficient is 0 over the whole pH range due to the missing variance in Glu194 or Glu204. Although the difference between the two M state structures is small with respect to the behavior of Glu194 and Glu204, the correlation behavior differs significantly.

N STRUCTURE

For N:1p8u, the correlation of Glu194 and Glu204 is -0.75 at pH=0. Thereafter, the correlation coefficient decreases. In the pH range from 2.5 to 7, the correlation is close to -1. At pH > 7, the correlation coefficient increases to 0. The protonation probability of Glu194 is close to 0 over the whole pH range. At pH < 8.5, the protonation probability of Glu204 is close to 1. In the pH range from 8.5 to 14, the protonation probability decreases to 0.25. For N:1p8u, the individual probability curves of Glu194 and Glu204 are not shown. The sum of their protonation probabilities is depicted in Figure 5.15 e.

The protonation behavior of Glu194 and Glu204 is antithetic in the pH range from 0 to 8.5. However, their correlation is perfectly negative only in the pH range from 2 to 7. In the pH range from 0 to 2 and 7 to 8.5, the two glutamate residues behave less (negatively) correlated than in the pH range from 2 to 7. At higher pH values, when Glu204 deprotonates and, thus, Glu194 and Glu204 have a similar protonation behavior Glu194, the two residues are, nevertheless, uncorrelated.

5.5.6 CORRELATION BETWEEN ASP85 AND GLU194/GLU204

Asp85 is located on the extracellular side of the retinal Schiff base in the center of BR. The residues of the proton release group, Glu194 and Glu204, are located adverse and close to each other on the extracellular side of BR. In the [bR] state, Asp85 is deprotonated and Glu194/Glu204 binds one proton. In the [M₁] state, Asp85 is protonated and Glu194/Glu204 still binds one proton. In the [M₂], [N] and [O] state Asp85 remains protonated, while Glu194/Glu204 are deprotonated. The photocycle and its proton transfer steps are depicted in Figure 5.2 and the intermediate protonation states are listed in Table 5.1.

The correlation between Asp85 and both Glu194 and Glu204 is depicted in Figure 5.23 for six BR structures. As can be seen, in the pH range from 3 to 12, the protonation

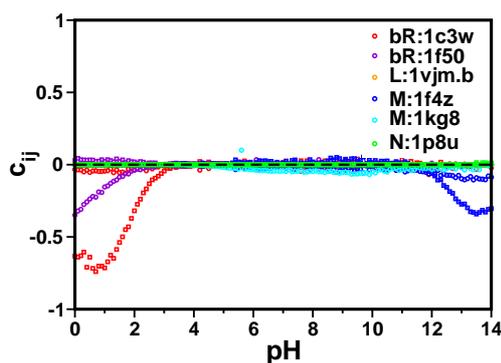


Figure 5.23. Correlation between Asp85 and Glu194/Glu204. The coefficient of the correlation between Asp85 and both Glu194 and Glu204 is plotted against pH for six BR structures.

behavior of Asp85 and Glu194/Glu204 is uncorrelated. At $\text{pH} < 3$, a negative correlation coefficient can be observed for Asp85 and Glu204 in bR:1c3w and for Asp85 and Glu194 in bR:1f4z. At $\text{pH} > 12$, a small negative correlation can be observed for Asp85 and both Glu194 and Glu204 in M:1f4z. In the following subsections, the correlation behavior is discussed in detail for the individual structures.

BR STRUCTURES

For bR:1c3w, the coefficient of the correlation between Asp85 and Glu194 fluctuates around 0 over the whole pH range. The correlation between Asp85 and Glu204 is -0.7 at $\text{pH} = 0$. Then, the coefficient increases to 0 at $\text{pH} = 3.5$ and fluctuates around 0 thereafter. For bR:1f50, the correlation between Asp85 and Glu194 is -0.35 at $\text{pH} = 0$. The coefficient increases to 0 at $\text{pH} = 2.5$. Thereafter, the correlation fluctuates around 0. The coefficient of the correlation between Asp85 and Glu204 fluctuates around 0 in the pH range from 0 to 12.5. Thereafter, the correlation coefficient is exactly 0 with one exception.

For both bR:1c3w and bR:1f4z, the protonation probability of Asp85 is slightly larger than 0.5 at $\text{pH} = 0$, but decreases rapidly with increasing pH as shown in Figure 5.12 a. At $\text{pH} > 3$, the protonation probability is close to 0 for both structures. In bR:1c3w, the protonation probability of Glu194 is 0.99 at $\text{pH} < 12.5$ (*cf.* Figure 5.16 a). Thereafter, it decreases slightly to 0.97. The protonation probability of Glu204 is 0.5 at $\text{pH} < 1$, decreases to 0 at $\text{pH} = 4$, and remains close to 0 up to $\text{pH} = 14$. In bR:1f50, the protonation probability of Glu194 is close to 0 over the whole pH range. The protonation probability of Glu204 is close to 1 in the pH range from 0 to 9.5. Then, the probability decreases to 0.15 at $\text{pH} = 14$. For bR:1f50, the individual probability curves of Glu194 and Glu204 are not shown. The sum of their protonation probabilities is depicted in Figure 5.15 a. In general, in bR:1c3w Glu194 is protonated and Glu204 is deprotonated and in bR:1f50 the protonation is reversed, while Asp85 is deprotonated in both structures.

In bR:1c3w, the protonation behavior of Asp85 and Glu194 is uncorrelated. The protonation of Asp85 and Glu204 is negatively correlated when both residues are protonated to some degree at low pH values. In bR:1f50, the protonation behavior of Asp85 and Glu194

is slightly negatively correlated when Asp85 is protonated in a number of states at low pH values. The protonation behavior of Asp85 and Glu204 is uncorrelated. Thus, Asp85 and either Glu194 or Glu204 are uncorrelated if they are deprotonated.

L STRUCTURE

For L:1vjm.b, the correlation between Asp85 and both Glu194 and Glu204 is 0 over the whole pH range. In the pH range from 0 to 11, Asp85 is fully protonated (*cf.* Figure 5.12 c). At $\text{pH} > 11$, its protonation probability is slightly smaller than 1 at 17 pH values. Glu194 is basically deprotonated in the pH range from 0 to 4. At $\text{pH} > 4$, the protonation probability is exactly 0. The protonation probability of Glu204 is close to 1 at $\text{pH}=0$. Then the protonation probability decreases to approximately 0.1 at $\text{pH}=4$ and approximately 0 at $\text{pH}=5$. In the pH range from 8.5 to 10, Glu204 is completely deprotonated. The individual probability curves of Glu194 and Glu204 are not shown. The sum of their protonation probabilities is depicted in Figure 5.15 c. Due to the absent variation of Asp85, Glu194 or Glu204, the correlation between Asp85 and both Glu194 and Glu204 is 0.

M STRUCTURES

For M:1f4z, the correlation between Asp85 and Glu194 is 0 at $\text{pH} \leq 3.2$ and fluctuates around 0 thereafter. In the pH range from 12 to 14 a small negative correlation of about -0.1 can be seen. Similarly, the correlation between Asp85 and Glu204 is 0 at $\text{pH} \leq 3.2$ and fluctuates around 0 thereafter. In the pH range of 12 to 14, the negative correlation is more pronounced than for Asp85 and Glu194 with a correlation coefficient of -0.3 at $\text{pH}=14$. For M:1kg8, the correlation between Asp85 and Glu194 is 0 at $\text{pH} < 4$. Thereafter a minimal negative correlation coefficient can be seen. The coefficient of the correlation between Asp85 and Glu204 is 0 over the whole pH range.

As shown in Figure 5.12 d, at $\text{pH} < 3.5$ the protonation probability of Asp85 is 1 for both M structures. At $\text{pH} > 3.5$, Asp85 deprotonates. In M:1kg8, the protonation probability remains above 0.8, while in M:1f4z, the protonation probability decreases to about 0.6 at $\text{pH}=14$. In M:1f4z, the protonation probability of Glu194 is close to 0 over the whole pH range (*cf.* Figure 5.16 c). The protonation probability of Glu204 is close to 1 in the pH range from 0 to 11. Then, Glu204 deprotonates and at $\text{pH}=14$, the protonation probability is 0.25. In M:1kg8, Glu194 is fully protonated at $\text{pH} \leq 3$. In the pH range from 3 to 7.5, the protonation probability decreases slightly to 0.95. Thereafter, Glu194 quickly deprotonates. In the pH range from 11.5 to 14, the probability is close to 0. Glu204 is fully deprotonated over the whole pH range. For M:1kg8, the individual probability curves of Glu194 and Glu204 are not shown. The sum of their protonation probabilities is depicted in Figure 5.15 d.

In M:1f4z, the protonation behavior of Asp85 and Glu194 is slightly negatively correlated when Asp85 is partly deprotonated at high pH values. Asp85 and Glu204 are negatively correlated, when both Asp85 and Glu204 are partly deprotonated at high pH values. In 1kg8, the protonation behavior of Asp85 and Glu194 is uncorrelated even when both show some variance. The correlation between Asp85 and Glu204 is 0 over the whole pH range due to the absent variation of Asp85 or Glu204.

N STRUCTURE

For N:1p8u, the correlation between Asp85 and Glu194 is 0 over the whole pH range with the exception of some values at $\text{pH} \geq 13.5$. Similarly, the correlation between Asp85 and Glu204 is 0 over the whole pH range with the exception of five values at $\text{pH} \geq 13.5$. The protonation probability of Asp85 is 1 in the pH range from 0 to 13.4, the (cf. Figure 5.12 e). At $\text{pH} \geq 13.5$ at five pH values, a probability of 0.99 is observed. The protonation probability of Glu194 is close to 0 over the whole pH range. In contrast, the protonation probability of Glu204 is close to 1 at $\text{pH} < 8.5$. In the pH range from 8.5 to 14, the protonation probability decreases to 0.25. For N:1p8u, the individual probability curves of Glu194 and Glu204 are not shown. The sum of their protonation probabilities is depicted in Figure 5.15 e. In general, the correlation between Asp85 and both Glu194 and Glu204 is 0 due to the absent variance of Asp85. However, at the five pH values where Asp85 does show a small variance the protonation behavior of Asp85 and both Glu194 and Glu204 is uncorrelated, too.

5.5.7 CORRELATION BETWEEN GLU194/GLU204 AND THE SCHIFF BASE

The retinal Schiff base is located in the center of BR. The residues of the proton release group, Glu194 and Glu204, are located adverse and close to each other on the extracellular side of BR. In the [bR] state, Glu194/Glu204 and the retinal Schiff base are protonated. In the [M₁] state, the Schiff base is deprotonated. In the [M₂] state, Glu194/Glu204 is deprotonated, too. In the [N] and [O] state, the Schiff base is protonated again, while Glu194/Glu204 remains deprotonated. The photocycle and its proton transfer steps are shown in Figure 5.2 and the intermediate protonation states are listed in Table 5.1.

The correlation between both Glu194 and Glu204 and the retinal Schiff base is depicted in Figure 5.24 for six BR structures. In general, the protonation behavior of Glu194/Glu204 and the retinal Schiff base is uncorrelated. However, interestingly a positive and also a negative correlation can be observed for Glu204 and the retinal Schiff base in L:1vjm.b. For M:1f4z and for N:1p8u, a small positive correlation can be seen for Glu194 or Glu204 and the retinal Schiff base at high pH values. For M:1kg8, the correlation between Glu194 and the retinal Schiff base is negative. In the following subsections, the correlation behavior is discussed in detail for the individual structures.

BR STRUCTURES

For bR:1c3w, the correlation coefficient of Glu194 or Glu204 and the retinal Schiff base fluctuates around 0 over the whole pH range. Similarly for bR:1f50, the correlation coefficient of Glu194 or Glu204 and the retinal Schiff base fluctuates around 0 over the whole pH range. The protonation probability of the retinal Schiff base is close to 1 over the complete pH range from 0 to 14 for both bR:1c3w and bR:1f50 (cf. Figure 5.11 a). In bR:1c3w, the protonation probability of Glu194 is 0.99 at $\text{pH} < 12.5$. Thereafter, it decreases slightly to 0.97 (cf. Figure 5.16 a). The protonation probability of Glu204 is 0.5 for the pH range of 0 to 1, decreases to 0 at $\text{pH}=4$, and remains close to 0 up to $\text{pH}=14$. In bR:1f50, the protonation probability of Glu194 is close to 0 over the whole

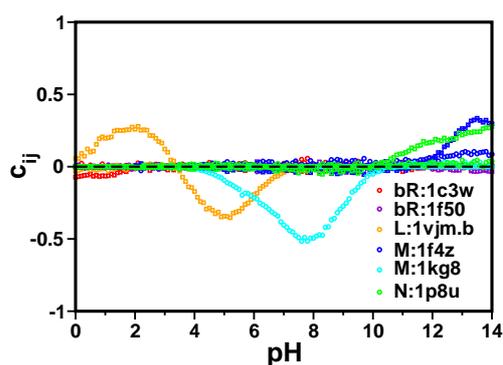


Figure 5.24. Correlation between Glu194/Glu204 and the Schiff base. The coefficient of the correlation between both E194 and E204 and the Schiff base is plotted against pH for six BR structures.

pH range. The protonation probability of Glu204 is close to 1 in the pH range from 0 to 9.5. Then, the probability decreases to 0.15 at pH = 14. For bR:1f50, the individual probability curves of Glu194 and Glu204 are not shown. The sum of their protonation probabilities is depicted in Figure 5.15 a. In short in bR:1c3w, Glu194 is protonated and Glu204 is deprotonated and in bR:1f50 the protonation is reversed. For bR:1c3w and bR:1f50, Glu194 or Glu204 and the retinal Schiff base are uncorrelated over the whole pH range.

L STRUCTURE

For L:1vjm.b, the correlation coefficient of Glu194 and the retinal Schiff base fluctuates around 0 at pH < 4 and is exactly 0 thereafter. The correlation between Glu204 and the retinal Schiff base is 0 at pH = 0. It increases up to 0.25 at pH = 2 and then decreases to -0.35 at pH = 5.2 with a zero point at pH = 3.4. At pH = 7.5, the coefficient again approximates 0 and at pH > 8, it is exactly 0. As shown in Figure 5.11 c, at pH < 2 the protonation probability of the retinal Schiff base is close to 1. In the pH range from 2 to 8, the Schiff base deprotonates. At pH > 8, the protonation probability is close to 0 or exactly 0. Glu194 is nearly deprotonated in the pH range from 0 to 4. At pH > 4, the protonation probability is exactly 0. The protonation probability of Glu204 is close to 1 at pH = 0. Then it decreases to approximately 0.1 at pH = 4 and approximately 0 at pH = 5. In the pH range from 8.5 to 14, Glu204 is completely deprotonated. The individual probability curves of Glu194 and Glu204 are not shown. The sum of their protonation probabilities is depicted in Figure 5.15 c.

In the pH range from 4 to 14, Glu194 and in part the retinal Schiff base show no variance. Their correlation coefficient is, therefore, 0. However, Glu194 and the retinal Schiff base are uncorrelated, too, when they show variance. In contrast, Glu204 and the Schiff base are slightly positively correlated in the pH range from 0 to 3.5, when both are mostly protonated. Then in the pH range from 3.5 to 7.5, when Glu204 is mostly deprotonated and the Schiff base mostly protonated, they are slightly negatively correlated. At pH > 8, the correlation coefficient is 0 due to the absent variance of Glu204.

M STRUCTURES

For M:1f4z, the correlation coefficient of Glu194 and the retinal Schiff base is approximately 0 at $\text{pH} < 12$. At high pH values the correlation increases slightly to 0.1 at $\text{pH} = 14$. Similarly, the correlation coefficient of Glu204 and the retinal Schiff base is approximately 0 at $\text{pH} < 12$. In the pH range from 12 to 14 the correlation increases to about 0.3.

For M:1kg8, the correlation between Glu194 and the retinal Schiff base is exactly 0 at $\text{pH} \leq 3$ and approximately 0 up to $\text{pH} 4.5$. Furthermore at $\text{pH} > 10$, the coefficient fluctuates around 0. However, in the pH range from 4.5 to 10, a negative correlation can be observed with a minimum of -0.5 at $\text{pH} = 7.7$. In contrast, the correlation between Glu204 and the retinal Schiff base is 0 over the whole pH range.

As shown in Figure 5.11 d, at $\text{pH} < 4$ the protonation probability of the retinal Schiff base is close to 1. In the pH range from 4 to 8, the Schiff base deprotonates. In M:1f4z, the minimum protonation probability is approximately 0.15. At $\text{pH} = 10$, the protonation probability increases up to approximately 0.5 at $\text{pH} = 14$. In M:1kg8, the protonation probability is close to 0 in the pH range from 8 to 14. In M:1f4z, the protonation probability of Glu194 is close to 0 over the whole pH range (*cf.* Figure 5.16 c). The protonation probability of Glu204 is close to 1 in the pH range from 0 to 11. Then, Glu204 deprotonates and at $\text{pH} = 14$, the protonation probability is 0.25. In M:1kg8, Glu194 is fully protonated at $\text{pH} \leq 3$. In the pH range from 3 to 7.5, the protonation probability decreases slightly to 0.95. Thereafter, Glu194 quickly deprotonates. In the pH range from 11.5 to 14, the probability is close to 0. Glu204 is fully deprotonated over the whole pH range. For M:1kg8, the individual probability curves of Glu194 and Glu204 are not shown. The sum of their protonation probabilities is depicted in Figure 5.15 d.

For M:1f4z, the behavior of Glu194 or Glu204 and the retinal Schiff base is slightly correlated at high pH values when both glutamate residues are mostly deprotonated and the Schiff base is partly protonated. The correlation between Glu204 and the Schiff base is more pronounced than between Glu194 and the Schiff base. For both Glu204 and the Schiff base the protonation probability is significantly higher in this pH range than for Glu194. Thus, the positive correlation may result from both residues being protonated at the same time.

For M:1kg8, Glu194 and the retinal Schiff base are negatively correlated when the first is still mostly protonated and the latter mostly deprotonated. At higher pH values, when both Glu194 and the Schiff base are virtually deprotonated their behavior is uncorrelated. The correlation coefficient of Glu204 and the Schiff base is 0 over the whole pH range due to the absent variance of Glu204.

N STRUCTURE

For N:1p8u, the correlation coefficient of Glu194 and the retinal Schiff base is 0 at $\text{pH} \leq 1.3$ and fluctuates around 0 thereafter. Similarly, the correlation coefficient of Glu204 and the retinal Schiff base is 0 at $\text{pH} \leq 1.3$ and fluctuates around 0 thereafter. However, at $\text{pH} > 10$, the correlation coefficient increases to about 0.25 at $\text{pH} = 14$.

As shown in Figure 5.11 e, at $\text{pH} < 3$ the retinal Schiff base is protonated. In the pH range from 3 to 8, the Schiff base deprotonates and at $\text{pH} = 8$, the protonation probability is close to 0. In the pH range from 8 to 14, the probability increases slightly to approximately 0.2. The protonation probability of Glu194 is close to 0 over the whole pH range. At $\text{pH} < 8.5$, the protonation probability of Glu204 is close to 1. In the pH range from 8.5 to 14, the protonation probability decreases to 0.25. For N:1p8u, the individual probability curves of Glu194 and Glu204 are not shown. The sum of their protonation probabilities is depicted in Figure 5.15 e.

For N:1p8u, the protonation behavior of Glu194 and the Schiff base is uncorrelated over the whole pH range. In contrast, Glu204 and the Schiff base which are uncorrelated at $\text{pH} < 10$ are slightly positively correlated at high pH values. At high pH values, the protonation probability of the Schiff base which is deprotonated at $\text{pH} = 8$ increases slightly. Similarly, Glu204 has a low probability to be protonated, whereas Glu194 is deprotonated. Thus, the positive correlation may result from the concurrent protonation of Glu204 and the Schiff base.

5.6 IMPLICATIONS FOR THE INTERMEDIATE STRUCTURES

The probabilities of functionally relevant protonation state subsets are analyzed for the different intermediate structures of bacteriorhodopsin. It could be shown, that the [bR] and [M₁] state are well described by the available high-resolution X-ray structures. However, the [M₂], the [N] and the [O] protonation state are less well represented by the structural data. While some intermediate structures have a high probability for the [M₂] state, mainly at high pH values, the L structure L:1vjm.b is the only structure that has a probability for the [M₂] state at physiological pH values. The [N] state has a minimal probability only in the N structure. The [O] state has a significant probability not in the O-like structure, but again in the L structure L:1vjm.b and a small probability in the M structure M:1kg8.

Additionally, the protonation behavior of the key residues is analyzed for the different intermediate structures. Again, the behavior expected for the early photocycle intermediates is, in general, in excellent agreement with the experimental data, while the protonation behavior expected for the late photocycle intermediates is not. For example, Asp96 is virtually protonated in all structures at physiological pH values. Although this aspartate is the proton donor of the retinal Schiff base and should be deprotonated in the N structure, only a slight deprotonation at high pH values can be observed for N:1p8u. Furthermore, the deprotonation of the proton release group Glu194/Glu204 at physiological pH can only be observed in two structures, the L state structures L:1e0p and L:1vjm.b, although the proton release group is deprotonated in the [M₂], [N] and [O] state. The respective intermediate structures, however, show a deprotonation of the proton release group only at high pH values. This is the reason, that the [M₂], [N] and [O] state are not well represented by the structural data. For future calculations, it may be of interest to use the region of Glu194 and Glu204 in L:1vjm.b as reference.

The correlation between the protonation behavior offers meaningful insight into the interaction between the key residues of proton transfer. For example, the protonation behavior of the retinal Schiff base and its proton acceptor Asp85 is strongly negatively correlated in several structures. Interestingly, a positive correlation can be observed for the protonation of Asp85 and Asp115 in M:1kg8. Furthermore, the residues of the proton release group, Glu194 and Glu204, are perfectly negatively correlated in several structures.

However, even though the most significant and interesting correlations are discussed here, often the correlation coefficient is close to 0 or the protonation is correlated only at low or high pH values. In the data not shown here, which is the bulk of the pair-correlations in BR, this behavior is prevalent. This correlation behavior reflects the protonation behavior of BR. The groups are often stably protonated or deprotonated with a probability close to 1 or 0 over a large pH range. Protonation changes are mainly seen at low pH values, when acids deprotonate, and at high pH values, when bases deprotonate. Consequently, the inter-relation between two protonatable residues may be too complex to be described by a single coefficient. Instead, it may be of interest to analyze their concerted behavior in more detail.

The calculations presented in this chapter were performed using the Metropolis Monte Carlo approach. They provided not only insight into the behavior of the key residues of proton transfer, but disclosed certain limitations of this method. In particular, the random sampling of low energy states that results in the successful description of thermodynamic properties cannot deterministically search all states accessible to a system. In the case of bacteriorhodopsin, the protonation states of lowest energy are of major interest. However, no method that allowed an unambiguous determination of these states existed. This offered an incentive to develop such a method. In the following chapter, an algorithm is introduced that deterministically builds list of the lowest energy states.

CHAPTER 6

EXTENDED DEAD-END ELIMINATION: AN ALGORITHM TO DETERMINE ALL STATES OF LOWEST ENERGY

'Space,' it says, 'is big. Really big. You just won't believe how vastly hugely mindbogglingly big it is.'

Douglas Adams, The Hitchhiker's Guide to the Galaxy

Proteins are the most versatile among the macromolecules of a living organism. They are involved in the catalysis of most metabolic processes, in the recognition of molecules (for example hormones and antigens), in signal transduction and in the transport of many substances. As a result of the multiple tasks they perform, proteins are central to the functioning of the cellular machine. Modified, mutated or missing proteins are the underlying cause for many diseases. The understanding of protein function on the atomic level is, therefore, of fundamental biological and medical interest. Ultimately, protein function depends on its three-dimensional structure. Due to its complexity, the investigation of the structure-function relationship of proteins remains one of the great challenges in structural biology.

The protein structure depends on the physical properties of the structural constituents and is influenced by environmental factors such as temperature and the pH value. Different experimental methods have been successful in providing structural information. Structures in atomic detail can, for example, be solved by NMR spectroscopy and X-ray scattering as described in more detail in Chapter 3. Experimental methods, however, are time consuming and limited with respect to the size and the properties of the molecule. Computational methods can supplement the experimental methods where these reach their limitations. Additionally, the comparison with a computationally derived structure can provide valuable information, if an experimental structure is available.

In accordance with the laws of thermodynamics, the protein structure of lowest free energy is the most probable. The determination of this structure is, therefore, the primary goal in computational structure prediction. As a first step, the conformational space of proteins is commonly discretized. The discretized space is finite, resulting in a combina-

torial problem. In principle, this strategy allows an exhaustive search for the structure of lowest energy. In an exhaustive search, all conformations or, more generally, all states of the protein are calculated and their energies compared to determine the structure or state of lowest energy. Proteins, however, are complex systems with many degrees of freedom. The direct calculation of the lowest energy structure is – due to the size of the combinatorial space – rarely possible.

Considering, for example, a small protein with 100 residues where each of these residues can adopt only 5 different rotamer positions. Then, the conformational space is of a size of $5^{100} \approx 10^{70}$. For a protein with 500 residues, the number of different conformations is close to 10^{350} . An increase in the number of residues leads to an exponential increase in the number of conformations. This phenomenon is termed combinatorial explosion.

Algorithms that deterministically find the state of lowest energy while circumventing an exhaustive search of the complete space of all states of the system, *i.e.*, the *state space*, are, therefore, of great interest. One solution to the problem of finding the state of lowest energy for complex protein systems, *i.e.*, dead-end elimination, is described in the next section. In the main part of this chapter, a novel algorithm is introduced that determines not only the state of lowest energy, but a complete list of all lowest energy states. This algorithm is termed extended dead-end elimination (X-DEE).

Algorithms that find the state of lowest energy are of general interest not only for protein structure prediction. Of major importance for the function of proteins is for example also their protonation and redox state. In the following sections, these algorithms are described in general terms. The goal is not only to specifically search the conformational space to determine the lowest energy structure of a protein. Instead, a defined state space is searched for a gap-free list of all lowest energy states.

6.1 DEAD-END ELIMINATION: AN ALGORITHM TO DETERMINE THE STATE OF LOWEST ENERGY

In 1992, Desmet *et al.* [209] introduced an algorithm termed *dead-end elimination* (DEE) that solves the combinatorial problem of identifying the global energy minimum for certain discrete systems. DEE has been successfully applied in protein structure prediction [209–215], protein design [216–218], sequence alignment [219], and also in the evaluation of protein solvent exposure [220]. Originally, DEE was applied in homology modeling to predict protein side chain positions. In this context, the protein is described as a set of residues, or *sites*, each of which may adopt a finite number of possible rotamer positions, or *forms*. An example is depicted in Figure 6.1.

For the DEE algorithm to be applicable, the system of interest has to be discretized. More precisely, the state energy has to be expressible in terms of intrinsic energies of individual sites and pairwise interactions between these sites. The sites can adopt discrete forms that differ in their contribution to the state energy. Each state of the system can be described by a vector \vec{x} that specifies the form x of each site μ in this state: $\vec{x} = (x_1, \dots, x_\mu, \dots, x_N)$, where N is the number of sites. A general formulation of the state

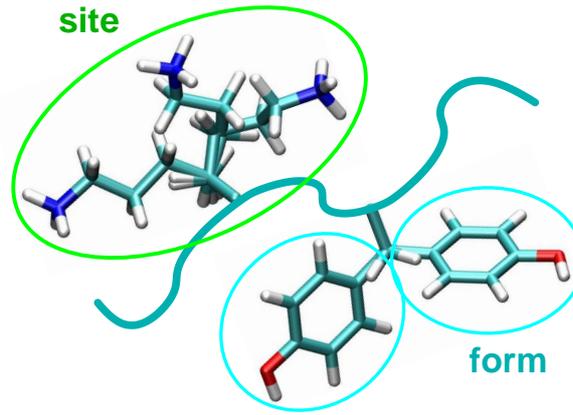


Figure 6.1. Definition of a protein site and a rotamer form. Each protein residue constitutes a site as indicated by the green ellipse for a lysine residue. Each site can adopt different forms, *i.e.*, side chain rotamer positions in the case of protein residues. Three different rotamer positions are shown for a lysine and two for a tyrosine residue. The two forms are indicated by the cyan ellipses for the tyrosine residue. The protein backbone is indicated by the irregular line.

energy is given by:

$$E(x) = \sum_{\mu=1}^N E_{\text{intr}}(x_{\mu}) + \frac{1}{2} \sum_{\mu=1}^N \sum_{\nu=1}^N I(x_{\mu}, x_{\nu}) \quad , \quad (6.1)$$

where $E_{\text{intr}}(x_{\mu})$ is the intrinsic energy of site μ in form x and $I(x_{\mu}, x_{\nu})$ is the interaction energy between the two sites μ and ν in their respective forms. The self interaction energy $I(x_{\mu}, x_{\mu})$ is assumed to be zero.

Assuming a constant number of forms P for each site, the number of possible states M equals P^N . As mentioned in the introduction to this chapter, an exhaustive search of a state space with cardinality P^N is impracticable even for small systems. DEE reduces the state space, thereby preventing an exhaustive search of the full space. The original idea of DEE is the identification of certain forms of sites that cannot contribute to the state of lowest energy. These forms, so-called *dead-ends*, are identified without explicitly calculating state energies. Dead-ends are excluded from further consideration. With each eliminated dead-end the size of the state space, in which the state of lowest energy can be found, decreases. Dead-ends are identified by the DEE criterion that compares the energy of two forms of one site. Successively, the DEE criterion is applied to all forms of all sites. The next subsection introduces the DEE criterion in more detail.

6.1.1 THE DEE CRITERION

The original DEE criterion formulated by Desmet *et al.* [209] defines form d of site α as a dead-end, if the state of lowest energy with d_{α} is higher than the state of highest energy that contains c_{α} . Hence, d_{α} is a dead-end if all states with d_{α} have a higher energy than all states that contain c_{α} , where c_{α} is another form c of site α . Later, Goldstein formulated a more powerful criterion that identifies d_{α} as dead-end if the state energy can always be

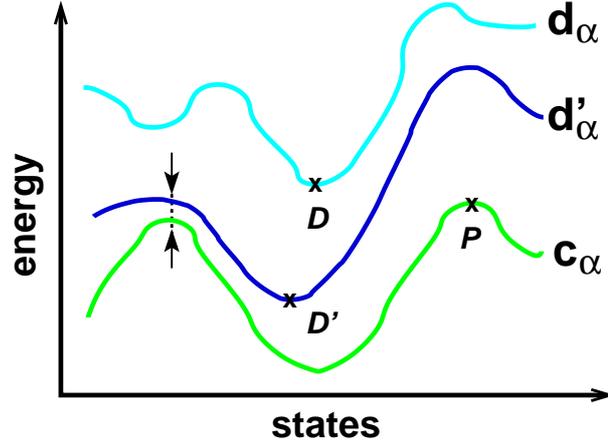


Figure 6.2. The DEE criterion. To identify dead-ends, the DEE criterion compares the contribution of two forms, *i.e.*, c and d , of one site α to the state energy. Desmet *et al.* [209] defined d_α as dead-end, if the state of lowest energy that contains d_α (indicated by D) has a higher energy than the state of highest energy that contains c_α (indicated by P). The Goldstein criterion identifies d_α as dead-end, if the state energy can always be lowered by changing the form of site α from d to c [210]. Thus, the Goldstein criterion additionally identifies d'_α as dead-end, although the original Desmet criterion is not fulfilled, since D' is lower in energy than P . Instead of calculating the curves depicted in the graph, the DEE criterion of Goldstein considers the minimum energy difference between the states that contain c_α and d_α , indicated by the arrows. The criterion is applied successively to all forms of all sites.

lowered by changing the form of site α from d to c [210]. Hence, d_α is a dead-end if all states that contain d_α have a higher energy than the corresponding states that contain c_α . A graphic representation of the two schemes is shown in Figure 6.2. In this work, the DEE criterion according to Goldstein is used.

Mathematically, the Goldstein DEE criterion is an inequality that compares the contribution of two forms of one site to the state energy. If this inequality is true, one form of this site, *i.e.*, the dead-end, can be eliminated from further consideration. To derive the DEE criterion, as a first step, the state energy given by Eq. (6.1) is defined for the specific cases that site α takes either the form c or d :

$$E_{c_\alpha}(x) = E_{\text{intr}}(c_\alpha) + \sum_{\mu \neq \alpha} E_{\text{intr}}(x_\mu) + \sum_{\mu \neq \alpha} I(c_\alpha, x_\mu) + \frac{1}{2} \sum_{\mu \neq \alpha} \sum_{\nu \neq \alpha} I(x_\mu, x_\nu) \quad (6.2)$$

$$E_{d_\alpha}(x) = E_{\text{intr}}(d_\alpha) + \sum_{\mu \neq \alpha} E_{\text{intr}}(x_\mu) + \sum_{\mu \neq \alpha} I(d_\alpha, x_\mu) + \frac{1}{2} \sum_{\mu \neq \alpha} \sum_{\nu \neq \alpha} I(x_\mu, x_\nu) \quad (6.3)$$

where x_μ and x_ν denote the form of all sites $\mu, \nu \neq \alpha$. Then, $E_{c_\alpha}(x)$ is subtracted from $E_{d_\alpha}(x)$, giving:

$$E_{d_\alpha}(x) - E_{c_\alpha}(x) = E_{\text{intr}}(d_\alpha) - E_{\text{intr}}(c_\alpha) + \sum_{\mu \neq \alpha} [I(d_\alpha, x_\mu) - I(c_\alpha, x_\mu)] \quad (6.4)$$

where x_μ denotes the form of all sites $\mu \neq \alpha$. The first two terms represent the difference in intrinsic energies, $E_{\text{intr}}(c_\alpha)$ and $E_{\text{intr}}(d_\alpha)$. The third term accounts for the difference in

interaction energy between the two forms c and d of site α and the forms x of all other sites μ .

According to the Goldstein DEE criterion, d_α is a dead-end, if the state energy can always be lowered by changing the form of site α from d to c . It is sufficient to prove that the minimum energy difference between the states that contain c_α and d_α is larger than zero as indicated in Figure 6.2 by the arrows. As Eq. (6.4) holds for all forms x , it follows that d_α is a dead end if there exists a c_α such that:

$$E_{\text{intr}}(d_\alpha) - E_{\text{intr}}(c_\alpha) + \sum_{\mu \neq \alpha} \min_x [I(d_\alpha, x_\mu) - I(c_\alpha, x_\mu)] > 0 . \quad (6.5)$$

As in Eq. (6.4), the first two terms account for the difference in intrinsic energies, $E_{\text{intr}}(c_\alpha)$ and $E_{\text{intr}}(d_\alpha)$. The minimum energy difference between the interaction of c_α with the other sites μ and d_α with the other sites μ is determined by searching over all possible forms x_μ . Ineq. (6.5) yields a lower bound of the change in energy upon changing the form of site α from d to c . If Ineq. (6.5) is true, *i.e.*, if this lower bound is larger than zero, the state energy can always be lowered by substituting d_α with c_α . Thus, d_α cannot appear in the state of lowest energy. The form d of site α is a dead-end that can be eliminated from further consideration.

Successively, the DEE criterion is applied to all forms of all sites. The computational cost to evaluate the DEE criterion is of the order of $N^2 P^3$, where N denotes the number of sites and P the number of forms per site [212]. This scaling behavior poses a significant advantage compared to the $\mathcal{O}(P^N)$ scaling behavior of a full exhaustive search of the complete state space.

In general, the state of lowest energy is not directly determined by DEE. Rather, if enough dead-ends can be excluded from the search, the size of the state space is reduced to a point, where a subsequent exhaustive search becomes feasible. Goldstein has shown that the DEE theorem does not work for random spin glasses [210]. The DEE algorithm has, however, been successfully applied to various biophysical systems as mentioned above. One condition for the effective and sufficient reduction of the state space seems to be the non-randomness of the system. This is the case in biological systems like protein structures and gene sequences.

6.1.2 OBJECTIVE TO OBTAIN A LIST OF THE LOWEST ENERGY STATES

For many systems not only the state of lowest energy, but a set of the lowest energy states is relevant. Proteins, for example, are dynamic systems that may adopt several functionally important states. To understand their mechanisms it is necessary to obtain a complete picture of the states that are accessible to the protein. Consequently, algorithms that are able to produce gap-free lists of low energy states, *i.e.*, lists of states that are complete up to a given energy distance from the global energy minimum, are of general interest in structural biology.

While Monte Carlo techniques allow to sample low energy states and generally provide an accurate description of thermal properties [126], they do not allow to obtain gap-free lists of the lowest energy states for a given energy range. This is a major drawback for

the investigation of the kinetics of a working enzyme. States with a low probability may well be functionally relevant [221, 222] and an incomplete sample of states available to the system may lead to wrong conclusions about the mechanism.

Based on DEE, a computational scheme is developed that generates a gap-free list of the lowest energy states. Furthermore, this new algorithm allows to complete any given list of low energy states by identifying those low energy states that are missing in the list. The remaining part of this chapter focuses on the description of this algorithm, termed X-DEE for extended DEE, and an analysis of its properties and its computational performance.

6.2 X-DEE: AN ALGORITHM TO DETERMINE A GAP-FREE LIST OF LOWEST ENERGY STATES

The basic idea of the X-DEE algorithm is to exclude a list of states from the search space explored by DEE. If a gap-free list of k low energy states $\{\vec{x}_1, \dots, \vec{x}_k\}$ is already known, the $(k+1)^{\text{th}}$ state can be found by restricting the search for the lowest energy state to the set of all states \mathcal{M} excluding the set of already known states, *i.e.*, by searching the state of lowest energy in $\mathcal{M} \setminus \{\vec{x}_1, \dots, \vec{x}_k\}$. More generally, X-DEE can be used to restrict the search to a set $\mathcal{M} \setminus \mathcal{L}$ for any given list \mathcal{L} of states. In case \mathcal{L} is not gap-free, X-DEE will identify the state of lowest energy not included in \mathcal{L} and can consequently complete \mathcal{L} , until a gap-free list of low energy states is obtained.

The DEE algorithm gains its computational efficiency from excluding sets of states from further consideration without explicitly calculating their energies. That is, the identification of d_α as dead-end allows to exclude a subset of states from further consideration. Specifically, all states that contain d_α are excluded while none of the state energies have to be calculated explicitly.

There is, however, no straightforward way to exclude an arbitrary list of states \mathcal{L} from the search space explored by DEE. Nevertheless, a DEE search can be restricted to a specific type of subset of \mathcal{M} . For example, let \mathcal{S} be a subset of \mathcal{M} containing all states, where site 1 has the form h : $\mathcal{S} = \{\vec{x} \mid x_1 = h\} \subseteq \mathcal{M}$. Applying DEE to this subset \mathcal{S} yields the state of lowest energy where site 1 has the form h .

The subset \mathcal{S} can be represented by a so-called *search key*: $\vec{b}_\mathcal{S} = (h_1, *, \dots, *, \mu, \dots, *, N)$, where h is the specified form of site 1. The asterisk $*$ indicates that this site is undefined and will be determined during the DEE search. For each site μ of the system, these search keys $\vec{b}_\mathcal{S}$ have a component b_μ which is either fixed to a specific form or undefined. Analogously, search keys where more than one site is fixed to a specific form represent the subsets that coincide in the form of these fixed sites. Keeping a number of sites fixed during a DEE search yields the state of lowest energy of a subset \mathcal{S} of all state \mathcal{M} characterized by the forms of those sites that are fixed.

The principal idea is to define search keys $\vec{b}_\mathcal{S} = (b_1, \dots, b_\mu, \dots, b_N)$ such that the subsets \mathcal{S} represented by the individual search keys together represent $\mathcal{M} \setminus \mathcal{L}$. Determining the state of lowest energy of all subsets via the DEE algorithm then yields the desired state of lowest energy of $\mathcal{M} \setminus \mathcal{L}$.

\mathcal{L} may constitute any list of states of the system. This list may for example contain states from a previous Monte Carlo calculation. By searching the lowest energy state in $\mathcal{M} \setminus \mathcal{L}$, then adding the newly determined state \vec{x}_{new} to the list and determining the lowest energy state of $\mathcal{M} \setminus (\mathcal{L} \cup \{\vec{x}_{\text{new}}\})$, the gaps in a given list of states are consecutively eliminated by X-DEE. One can also start by determining the state of lowest energy of \mathcal{M} using DEE. By adding this state to \mathcal{L} , which then contains one state, the state of lowest energy in $\mathcal{M} \setminus \mathcal{L}$ can be determined by X-DEE. By successively adding the newly determined state to \mathcal{L} , the list is consecutively filled with the states of lowest energy.

The main challenge of X-DEE lies in creating a search basis \mathcal{B} , *i.e.*, a set of search keys \vec{b}_S , such that \mathcal{L} is excluded from the search and the complete set $\mathcal{M} \setminus \mathcal{L}$ is searched. In the following subsection, a procedure to generate \mathcal{B} is presented that fulfills both conditions.

6.2.1 CONSTRUCTION OF A SEARCH BASIS \mathcal{B} FOR $\mathcal{M} \setminus \mathcal{L}$

A schematic representation of the procedure CreateSearchBasis is shown in Figure 6.3. Given the list of states \mathcal{L} to be excluded from the search, a search basis \mathcal{B} is constructed as follows:

Input of the recursive procedure CreateSearchBasis is a list of states \mathcal{L} and a list vector \vec{t} that is associated to \mathcal{L} . In the first step, \mathcal{L} corresponds to the complete list \mathcal{L} . The list vector \vec{t} contains an element for each site μ : $\vec{t} = (t_1, \dots, t_\mu, \dots, t_N)$. This list vector keeps track of the sites that are already fixed to specific forms. In the first step, all sites in \vec{t} are undefined, $\vec{t} = (*_1, \dots, *_\mu, \dots, *_N)$.

The algorithm proceeds in five steps:

1. It is verified that the list vector \vec{t} associated to \mathcal{L} contains at least one undefined site. Otherwise, the procedure stops.
2. Among the undefined sites, the algorithm searches for a site μ with forms that do not occur in any of the state vectors in \mathcal{L} , *i.e.*, a site μ with unused forms h . Once such a site μ is found, the algorithm proceeds to step 3. If no such site exists, the first undefined site in \vec{t} is chosen as site μ and the algorithm proceeds directly to step 4.
3. For each unused form h of site μ , a search key \vec{b} is defined by copying the list vector \vec{t} to \vec{b} and fixing site μ to form h , *i.e.*, $\vec{b} \mid b_\mu = h$. Each search key \vec{b} , thus, differs from the current list vector only at site μ . Fixing site μ to forms h not occurring in \mathcal{L} , guarantees that the subset represented by \vec{b} and \mathcal{L} are disjoint, *i.e.*, \vec{b} represents a subset of $\mathcal{M} \setminus \mathcal{L}$. The search keys \vec{b} are added to the search basis \mathcal{B} .
4. The vectors in \mathcal{L} are divided into sublists $\mathcal{L}_{\text{sub},i}$ such that site μ has form g_i in all state vectors \vec{x} in $\mathcal{L}_{\text{sub},i}$, $\mathcal{L}_{\text{sub},i} = \{\vec{x} \mid x_\mu = g_i\}$. To each sublist $\mathcal{L}_{\text{sub},i}$, a separate list vector $\vec{t}_{\text{sub},i}$ is assigned by copying list vector \vec{t} to $\vec{t}_{\text{sub},i}$ and fixing site μ to the form g_i common to all state vectors in $\mathcal{L}_{\text{sub},i}$, $\vec{t}_{\text{sub},i} \mid t_\mu = g_i$.
5. For each sublist $\mathcal{L}_{\text{sub},i}$ and its list vector $\vec{t}_{\text{sub},i}$, the steps 1 to 5 are executed again.

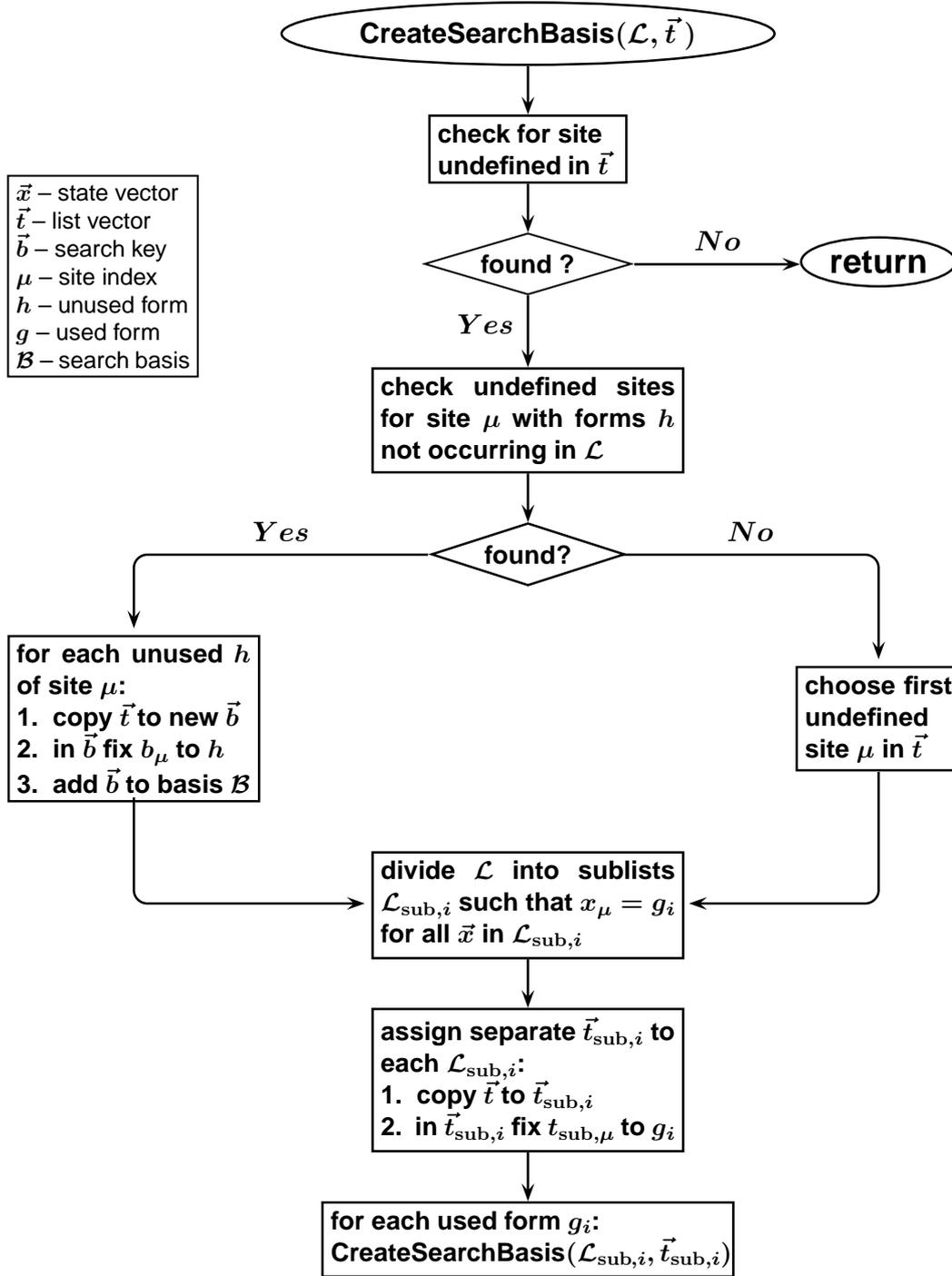


Figure 6.3. Flowchart of the recursive procedure CreateSearchBasis. Input of CreateSearchBasis is a list of states \mathcal{L} and an associated list vector \vec{t} . With each recursion, \mathcal{L} is divided into sublists $\mathcal{L}_{\text{sub},i}$ and one additional site is fixed in the associated list vectors $\vec{t}_{\text{sub},i}$. CreateSearchBasis terminates when all sites of a list vector $\vec{t}_{\text{sub},i}$ are fixed. With each recursion, search keys \vec{b}_S can be generated that differ from the list vector $\vec{t}_{\text{sub},i}$ in the form of site μ . The search keys are added to the search basis \mathcal{B} . In total, CreateSearchBasis generates a set of search keys \vec{b}_S characterizing subsets S whose union represent $\mathcal{M} \setminus \mathcal{L}$.

The recursive procedure `CreateSearchBasis` starts with a list \mathcal{L} and an associated list vector \vec{t} and terminates with a set of sublists $\mathcal{L}_{\text{sub},i}$ with associated list vectors $\vec{t}_{\text{sub},i}$. If step 3 is executed, new search keys are added to the search basis \mathcal{B} .

The search space of `CreateSearchBasis` is finite. Thus, the procedure eventually terminates. Specifically, `CreateSearchBasis` terminates for a given sublist, if all sites are fixed in the associated list vector (*cf.* step 1). An example of creating a search basis is given in Figure 6.4.

All search keys in \mathcal{B} are subjected to a DEE search yielding the states of lowest energies of the represented subsets \mathcal{S} . These states include the state of lowest energy of $\mathcal{M} \setminus \mathcal{L}$.

6.2.2 COMPLETENESS OF THE SEARCH BASIS \mathcal{B}

For the search basis \mathcal{B} , it remains to be shown that all subsets of states \mathcal{S} represented by the search keys are subsets of $\mathcal{M} \setminus \mathcal{L}$ and that the union of all subsets \mathcal{S} represent the complete set $\mathcal{M} \setminus \mathcal{L}$.

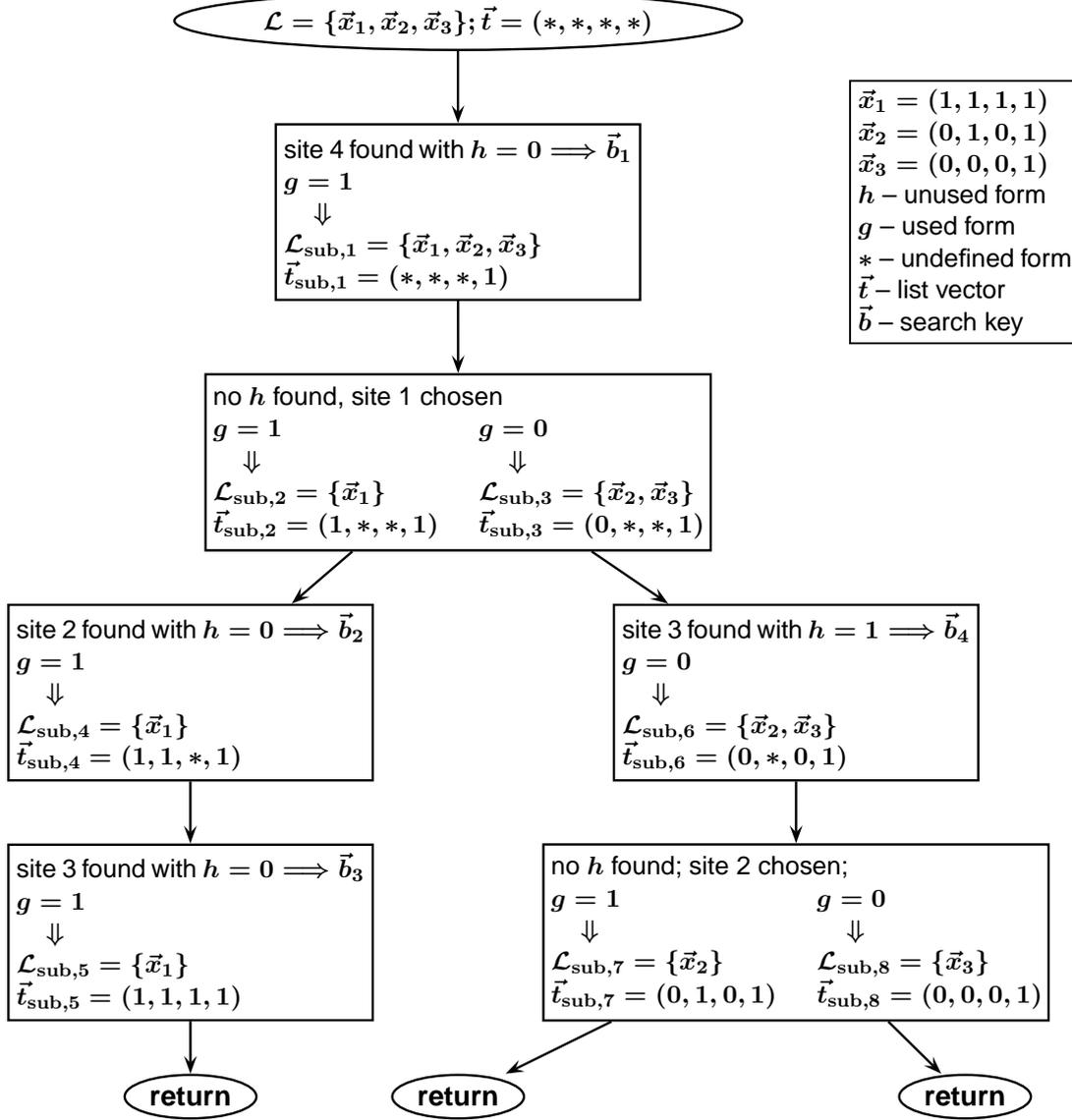
1. Each search key is derived from a list vector \vec{t} which is associated with a list \mathcal{L} . By construction, a search key derived from \vec{t} at most searches the subset \mathcal{L} . However, each search key differs at a given site μ from all state vectors in \mathcal{L} and, thus, the set \mathcal{S} represented by this search key and \mathcal{L} are disjoint. It follows that each search key represents a subset \mathcal{S} of $\mathcal{M} \setminus \mathcal{L}$.
2. To prove that the complete set $\mathcal{M} \setminus \mathcal{L}$ is represented by the search basis \mathcal{B} , it is sufficient to show that for any state \vec{x} in $\mathcal{M} \setminus \mathcal{L}$, there is a subset \mathcal{S} of $\mathcal{M} \setminus \mathcal{L}$ represented by a search key $\vec{b}_{\mathcal{S}}$ such that $\vec{x} \in \mathcal{S}$.

After the first recursion of `CreateSearchBasis`, there are two possibilities: either a search key is generated with only one fixed site which adopts the same form as in \vec{x} , or a sublist \mathcal{L}_{sub} and an associated list vector \vec{t}_{sub} exist such that \vec{t}_{sub} has its fixed sites in common with \vec{x} . In the first case, the search key represents the subset \mathcal{S} which includes \vec{x} . Hence, \vec{x} is represented by \mathcal{B} .

The second case can occur for two reasons: either a site μ was found with unused forms but \vec{x} has a used form at site μ , or no site with an unused form was found and an undefined site was chosen. In this case, the algorithm proceeds further. After the next recursion, either an adequate search key is constructed or a sublist exists with an associated list vector having its two fixed sites in common with \vec{x} .

Proceeding further, either a search key $\vec{b}_{\mathcal{S}}$ is generated such that $\vec{x} \in \mathcal{S}$, or finally a sublist \mathcal{L}_{sub} of \mathcal{L} exists with an associated list vector \vec{t}_{sub} which has only one undefined site and is identical to \vec{x} at all other $N-1$ sites. Since \vec{x} is assumed to be in $\mathcal{M} \setminus \mathcal{L}$, it follows that \vec{x} has to differ from all vectors in \mathcal{L}_{sub} at the site not fixed in \vec{t}_{sub} . Consequently, the form that \vec{x} adopts at this undefined site is unused and a search key $\vec{b}_{\mathcal{S}}$ is created that represents a subset \mathcal{S} of $\mathcal{M} \setminus \mathcal{L}$ which contains \vec{x} . In this case, all sites are fixed in $\vec{b}_{\mathcal{S}}$ and the subset \mathcal{S} contains only the state vector \vec{x} .

Thus, by construction, all states in \mathcal{L} are excluded from \mathcal{B} and the union of the subsets \mathcal{S} , characterized by the search keys in \mathcal{B} , represent the complete set $\mathcal{M} \setminus \mathcal{L}$.



Output: $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \vec{b}_3, \vec{b}_4\}$ with

$$\begin{aligned} \vec{b}_1 &= (*, *, *, 0) \\ \vec{b}_2 &= (1, 0, *, 1) \\ \vec{b}_3 &= (1, 1, 0, 1) \\ \vec{b}_4 &= (0, *, 1, 1) \end{aligned}$$

Figure 6.4. Example for the generation of a search basis \mathcal{B} . The system has four sites that can adopt the forms 1 and 0. The list \mathcal{L} contains three states: $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$. An unused form h is a form that does not appear in \mathcal{L} or $\mathcal{L}_{\text{sub},i}$ for a given site μ . In the list vectors \vec{t} and the search keys \vec{b} , sites can be either fixed to a form, i.e., 1 or 0, or undefined, *. The output is the search basis \mathcal{B} .

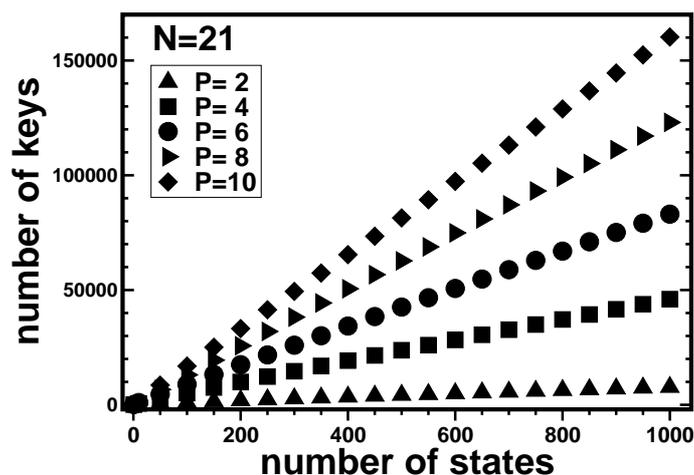


Figure 6.5. Statistics of the search basis I. The total number of search keys plotted in dependence on the number of states in lists \mathcal{L} . The system has $N=21$ sites. The number of forms P ranges from 2 to 10. The lists \mathcal{L} contain random states.

6.3 COMPUTATIONAL DETAILS

CreateSearchBasis is analyzed using random systems with $N=20-40$ sites and $P=2-10$ forms. For these systems, lists were generated that contain between 0 and 1000 random states. Additionally, a high resolution X-ray structure of bacteriorhodopsin's M intermediate (PDB code: 1kg8, *cf.* Table 3.1 [43]) was used. The protonatable residues arginine, aspartate, glutamate, lysine and tyrosine and the retinal Schiff base are defined as sites. Each of these protonatable sites can bind one proton and, thus, can adopt $P=2$ forms, protonated with $x=1$ and deprotonated with $x=0$. The application of X-DEE to protonation state calculation of proteins is presented in detail in the next chapter. In this chapter, the bacteriorhodopsin system is used to compare the behavior of a non-random to a random system. The M structure contains $N=39$ sites resulting in a total of $M=P^N=2^{39} \approx 5 \cdot 10^{11}$ states. The M structure will be referred to as full M system. Additionally, a system small enough to allow an exhaustive search of the state space was derived from the M structure. As sites were chosen: Asp36, Asp38, Lys40, Lys41, Glu74, Arg82, Asp85, Asp96, Asp102, Asp104, Asp115, Lys129, Arg164, Glu164, Glu166, Arg175, Glu194, Glu204, Asp212, Arg225 and the retinal Schiff base. In total, $N=21$ sites were chosen resulting in $M=P^N \approx 2\,000\,000$ states. All other sites were fixed to their standard protonation form at $\text{pH}=7$. This systems will be referred to as small M system.

6.4 STATISTICS OF THE SEARCH BASIS \mathcal{B}

X-DEE divides the set $\mathcal{M} \setminus \mathcal{L}$ into suitable subsets \mathcal{S} each characterized by a so-called search key $\vec{v}_{\mathcal{S}}$ (*cf.* Figure 6.3 and 6.4). \mathcal{M} is the complete set of states of the system and \mathcal{L} is a list of states to be excluded from the search for the lowest energy states. Together,

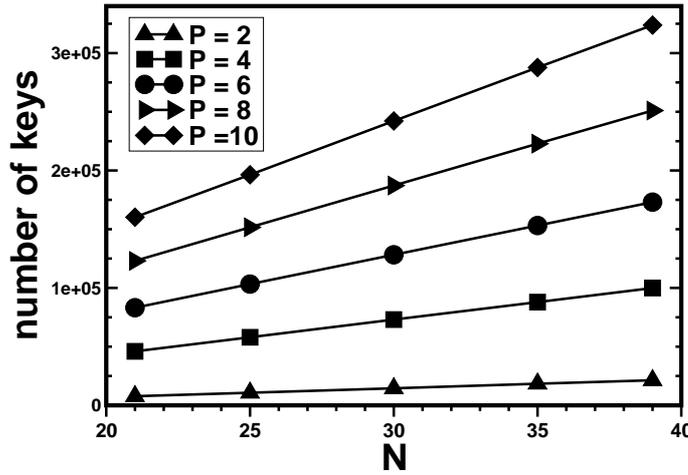


Figure 6.6. Statistics of the search basis II. Total number of search keys plotted in dependence on the number of sites N for numbers of forms P ranging from 2 to 10. The search keys were generated for lists of 1000 random states.

the search keys \vec{b}_S form the search basis \mathcal{B} . Each subset \mathcal{S} represented by the respective search key \vec{b}_S is subjected to a separate DEE search. Since the DEE search is the most time consuming operation of X-DEE, the computational performance of X-DEE depends on the number of search keys and their properties. The search basis \mathcal{B} is analyzed in terms of the absolute number of keys, the number of undefined sites per key and the number of new keys per state added to \mathcal{L} .

6.4.1 TOTAL NUMBER OF SEARCH KEYS

In Figures 6.5 and 6.6, the total number of search keys is plotted in dependence on the number of states in \mathcal{L} and on the number of sites N , respectively. These graphs show the dependence of the number of search keys on (i) the number of states in \mathcal{L} , (ii) the number of forms P , and (iii) the number of sites N . The depicted systems have $N=20$ to 40 sites and $P=2$ to 10 forms. The lists contain between 0 and 1000 random states.

(i) Figure 6.5 shows that the total number of search keys depends linearly on the number of states in \mathcal{L} . The number of states in \mathcal{L} influences the number of search keys in two different ways: on the one hand, each additional state in \mathcal{L} increases the number of states to be excluded from the search and thereby tends to increase the number of generated keys. On the other hand, each additional state in \mathcal{L} decreases the search space $\mathcal{M} \setminus \mathcal{L}$ and thereby tends to decrease the number of generated keys. Ultimately, the number of search keys will decrease with the number of states in \mathcal{L} . However, as long as \mathcal{L} is small compared to $\mathcal{M} \setminus \mathcal{L}$, an approximately linear increase of the total number of search keys can be observed.

(ii) Furthermore, Figure 6.5 shows the total number of search keys in dependence on the number of states in \mathcal{L} for systems with $N=21$ sites and a number of forms P ranging from 2 to 10. For each additional form, one additional key can be generated per site. Thus,

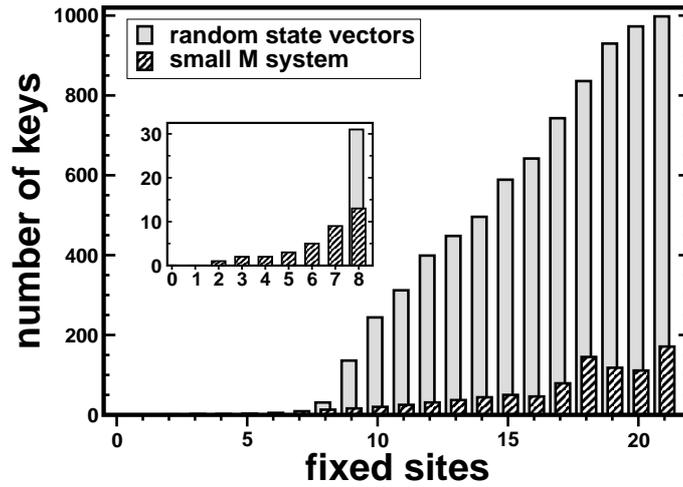


Figure 6.7. Properties of the search keys I. Distribution of search keys with respect to the number of fixed sites. The search basis was generated for two systems with $N=21$ sites and $P=2$ forms, for a list of 1000 states. One list contained random state vectors, and the other contained the 1000 lowest energy states of the small M system. The insert shows a magnification of the range of fixed sites from 0 to 8.

increasing the number of forms by 1 increases the slope of the curve by approximately N .

(iii) Figure 6.6 depicts the total number of search keys in dependence on the number of sites N for a number of forms P ranging from 2 to 10. The search keys were generated for lists of 1000 random states. The number of recursion levels of CreateSearchBasis is equal to the number of sites. Thus, the number of generated keys increases approximately linearly with N .

From Figures 6.5 and 6.6 it can be seen that the number of search keys increases linearly with the size of the system. While the total number of keys can be relatively high, in the next subsection it will be shown that the computational cost of X-DEE does not increase as quickly as the number of keys.

6.4.2 PROPERTIES OF THE SEARCH KEYS

The computational cost of a DEE search depends on the size of the subset \mathcal{S} that has to be searched. The size of \mathcal{S} is directly correlated to the number of undefined sites in the search key $\vec{b}_{\mathcal{S}}$. In other words, the more sites are fixed in $\vec{b}_{\mathcal{S}}$, the faster is the DEE search due to the scaling behavior of the Goldstein algorithm [212]. In Figure 6.7, the total number of search keys is plotted in dependence on the number of fixed sites per key for a system with $N=21$ sites and $P=2$ forms. The search keys were generated for two lists: a list of 1000 random states and the gap-free list of 1000 low energy states of the small M system. As can be seen, the majority of search keys has a large number of fixed sites. Each recursion level of CreateSearchBasis potentially increases the number of sublists \mathcal{L}_{sub} for which search keys can be generated. The higher the recursion level, the more sites will be fixed in the generated keys. Thus, with each recursion level more keys can be generated, but more sites will be fixed in these search keys. DEE searches

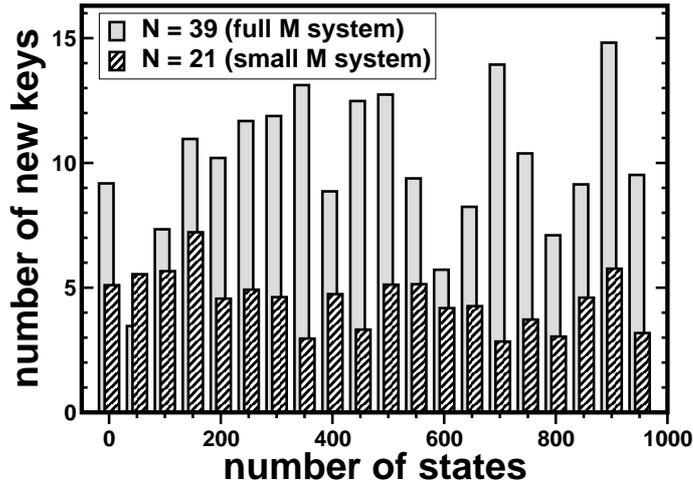


Figure 6.8. Properties of the search keys II. The average number of new keys in dependence on the number of states in \mathcal{L} for systems with $P=2$ forms and $N=21$ and 39 sites, *i.e.*, the small and full M system, respectively. The average taken over 50 consecutive states is plotted.

of the keys generated during higher recursion levels will, therefore, be increasingly less expensive.

Furthermore, Figure 6.7 also shows that significantly more search keys are generated for a list of random states than for a list of low energy states. This behavior is due to the fact that, in contrast to random states, the consecutive lowest energy states of a non-random system resemble each other. Lists of similar states lead to fewer sublists \mathcal{L}_{sub} during early recursion levels, since less used forms will exist for a given site in a given sublist \mathcal{L}_{sub} . During the early recursion levels, this smaller number of sublists leads to the generation of more keys for lists of lowest energy states (see inset in Figure 6.7). Since the final number of sublists \mathcal{L}_{sub} is equal to the number of states in \mathcal{L} , during later recursion levels, there will be a larger number of subdivisions for a list of lowest energy states. This implies that for the list of lowest energy states significantly less search keys are generated than for lists of random states. Thus, X-DEE is significantly more efficient for non-random (biological) systems.

6.4.3 NUMBER OF NEW KEYS FOR EACH NEW STATE IN \mathcal{L}

X-DEE can be used to build up a list of the lowest energy states one after the other. Then, each $\mathcal{L} \cup \{\vec{x}_{\text{new}}\}$ differs from \mathcal{L} only by the new state \vec{x}_{new} . The search keys representing $\mathcal{M} \setminus (\mathcal{L} \cup \{\vec{x}_{\text{new}}\})$ will be mostly identical to those keys representing $\mathcal{M} \setminus \mathcal{L}$. To obtain the next state, only subsets represented by new keys have to be subjected to a DEE search. Therefore, the computational cost to determine the next state will be proportional to the number of new keys necessary to characterize $\mathcal{M} \setminus (\mathcal{L} \cup \{\vec{x}_{\text{new}}\})$.

In Figure 6.8, the number of new keys is plotted in dependence on the number of states in \mathcal{L} for systems with $P=2$ forms and $N=21$ and 39 sites, *i.e.*, for the small and full M

systems, respectively. The number of new keys is averaged over 50 consecutive states. As can be seen, the number of new keys for every new state in \mathcal{L} remains approximately constant. Peaks in the number of new keys can be observed that amount to a maximum of 112 and 253 new keys for the small and the full M system, respectively. These peaks occur, if the new state that was added to \mathcal{L} originated from a search key that represents a large subset \mathcal{S} of $\mathcal{M} \setminus \mathcal{L}$, *i.e.*, if the state was found by a search key with a large number of undefined sites. This large subset \mathcal{S} , previously represented by only one search key, is split up into many smaller subsets, each represented by a new search key. The frequency and also the amplitude of the peaks decrease with the number of states in \mathcal{L} , since search keys with mostly undefined sites become less probable. On average, for the small 5 new keys and for the full M systems 10 new keys are generated per state added to \mathcal{L} . The average computational cost to determine an additional state is constant.

6.5 APPLICABILITY OF X-DEE

The X-DEE algorithm introduced in this chapter allows to determine a gap-free list of the lowest energy states. For X-DEE to be applicable, the systems have to fulfill certain conditions. In principle, X-DEE can be applied to all systems that can be treated with the DEE algorithm. That is, their state energy must be expressible as pairwise interaction between sites and these sites assume discrete forms that differ in their contribution to the state energy. Figure 6.1 shows an example of these sites, *i.e.*, protein residues and forms, *i.e.*, side chain positions.

X-DEE can be used to build up a gap-free list of the lowest energy states. Applying X-DEE to a list of lowest energy states, the state next in energy is found. Adding this state to the list and consecutively applying X-DEE to the new list of lowest energy states, a complete list of the lowest energy states is determined one after the other. Alternatively, X-DEE can be used to complete any list of states by eliminating gaps. If X-DEE is applied to an incomplete list of states, the algorithm will find the state of lowest energy missing from this list. Adding this state to the list and applying X-DEE to the new list, the gaps of the list will be filled one state after the other. Such lists can for instance be generated by Monte Carlo sampling. As described in Chapter 2, Metropolis Monte Carlo samples low energy states by approximating the Boltzmann distribution of states. There is, however, no verification as to whether the state of lowest energy is found or whether the sampled states represents a gap-free list of the lowest energy states. X-DEE can be used to complete this list or verify a chance completeness.

X-DEE will be of general interest, when a gap-free list of the lowest energy states is required to advance the understanding of a system. Such a list allows to analyze the ensemble of states accessible to a system in a certain energy range and, thus, can provide useful information on the functional mechanism. X-DEE can for example be applied to obtain a gap-free list of the lowest energy states in homology modeling and X-ray refinement. A particularly interesting application is enzyme design. A good catalyst needs not only a high binding affinity for the reactant but, in particular, has to stabilize the transition state of its substrate. X-DEE can be a powerful tool in enzyme design by optimizing the energy difference between these states. Furthermore, in enzymes the

protonation of the active site commonly plays a crucial role for its function. Calculating the lowest energy protonation states with X-DEE is, thus, of major interest. In addition, X-DEE provides useful starting information for the theoretical investigation of kinetic properties. By providing a gap-free list of the lowest energy states, the complete pool of states that may be occupied by the system is available. The system kinetics can then be analyzed by determining the transitions between these states.

In this work, X-DEE was implemented for the calculation of protonation states. The theoretical approach is discussed in the next chapter. Furthermore, a first application to the proton pump bacteriorhodopsin is presented and an analysis of the lowest energy protonation states derived with X-DEE is given.

CHAPTER 7

APPLICATION OF X-DEE TO BACTERIORHODOPSIN: DETERMINATION OF ALL LOWEST ENERGY PROTONATION STATES

Energy is eternal delight.

William Blake

Protonation state changes are crucial for many enzymatic reactions and other biochemical processes such as signal transduction, photosynthesis and proton transfer across membranes [65, 77, 113, 223, 224]. Even without protonation state changes, the charge distribution of a protein is essential for its function: In the context of the relatively apolar protein matrix a full charge can have a strong influence on the structure and the function. Furthermore in proteins, commonly not only the state of lowest energy but also the next higher protonation states are significantly populated. The active site of enzymes often contains charged groups, for example, deprotonated aspartate or protonated histidine residues. A change in the charge of such a crucial residue may be functionally important or may inactivate an enzyme. The knowledge of all protonation states that are energetically accessible to a protein can significantly advance the understanding of its functional mechanism. However, due to the large number of all possible states even for small proteins, an exhaustive search of all protonation states is usually not feasible. The extended dead-end elimination (X-DEE) algorithm introduced in the previous chapter provides a solution to this problem.

X-DEE was developed to generate gap-free lists of the lowest energy states. In principle, the algorithm is applicable to all molecular systems that can be treated by the dead-end elimination (DEE) algorithm [209]. More precisely, their state energy is required to be expressible in terms of pairwise interaction between sites. Furthermore, the energy contribution of these sites to the state energy must be discrete or discretized. That is, it must be possible to define a limited number of forms these sites can adopt that differ in their energy contribution to the state energy. X-DEE can for example be used in the homology modeling of protein structures. Then, the protein residues constitute the sites and their rotamer positions the forms of these sites. An illustration can be found in Figure 6.1 in the previous chapter. An in depth discussion of the theory underlying the X-DEE algorithm and an analysis of its computational performance is given in the previous chapter as well. This chapter presents the first application of the newly introduced X-DEE algorithm to a biological system.

Here, X-DEE is implemented to generate gap-free lists of the lowest energy protonation states of proteins. In the following section, the implementation of the X-DEE algorithm for protonation state calculations will be discussed in detail. The computational details of the calculations presented in this chapter are given in Section 7.2. X-DEE is applied to bacteriorhodopsin (BR), the prototypical example of a proton pump. Absorption of light energy by BR initiates the so-called photocycle during which a proton is translocated across the membrane (depicted in Figure 1.8 in Chapter 1). Using X-DEE, the lowest energy protonation states are calculated for two BR structures. These structures represent the intermediate state before and after the first proton transfer step, *i.e.*, the bR and the M₁ state. A general analysis of the determined states will be given in Section 7.3. Section 7.4 discusses the calculated protonation states with respect to the proton transfer mechanism of BR.

7.1 IMPLEMENTATION OF X-DEE

Proteins contain protonatable residues whose charged state depends on their interaction with the protein environment. These protonatable residues are treated as the sites of the system. The sites can assume different protonation forms. Depending on the pH value and their environment, aspartate, glutamate, cysteine and tyrosine residues can be negatively charged. In contrast, arginine, histidine and lysine residues can be positively charged.

While most protonatable residues can adopt two forms, protonated and deprotonated, the histidine residue is a special case. It can assume four different forms: doubly deprotonated, singly protonated at the δ -nitrogen atom, singly protonated at the ϵ -nitrogen atom, and doubly protonated. For many applications of X-DEE, different number of forms for the sites of the system will be the rule rather than the exception. The number of forms per site can be easily and individually defined and pose no problem for the X-DEE algorithm. However, the protein investigated in this study, BR, contains no histidine residue. The following discussion will, therefore, focus on the simple case that all sites can adopt two different forms: protonated and deprotonated.

THE PROTONATION STATE ENERGY

A general formulation of the state energy for systems treatable with X-DEE is given by Eq. (6.1) in the previous chapter. In equivalence, the energy for a protonation state is defined on the basis of pairwise interaction between the protonatable sites:

$$G(x) = RT \ln 10 \sum_{\mu=1}^N \left((x_{\mu} - x_{\mu}^{\circ}) (\text{pH} - \text{p}K_{\text{intr},\mu}) \right) + \frac{1}{2} \sum_{\mu=1}^N \sum_{\nu=1}^N \left(W_{\mu\nu} (x_{\mu} - x_{\mu}^{\circ}) (x_{\nu} - x_{\nu}^{\circ}) \right) . \quad (7.1)$$

The first term represents the intrinsic protonation energy. The $\text{p}K_{\text{intr}}$ is the $\text{p}K_{\text{a}}$ value a protonatable residue has in the protein environment, if all other protonatable residues are in their neutral form. As described by Eq. (2.17) in Chapter 2, the $\text{p}K_{\text{intr}}$ value is determined by constant background charges and the permittivity of the medium. The second term considers the interaction energy $W_{\mu\nu}$ between a pair of sites μ and ν . The energy contribution of each protonatable residue to the state energy depends on its protonation form. The discrete forms of these sites x_{μ} are characterized with respect to the number of protons bound: the value of x_{μ} is 1 if site μ is protonated and 0 if site μ is deprotonated in the considered state. The reference protonation form of the site μ , *i.e.*, its neutral form, is given by x_{μ}° , where $x_{\mu}^{\circ} = 1$ for acids and $x_{\mu}^{\circ} = 0$ for bases. For site ν , the protonation form x_{ν} and the reference protonation form x_{ν}° is defined equivalently.

THE DEE CRITERION

X-DEE is based on the DEE theorem introduced by Desmet *et al.* [209]. DEE allows to determine the lowest energy state of a system by identifying so-called *dead-ends*. These dead-ends are high-energy forms of sites that are incompatible with the state of lowest energy. Dead-ends are excluded from further consideration. Each eliminated dead-end lowers the size of the state space in which the state of lowest energy can be found. Thereby, the combinatorial problem is reduced, finally, allowing an exhaustive search for the lowest energy state in the remaining state space. Those forms of sites that can be considered dead-ends in the search for the lowest energy state are identified by the so-called DEE criterion.

The DEE criterion is an inequality that compares the energy contribution of two forms, c and d of one site α to the state energy. A certain form d_{α} is defined as a dead-end, if the state energy can always be lowered by changing the form of site α from d to another form c . A detailed account of the underlying theory and a general formulation of the DEE criterion can be found in Section 6.1 in Chapter 6.

To derive the DEE criterion, the protonation state energy given by Eq. (7.1) is defined for

the case that site α takes the specific form f :

$$G_{f_\alpha}(x) = RT \ln 10 \left((f_\alpha - x_\alpha^\circ) (\text{pH} - \text{p}K_{\text{intr},\alpha}) + \sum_{\mu \neq \alpha} (x_\mu - x_\mu^\circ) (\text{pH} - \text{p}K_{\text{intr},\mu}) \right) \\ + \sum_{\mu \neq \alpha} W_{\alpha\mu} (f_\alpha - x_\alpha^\circ)(x_\mu - x_\mu^\circ) + \frac{1}{2} \sum_{\mu \neq \alpha} \sum_{\nu \neq \alpha} W_{\mu\nu} (x_\mu - x_\mu^\circ)(x_\nu - x_\nu^\circ) . \quad (7.2)$$

The first two terms represent the intrinsic protonation energy of site α and all other sites μ , respectively. The third and fourth term describe the interaction energy between site α and all other sites μ and the interaction between all other sites μ and ν , respectively. From Eq. (7.2), the energy of every protonation state of the system that contains site α in the specific protonation form f can be calculated.

To compare the energy contribution of two specific forms c and d of site α to the state energy, these are substituted in Eq. (7.2) for f . Subtracting $G_{c_\alpha}(x)$ from $G_{d_\alpha}(x)$ leads to:

$$G_{d_\alpha}(x) - G_{c_\alpha}(x) = RT \ln 10 (d_\alpha - c_\alpha) (\text{pH} - \text{p}K_{\text{intr},\alpha}) + \sum_{\mu \neq \alpha} W_{\alpha\mu} (d_\alpha - c_\alpha)(x_\mu - x_\mu^\circ) . \quad (7.3)$$

To prove that d_α is a dead-end, it is sufficient to show that the minimum energy difference between the states that contain either c_α or d_α is larger than zero. In other words, d_α is a dead-end, if the state energy can always be lowered by changing the form of site α from d to c . In analogy to the general DEE criterion given by Ineq. (6.5) in Chapter 6, the criterion for protonation state calculations takes the following form:

$$RT \ln 10 (d_\alpha - c_\alpha)(\text{pH} - \text{p}K_{\text{intr},\alpha}) + \sum_{\mu \neq \alpha} \min_x \left[W_{\alpha\mu} (d_\alpha - c_\alpha) (x_\mu - x_\mu^\circ) \right] > 0 . \quad (7.4)$$

The minimum energy difference between the interaction of site α with all other sites μ is determined by searching over all possible forms x_μ . When Ineq. (7.4) is true, the DEE criterion is fulfilled. That is, the state energy can always be lowered by changing the form of site α from d to c . Then, d_α is a dead-end that can be excluded from further considerations. The DEE criterion is successively applied to all forms of all sites. A graphic representation of the DEE criterion is given in Figure 6.2 in Chapter 6.

Identifying one form of a site α as a dead-end fixes this site to the remaining form $f_\alpha = 1 - d_\alpha$. As a consequence of fixing site α to the remaining form f , the DEE criterion is not applied to site α again, since the DEE criterion compares the contribution to the state energy of two forms of one site. Furthermore, the DEE criterion given by Ineq. (7.4) can be evaluated computationally more efficiently, if the sum in the second term is taken only over so-called unfixed sites. For sites that are fixed to a certain form, the contribution to the state energy is constant. To account for the constant contribution of the so-called fixed sites, the $\text{p}K_{\text{intr}}$ value of the unfixed sites μ can be corrected:

$$\text{p}K_{\text{intr},\mu}^{\text{new}} = \text{p}K_{\text{intr},\mu} - \frac{1}{RT \ln 10} \sum_{\alpha}^{N_{\text{fixed}}} W_{\mu\alpha} (f_\alpha - x_\alpha^\circ) \quad (7.5)$$

where the value of f is either 1 or 0 depending on whether site α is fixed in the protonated or deprotonated form. The corrected $\text{p}K_{\text{intr},\mu}^{\text{new}}$ value of the unfixed sites includes the in-

teraction with the fixed sites as a constant energy contribution. The correction has to be carried out each time a dead-end is identified, before the DEE criterion is applied again.

In the special case described here, the sites have only two forms, protonated and deprotonated. However, it is possible to generalize this example for systems whose sites have more than two forms. In general, if all forms of one site except one are identified as dead-ends, this site can be fixed. The remaining DEE search can then be evaluated computationally more efficiently by including the energy contribution of the fixed site as a constant term and taking the sum in the DEE criterion only over the unfixed sites.

DETERMINATION OF A GAP-FREE LIST OF LOWEST ENERGY STATES

The application of DEE allows to determine the state of lowest energy. This study aims, however, at identifying a gap-free list of all lowest energy states. The X-DEE algorithm introduced in the previous chapter utilizes the function of DEE to generate such gap-free lists of lowest energy states. The principal idea of X-DEE is to exempt a certain list of states from a DEE search. For example, by excluding the lowest energy state of the system, the DEE search will then determine the second lowest energy state. Thus, one after another a gap-free list of the lowest energy states can be generated. If the list of states exempt from the DEE search is not gap-free, the state of lowest energy not included in the list will be determined. X-DEE can consequently be used to complete any given list of states until a complete list of all lowest energy states is known.

Excluding an arbitrary list of states from a DEE search is, however, not trivial. This problem is solved by X-DEE. Input of the core function of X-DEE, `CreateSearchBasis`, is a list of states \mathcal{L} that shall be excluded from the DEE search. `CreateSearchBasis` then generates so-called search keys that represent all states of the system that are not in \mathcal{L} . Subjecting these search keys to a DEE search will identify the state of lowest energy of $\mathcal{M} \setminus \mathcal{L}$, where \mathcal{M} are all states of the system. An in depth description of the X-DEE algorithm is given in Section 6.2 in Chapter 6.

7.2 COMPUTATIONAL DETAILS

Calculations were performed on two BR structures that represent the bR state, *i.e.*, PDB code 1c3w and the M state, *i.e.*, PDB code 1kg8 (*cf.* Table 3.1). The protonatable residues arginine, aspartate, glutamate, lysine and tyrosine and the retinal Schiff base are defined as sites. These residues can bind one proton and, thus, can adopt $P=2$ forms, protonated with $x=1$ and deprotonated with $x=0$. The two BR structures each contain $N=39$ sites resulting in $M=P^N \approx 5 \cdot 10^{11}$ states. In the following, the bR and M structure will be referred to as bR system and M system, respectively. To differentiate between the bR and M systems, *i.e.*, BR structures, and the functionally relevant protonation states, the latter are enclosed in squared brackets ([bR], [M₁], and [M₂]). The protonation states are listed in Table 5.1 in Chapter 5.

Additionally, systems small enough to allow an exhaustive search of the combinatorial space were derived from the bR and M structure. As sites were chosen: Asp36, Asp38, Lys40, Lys41, Glu74, Arg82, Asp85, Asp96, Asp102, Asp104, Asp115, Lys129, Arg164,

Glu164, Glu166, Arg175, Glu194, Glu204, Asp212, Arg225 and the retinal Schiff base. In total, $N=21$ sites were chosen resulting in $M=P^N \approx 2\,000\,000$ states. All other sites were fixed to their standard protonation form at $\text{pH}=7$. The fixed sites are treated as a constant energy contribution using Eq. (7.5). These systems will be referred to as small bR system and small M system.

The $\text{p}K_{\text{intr}}$ values and the interaction energy matrix $W_{\mu\nu}$ were calculated for the bR and M system and their small derivatives as described in Chapter 3. All states of the small bR and small M system are calculated. For the bR and M_1 system, the lowest energy states are calculated in an energy range of $15 k_B T_{300\text{K}}$.

7.3 GENERAL ANALYSIS OF THE LOWEST ENERGY STATES

In this section, a general analysis of the determined lowest energy states will be presented. First, it will be demonstrated that a relatively small number of the protonation states can be used to derive the thermodynamic properties of the system. Then, the energy distribution of the generated X-DEE states will be discussed. Furthermore, the probability of the protonation states can be approximated as described in Subsection 7.3.3.

7.3.1 CONVERGENCE OF THE PARTITION FUNCTION

Thermodynamic properties are efficiently calculated using Metropolis Monte Carlo. This method achieves a convergence of thermodynamic properties by sampling a large fraction of the phase space biased towards the Boltzmann distribution of states [225]. In this work, Metropolis Monte Carlo was applied to BR. The successful description of protonation probabilities for BR is presented in Chapter 5. The Metropolis Monte Carlo approach is described in detail in Chapter 2.

While Metropolis Monte Carlo can provide a set of low energy states, there is no guarantee that this set is gap-free, *i.e.*, that no states with lower energy exist. X-DEE, in contrast, was developed to provide a complete description of the lowest energy states. X-DEE will primarily be utilized to analyze the lowest energy states individually and not as an ensemble. However, since the thermodynamic properties of a system depend mainly on the states of lowest energy, additionally, X-DEE may be used to derive thermodynamic properties.

Thermodynamic properties depend on the partition function Z which is the sum of the Boltzmann factors over all possible states of the system:

$$Z = \sum_n^M \exp(-\beta G_n) . \quad (7.6)$$

where M is the total number of states, G_n is the state energy and β equals $1/(k_B T)$; k_B is the Boltzmann constant and T the absolute temperature. The exact value of Z can only be obtained, if all state energies of the system are known. Then, the exact probability P

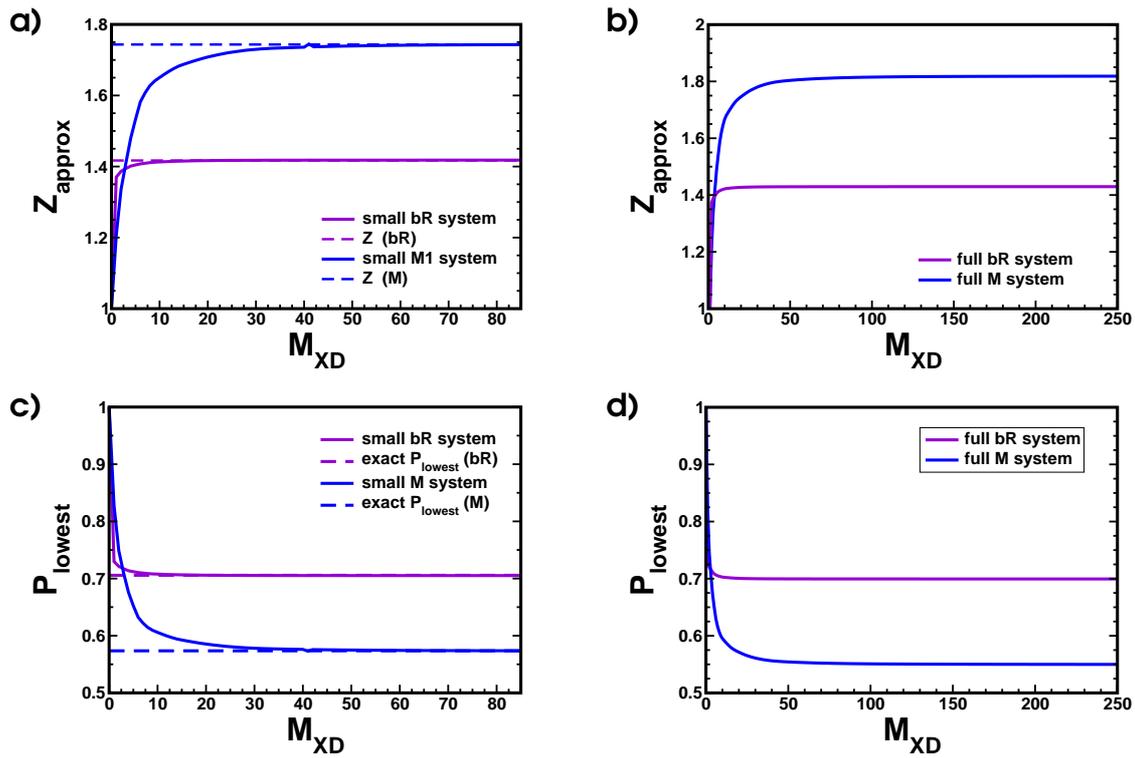


Figure 7.1. Convergence of Z_{approx} and the probability of the lowest energy state. Z_{approx} (cf. Eq. (7.8)) is plotted in dependence on the number of states M_{XD} considered to approximate the partition function: **a)** small systems, and **b)** full systems. The dotted lines indicate Z calculated from all states for the small systems. The probability of the lowest energy state P_{lowest} is plotted in dependence on the number of states M_{XD} considered to approximate the partition function: **c)** small systems, and **d)** full systems. For the small systems, the dotted lines indicate the exact probability of the lowest energy state calculated from all states. All calculations were performed at pH = 7.

of each state n can be calculated as:

$$P_n = \frac{\exp(-\beta G_n)}{Z} . \quad (7.7)$$

The calculation of all state energies is, however, not feasible for complex systems. Here, the partition function is approximated as:

$$Z_{\text{approx}} = \sum_n^{M_{\text{XD}}} \exp(-\beta G_n) , \quad (7.8)$$

where the sum is taken over the number of states M_{XD} that were determined by X-DEE. The approximate probability P of each state n is calculated as:

$$P_n = \frac{\exp(-\beta G_n)}{Z_{\text{approx}}} . \quad (7.9)$$

Since all lowest energy states are known, the convergence of the partition function can be analyzed with respect to the number of states that have to be considered.

Lists of lowest energy states are calculated for the full bR and M systems and for the small bR and M systems. For the small systems the total number of states is approximately 2 million. Thus, the calculation of all state energies and, therefore, of the exact partition function is possible.

Figure 7.1 a and 7.1 b show the convergence of the partition function for all systems in dependence on the number of considered states. In Figure 7.1 a, the exact partition function is indicated for the small systems. As can be seen in Figure 7.1 a and 7.1 b, the approximated partition function Z_{approx} converges quickly with the number of considered states. Likewise the probability of the lowest energy state P_{lowest} converges quickly to the exact probability for the small systems as shown in Figure 7.1 c. Figure 7.1 d shows the convergence behavior of the probability of the lowest energy state P_{lowest} for the full systems. For these, a convergence of the approximated partition function and the probability of the lowest energy state to a constant value can be observed when approximately 50 lowest energy states are considered. Z_{approx} does not change anymore between the 1500th and the 3000th state. Both, the full and the small M system converge slower than the respective bR system, indicating a larger density of states in the low energy range for the M system.

7.3.2 ENERGY DISTRIBUTION OF THE LOWEST ENERGY STATES

In Figure 7.2, the energy distribution of the lowest energy states is shown for the bR and M system. All states in an energy range of $15 k_B T_{300K}$ are depicted. As can be seen, the density of states is significantly higher for the M system than for the bR system. For the bR system approximately 200 states fall into the $15 k_B T_{300K}$ range, while the M system has approximately 700 states in the same energy range. An indication of this behavior was already given by the slower convergence of the partition function for the M system (*cf.* Figure 7.1 b). Whether a system is well represented by the lowest energy state or by several lowest energy states, can only be decided when these states are known. Interestingly, for the bR system the first two states are closer in energy than for the M system (see inset in Figure 7.2).

7.3.3 PROBABILITY OF THE LOWEST ENERGY STATES

As shown in Figure 7.1, the partition function can be approximated by relatively few lowest energy states. Thus, the probability of each state can be calculated to a good approximation when only a few lowest energy states are known. Therefore, for the full bR and M system, the probability for each lowest energy state P_n was calculated from the converged partition function using Eq. (7.9). The calculations were performed for different pH values. Here, the cumulative probability P_{tot} is defined as the sum of probabilities up to a given number of states M_{XD} :

$$P_{\text{tot}} = \sum_{n=1}^{M_{\text{XD}}} P_n = \sum_{n=1}^{M_{\text{XD}}} \frac{\exp(-\beta G_n)}{Z_{\text{approx}}} \quad , \quad (7.10)$$

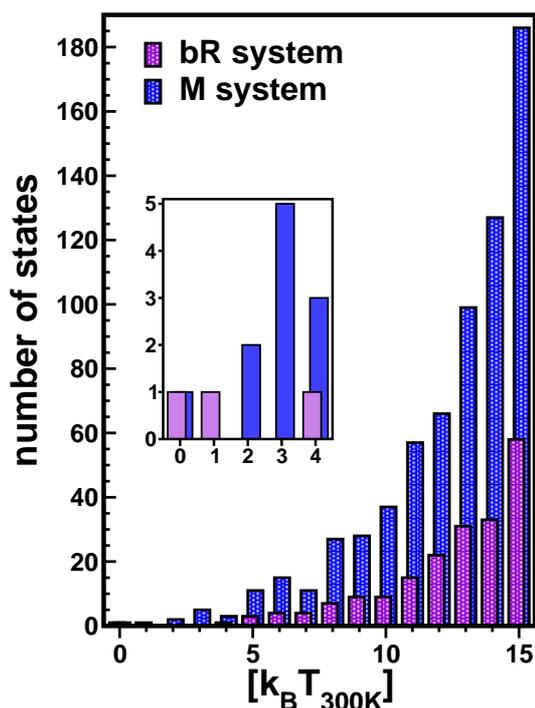


Figure 7.2. Distribution of state energies. All lowest energy states in an energy range of $15k_B T_{300K}$ are considered for the bR and the M system. The state energies were calculated at $\text{pH}=7$. The insert shows a magnification of the $k_B T_{300K}$ range from 0 to 4.

where the sum is taken in consecutive order with respect to the energies of the states calculated using X-DEE.

The graphs in Figure 7.3 show P_{tot} as a function of M_{XD} . It can be seen that the cumulative probability converges rapidly to 1 at most pH values showing that relatively few lowest energy states have a significant probability. At physiological pH values, the bR structure is well represented by the lowest energy state. For a good representation of the M system, however, at least 10 states should be considered.

Additionally, the contribution of each state to P_{tot} can be deduced. The number of states with a significant probability depends on pH. At pH values of 5, 6 and 9, more states have a significant probability. Therefore, the convergence of P_{tot} is slower compared to the convergence at $\text{pH}=7$. Lowest energy states of similar energy and, thus, similar probability result from the titration behavior of the protonatable residues. For example, aspartate and glutamate residues commonly titrate at low pH values. At high pH values, lysine and arginine residues commonly titrate. States that only differ in the protonation form of these residues can be expected to have similar state energies in the respective pH range. Other states have no significant probability.

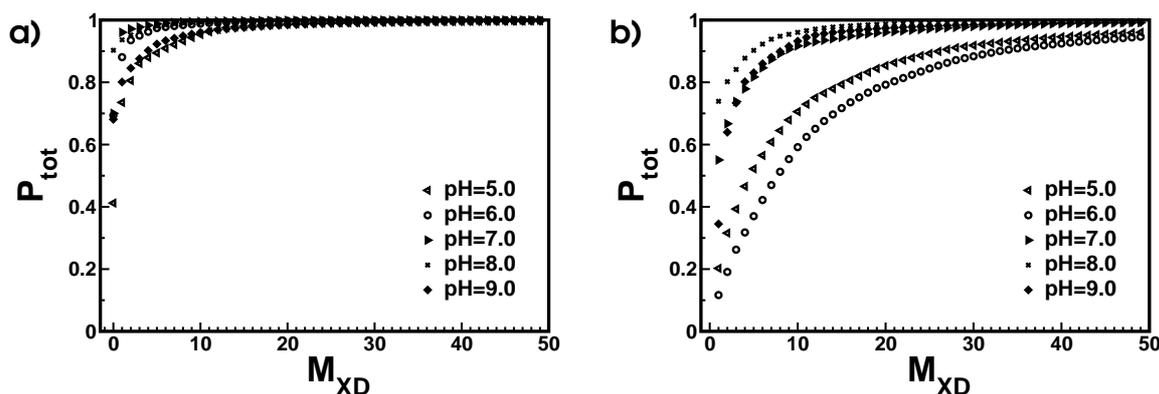


Figure 7.3. Probability of states. Cumulative probability P_{tot} as a function of M_{XD} (cf. Eq. (7.10)). **a)** bR system and **b)** M system. P_{tot} is plotted for different pH values.

7.4 THE PROTON PUMP BACTERIORHODOPSIN

BR is a light-activated proton pump that transfers protons from the cytoplasm to the extracellular space. The key residues involved in the proton transfer are from the cytoplasmic to the extracellular side: Asp96, the retinal Schiff base, Asp85, Glu194/Glu204. The protonation of Asp115, Arg115 and Asp82 is also functionally relevant. Their protonation form does, however, not change during the proton transfer. These key residues are depicted in Figure 1.7 in Chapter 1. After light absorption in the bR state (the ground state), BR undergoes a series of intermediate states, the K, L, M_1 , M_2 , N and O state. The intermediate states are characterized by their absorption maximum and changes in the protonation of the key residues. The first proton transfer step takes place in the transition from the L to M_1 intermediate. From the O intermediate BR returns to its ground state. In the next subsection, the functionally relevant protonation states of BR will be briefly reviewed. Then, the protonation states calculated for the bR and M system will be analyzed with respect to the proton transfer of BR.

7.4.1 FUNCTIONALLY RELEVANT PROTONATION STATES OF BR

To differentiate between the bR and M system (BR structures) and the [bR], [M_1], and [M_2] protonation state, the latter are enclosed in squared brackets. The [bR] protonation state is characterized by a protonated retinal Schiff base and a deprotonated Asp85. Asp96, Asp115 and Arg82 were experimentally shown to be protonated, while Asp212 was shown to be deprotonated. The two glutamate residues, Glu194 and Glu204, are part of the proton release complex which is protonated in the bR state. After the first proton transfer step, BR is in the [M_1] protonation state which is characterized by a deprotonated retinal Schiff base and a protonated Asp85. No other key residue changes its protonation form. The [M_1] state is followed by the [M_2] protonation state, where a proton has been released to the extracellular side. The [M_2] protonation state is characterized by a deprotonated retinal Schiff base, a protonated Asp85 and a deprotonated proton release complex. Subsequently, the retinal Schiff base receives a proton from Asp96 close

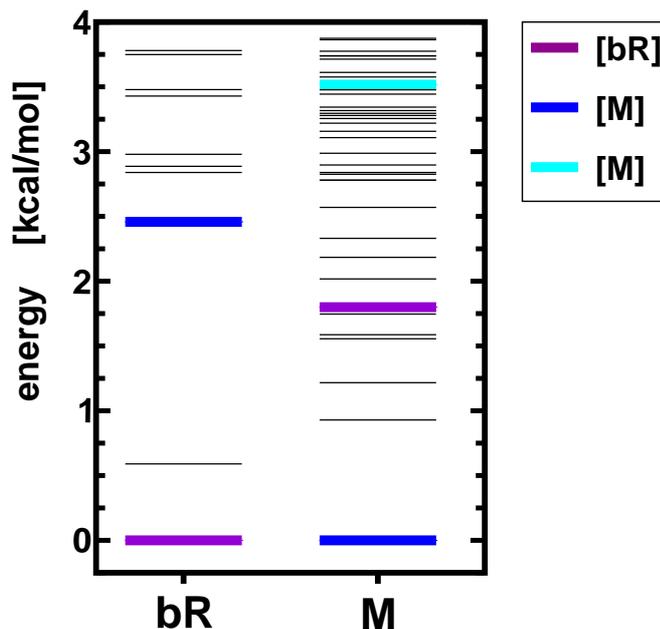


Figure 7.4. Distribution of the protonation states of BR. For the bR and the M system all state energies are plotted up to 4 kcal/mol. The lowest [bR], [M₁] and [M₂] protonation states are highlighted. These protonation states reflect the first two proton transfer steps of BR. The [bR] protonation state is characterized by a protonated retinal Schiff base and a deprotonated Asp85, the [M₁] protonation state by a deprotonated retinal Schiff base and a protonated Asp85. The [M₂] state has in addition a deprotonated proton release group.

to the cytoplasmic site leading to the [N] protonation state. Thereafter, Asp96 takes up a proton from the cytoplasm and the protein is in the [O] protonation state. Finally, Asp85 deprotonates, the proton release complex becomes protonated again and the BR returns to the [bR] state. The functionally relevant protonation states are listed in Table 5.1 in Chapter 5.

7.4.2 DISTRIBUTION OF FUNCTIONALLY RELEVANT PROTONATION STATES

Figure 7.4 shows the energy of the protonation states for the bR and the M system. The states that represent the lowest [bR], [M₁], and [M₂] protonation state are highlighted. As can be seen, for the bR system the state of lowest energy is the [bR] protonation state, and the [M₁] protonation state is the third state. The energy difference between these states is approximately 2.5 kcal/mol. The [M₂] protonation state differs from the state of lowest energy by 10.8 kcal/mol. The second state is a [bR] protonation state that differs from the lowest energy state in the protonation form of an arginine residue that is exposed to the bulk water and close to the membrane. This arginine does not interact with the key residues of proton transfer.

For the M system, the lowest energy state is the [M₁] protonation state. The [bR] protonation state is the seventh state differing from the lowest energy state by less than 2 kcal/mol. The [M₂] protonation state is about 3.5 kcal/mol higher in energy than the

state of lowest energy. Thus, for the M system the energy difference between $[M_1]$ and $[M_2]$ state is significantly smaller than for the bR system. The second state of the M system has a protonated retinal Schiff base and a protonated Asp85, *i.e.*, constituting a state that does not correspond to a functionally relevant protonation state. The following three states are $[M_1]$ protonation states where surface residues change their protonation. The sixth state has a deprotonated proton release complex, while both retinal Schiff base and Asp85 are protonated. Thus, Like the second state this state does not correspond to a functionally relevant protonation state.

7.4.3 IMPLICATIONS FOR THE PROTON TRANSFER IN BR

As shown in Figure 7.4, the states of lowest energy of the bR and M systems are in agreement with their experimental classification. For the bR system the second functionally relevant protonation state is $[M_1]$ which follows the [bR] protonation state during the proton transfer. For the M system, the next two functionally relevant protonation states are the [bR] state, corresponding to the preceding state during the proton transfer, and the $[M_2]$ state that corresponds to the following state during the proton transfer. For the M system, the energy differences between the lowest [bR], $[M_1]$, and $[M_2]$ protonation state are relatively small.

Interestingly, for both systems these lowest functionally important states differ only in the protonation of key residues of proton transfer. This observation implies that protonation changes of the key residues during the first proton transfer step are decoupled from protonation changes of the other protonatable residues. In this respect, the first proton transfer step in BR may be a rather special case. Other transfer steps may show a more complex behavior, since protonation changes of functionally relevant groups can be coupled to changes in other groups that have not yet been considered before. It will be interesting to analyze whether the behavior of other proton transfer steps in BR and in different systems is indeed more complex. Since X-DEE is able to determine gap-free lists of lowest energy states, it is well suited for analyzing the energetics of charge transfer processes.

7.5 CALCULATING THE KINETICS OF PROTON TRANSFER

Using X-DEE, gap-free lists of the lowest energy protonation states are obtained by subsequently excluding the determined states from the search space. X-DEE provides necessary starting information for theoretical investigations of kinetic properties. By providing a gap-free list of the lowest energy states, the complete pool of possible states is available, *i.e.*, states that may be occupied by the system. The system kinetics may be analyzed by determining the transitions between these states using for example differential equations or a dynamical Monte Carlo approach.

CHAPTER 8

CONCLUDING REMARKS AND OUTLOOK

*Sicher ist, daß nichts sicher ist.
Selbst das nicht.*

Joachim Ringelnatz

Gallia est omnis divisa in partes tres. . .

Gaius Iulius Caesar

Several aspects of archaeal rhodopsin proteins have been studied in this thesis, providing a more detailed picture of the functional characteristics of these proteins and of rhodopsins in general. The results can be divided into three parts. First, the absorption tuning of the retinal chromophore has been analyzed. Absorption tuning of the retinal chromophore is essential for light signaling, *e.g.*, color vision in higher organisms and phototaxis in single cell organisms. A comparative analysis of the electrostatic basis of the tuning mechanism of archaeal rhodopsin was presented in Chapter 4. Secondly in Chapter 5, the protonation behavior of the proton pump bacteriorhodopsin has been discussed. Bacteriorhodopsin has become a prototype for the investigation of the ion pumping mechanism. During this thesis, the protonation behavior of functionally important groups, their pair-correlation and the probability of functionally relevant protonation states has been examined in detail for available bacteriorhodopsin structures. Finally, an algorithm has been developed that determines a gap-free list of the lowest energy states. The so-called X-DEE algorithm, introduced in Chapter 6, is applicable to biological systems. Here, X-DEE was implemented to determine the lowest energy protonation states of proteins. The algorithm was applied to bacteriorhodopsin and the results were presented in Chapter 7.

Each Chapter closes with a conclusion related to its respective matter. This chapter provides a general synopsis of the key findings presented in this thesis and discusses questions to be addressed in the future. To begin with, the model underlying all calculations and unifying the different parts of this work is briefly reviewed in the next section.

THE PROTEIN MODEL

A continuum model is used to describe the electrostatics of a protein in aqueous solution. The protein region is modeled as a low dielectric embedded in a high dielectric solvent medium, *i.e.*, the polarizability of the protein and the solvent is represented by a low and high permittivity, respectively. Additionally, the protein atoms are explicitly described as point charges, *i.e.*, their atomic partial charges, within their van der Waals radii. The charge distribution of solvent ions is modeled by a Boltzmann distribution. Rhodopsin proteins are transmembrane proteins. Therefore, the membrane has to be included in the model. Here, the apolar region of the membrane is represented by uncharged dummy atoms within a region of low permittivity. The electrostatic potential at every point is mathematically described by the Poisson-Boltzmann equation. Finite differences are used to solve this partial differential equation numerically for complex protein systems. This model underlies all calculations performed during this work.

THE INTER-PROTEIN ABSORPTION SHIFT

Rhodopsins are part of a variety of light sensing systems ranging from phototaxis in single cell organisms to eyesight in higher organisms. The rhodopsin family of photosensitive proteins uses retinal as chromophore. The absorption spectrum of the retinal is characteristic for individual rhodopsins, *i.e.*, the proteins tune the light absorption of the chromophore. Due to its implication in color vision, the absorption tuning mechanism of the rhodopsin protein family is of medical interest and is intensively researched. However, the structural basis of the tuning mechanism and of the absorption shift between different rhodopsins is not yet understood. The three archaeal rhodopsins, bacteriorhodopsin, halorhodopsin and sensory rhodopsin II, provide an excellent test case. High-resolution structures are available for all three proteins, allowing a comparative theoretical analysis of the absorption shift between the rhodopsins in atomic detail. Furthermore, the absorption maxima of bacteriorhodopsin and halorhodopsin are nearly identical, while the spectrum of sensory rhodopsin II is considerably blue shifted, although these three proteins are structurally similar. Additionally, the conformation of the retinal Schiff base is virtually identical in all archaeal rhodopsins. Electrostatic interactions between the retinal Schiff base and the protein appear, therefore, as an essential factor for the absorption shift between the archaeal rhodopsins.

In this work, the electrostatic potential that the proteins cause at the retinal was calculated. The comparison of these potentials showed a clear distinction between bacteriorhodopsin and halorhodopsin on the one hand and sensory rhodopsin II on the other hand. This observation is in accordance with their distinct absorption maxima. Therefore, the assumption is supported that electrostatic interactions between protein and chromophore are the determinant for the absorption shift between the archaeal rhodopsins.

Based on this result, the electrostatic potential of the proteins was decomposed into the contribution of each individual protein residue. After a superposition and sequence alignment of bacteriorhodopsin, halorhodopsin and sensory rhodopsin II, the electrostatic

potential of each residue was compared. Seven residues were identified that cause the distinct differences in electrostatic potential between the proteins. It could be shown that the electrostatic potential at the retinal is virtually identical omitting these residues. Three of the residues are located in the retinal binding pocket. These have in fact been shown to account for part of the absorption shift between bacteriorhodopsin and sensory rhodopsin II by mutational studies. Another residue is located close to the β -ionone ring of retinal, but is not part of its binding pocket. The remaining three residues are located more than 8 Å away from the retinal. These four residues have been implicated in the inter-protein shift here for the first time. One reason lies in the fact that they are, because of their location, no obvious candidates for the tuning mechanism in general and the absorption shift between the archaeal rhodopsins in particular. Mutational studies may help to verify the findings of this work.

THE MODEL OF A PARTICLE IN A BOX WITH STEP POTENTIAL

The comparative analysis of the electrostatic potential indicated a strong connection between the potential and the absorption maximum for the archaeal rhodopsins. The evidence was, however, circumstantial and the relation between potential and absorption intuitive. Conveniently, for polyenes such as retinal, a simple mathematical model of a particle in a box is known to describe the allowed energy levels. In this model, the one-dimensional translation of an electron along the conjugated π -system of a polyene is considered. The length of the π -system is represented by the extension of the box. The absorption maximum can then be derived from the energy difference of the first two energy levels of this electron.

In this work, the influence of the electrostatic potential of the archaeal rhodopsins on the retinal π -system had to be accounted for. Hence, the electrostatic potential of the protein was included as an additional parameter into the model of a particle in a box. The *model of a particle in a box with step potential* allowed to calculate qualitative absorption maxima from the observed electrostatic potential of the archaeal rhodopsins. The calculated differences between the first two energy levels are in excellent agreement with the experimental absorption maxima: the energy difference is nearly identical for bacteriorhodopsin and halorhodopsin, while it is significantly higher for sensory rhodopsin II.

The approach used in this work, *i.e.*, a combination of continuum electrostatic calculations and the quantum mechanical model of a particle in a box with step potential may be applied to other light absorbing proteins such as the visual pigments or the green fluorescent protein. It is possible to generalize the model of a particle in a box to more complex potential functions than a step potential. Then, protein potentials with a complex behavior may be related to the absorption maximum of their chromophore.

Furthermore, the model can be extended to two or three dimensions which are not restricted to box-shaped regions. This will be of interest for applications for example to chlorophyll, *i.e.*, the central photosensitive pigment in photosynthesis. This can be achieved by discretizing the Hamilton Operator in Schrödinger's equation in the geometry of the chromophore. The discretization may be obtained from finite differences. By calculating the electrostatic potential of the protein at each lattice point in the geometry

of the chromophore, this potential can be incorporated in the discretized Hamilton Operator. The allowed energies of a particle in such a system are given by the eigenvalues of the operator.

One approximation of continuum electrostatics is the representation of atoms as point charges thereby neglecting the electron distribution. In principal, the explicit description of electron behavior requires quantum mechanical methods. Quantum mechanical approaches are, however, limited to systems of relatively few atoms. Here, it could be shown that a classical method, namely continuum electrostatics, offers meaningful insights also for excited state processes. Electrostatic interactions of a protein with a chromophore could be qualitatively related to the absorption maximum using classical electrostatic calculations. The analysis of an observed phenomenon can be greatly advanced by applying different methods and only the interplay between these investigations may lead to a true understanding of a phenomenon.

PROTONATION BEHAVIOR AND CORRELATION IN BACTERIORHODOPSIN

The proton transfer of bacteriorhodopsin is one of the most intensely studied processes in the molecular life sciences. For one, this is due to the fundamental importance of ion transport for the metabolism of a cell. Furthermore, bacteriorhodopsin has been proven to be a relatively simple ion pump that can be easily studied. The investigation of bacteriorhodopsin has led to a large amount of information about this protein, its photocycle and its structure. Upon light activation, bacteriorhodopsin undergoes a photocycle with several intermediate states: bR (ground state), K, L, M_1 , M_2 , N and O state. During this photocycle a proton is transferred from the cytoplasm via several key residues of bacteriorhodopsin to the extracellular space. However, open questions remain as to the exact proton transfer path and the chronological order of the transfer events. In this work, continuum electrostatics in combination with a Metropolis Monte Carlo approach was used to characterize the protonation behavior of bacteriorhodopsin.

The protonation behavior of the key residues of proton transfer was analyzed for various photocycle intermediate structures of bacteriorhodopsin. First, the probabilities of functionally important protonation states were presented. In principle, the probabilities of protonation state subsets were calculated, where the protonation of all key residues was specified, while the behavior of the other protonatable residues was not considered. This investigation allowed a classification of the structures with respect to the photocycle intermediate they represent. It could be shown that the intermediate protonation states of the first half of the photocycle, *i.e.*, the bR, K, L and M_1 protonation state, are excellently represented by the available structural data. Since the bR, K and L intermediate assume the same protonation state, the structures well represent two functionally important protonation states: the bR and the M_1 protonation state.

The later photocycle intermediate protonation states, *i.e.*, the M_2 , N and O protonation state, are less well accounted for. At physiological pH values, only an L structure shows

significant probability for the M₂ and the O protonation state. Several M structures and also the single available N structure have a probability for the M₂ protonation state, though only at higher pH values. Additionally at higher pH values, the N structure has a small probability for the N protonation state. Furthermore at physiological pH values, a small probability for the O protonation state could be observed for an M structure. The three protonation states of the later photocycle intermediates, the M₂, N and O protonation state, have a deprotonated proton release group in common. This group consists of two glutamate residues that are located adverse and close to each other near the extracellular side of bacteriorhodopsin. The protonation probabilities of the glutamate residues clearly showed that in most structures at least one proton is bound to these two residues in the physiological pH range. This protonation behavior explains why no significant probability could be observed for the M₂, N and O protonation state at physiological pH values with the exception of the L structure where indeed both glutamate residues are deprotonated.

Secondly, the protonation behavior of the key residues of bacteriorhodopsin was analyzed individually. As already mentioned, the two glutamate residues that constitute the proton release group are partly protonated in most structures. This behavior is not in accordance with the later photocycle intermediate states, the M₂, the N and the O protonation state. However, it could be shown that the calculated protonation behavior is in general in excellent agreement with previous calculations and experimental data.

Finally, the Pearson product-moment correlation coefficient was calculated to assess the correlation of the protonation behavior of two residues. This coefficient is obtained by dividing the covariance of the protonation behavior of two residues by the product of their standard deviations. Strong negative correlation could be observed for several residue pairs, for example Asp85 and the retinal Schiff base. Asp85 is the acceptor for the proton bound to the Schiff base. In several functionally relevant protonation states one proton resides either on Asp85 or on the retinal Schiff base in agreement with negative correlation behavior. Furthermore, the two glutamate residues of the proton release group are in some structures perfectly negatively correlated. This correlation behavior indicated that one glutamate residue is protonated, while the other is deprotonated in the sampled protonation states. Interestingly, some residue pairs are positively correlated, for instance Asp85 and Asp115 in two M structures. A positive correlation indicates that Asp85 and Asp115 behave similarly. Indeed, Asp85 is protonated in the M protonation state, while Asp115 remains protonated throughout the photocycle.

As could be shown, the protonation probabilities often take extreme values of nearly 0 or 1, *i.e.*, the residues are virtually deprotonated or protonated over a large pH range. If the protonation probability of a residue approaches its limit of 0 or 1, the variance of the protonation behavior and consequently the covariance with another residue approaches 0. However, even though the covariance is 0, the correlation coefficient may, nevertheless, take values between -1 and 1. Be assumed that the protonation of the first residue μ is equal to that of the second residue ν , *i.e.*, $x_\mu = x_\nu$. Then, the correlation coefficient is 1. Further assuming that the probabilities approach 0, then, the covariance approaches 0, while the correlation coefficient remains 1. The same holds true for perfectly negatively correlated behavior, *i.e.*, $x_\mu \neq x_\nu$, with probabilities approaching 0 and 1, respectively.

Then, the covariance approaches 0, while the correlation coefficient is -1. Thus, the correlation coefficient may differ substantially from the covariance.

However, in many cases the correlation coefficient is 0 even though correlated behavior would be expected. For example, Asp85 and the retinal Schiff base behave contrarily in the bR, M₁ and M₂ protonation state. Considering only the individual protonation probabilities, perfectly negative correlation would be expected for most structures. In one case, however, the protonation probability of the Schiff base is nearly 0 over the whole pH range, while Asp85 is completely protonated with very few exceptions, when its protonation probability is 0.99. Nevertheless, the correlation coefficient for these two residues is 0 except for the few times when Asp85 is not completely protonated. Then the correlation is indeed close to -1.

As could be shown in Chapter 2 using the example of DTPA, meaningful insight can be gained from the evaluation of the correlation coefficient. However, the probabilities of the considered protonatable sites do not take extreme values, whereas this is commonly the case in bacteriorhodopsin. To understand the inter-relation between the protonation behavior of two residues, the Pearson product-moment correlation coefficient may be too limiting and misleading. A more complete picture of the inter-relation of the protonation behavior of two residues might be gained from calculating the probabilities of the following protonation state subsets: (i) both residues being protonated, (ii) both residues being deprotonated, (iii) one residue being protonated, while the other is deprotonated, and (iv) *vice versa*. By comparing these four probabilities, it can be understood how the two residues behave together as well as individually.

THE X-DEE ALGORITHM

During the investigation of the protonation behavior of bacteriorhodopsin the usefulness of knowing all accessible protonation states of this protein became obvious. Employing Metropolis Monte Carlo, the protonation states were randomly sampled approximating the Boltzmann distribution. This state ensemble is then used to successfully describe the thermodynamic properties of the system. However, due to its concept of random sampling, Metropolis Monte Carlo cannot deterministically search all states accessible to a system. For small systems it is possible to generate a complete list of states using an exhaustive search. However, this approach is not feasible for more complex systems. For example, calculating all protonation states of the relatively small bacteriorhodopsin, *i.e.*, $2^{39} \approx 5 \cdot 10^{11}$ protonation states, where 39 is the number of protonatable residues, far exceeds today's computer resources. For protein structure predictions, where residues can assume several rotamer positions, the complexity of the system multiplies. Furthermore, only a marginal number of all states of a system, *i.e.*, the lowest energy states, are of functional relevance. The effort to calculate all states to gain the states of interest appears disproportionate. No efficient method existed to determine all functionally relevant states of a complex system – a main challenge in structural biology. The perception of the necessity of a complete list of the lowest energy states motivated the development of X-DEE. The theory underlying X-DEE was introduced in Chapter 6 and a first application of this algorithm to bacteriorhodopsin was presented in Chapter 7.

X-DEE allows the determination of a gap-free list of the lowest energy states for a system under conditions that are commonly fulfilled by biological systems. The state energy of the system must be expressible in terms of pairwise interactions between sites, *e.g.*, protein residues. These sites can assume only a finite number of different forms, *e.g.*, protonation form or rotamer position, that differ in their energy contribution. The newly introduced algorithm is based on the dead-end elimination theorem which led to its name X-DEE (extended dead-end elimination). Dead-end elimination provides a criterion to identify the high energy forms of sites that cannot be part of the lowest energy state. These high energy forms, so-called dead-ends, are excluded from the search. By reducing the combinatorial problem to a manageable size, dead-end elimination allows to determine the state of lowest energy.

Based on dead-end elimination, the aim was to determine not only the lowest energy state, but a complete list of succeeding low energy states, one after the other starting from the state of lowest energy. The fundamental idea of X-DEE is the exclusion of certain states from the search space. When for example the lowest energy state is known, this state is excluded from the search space. Consequently, the state of lowest energy determined from this reduced search space is the second lowest energy state of the system. That is, if states are excluded from the search space, always the lowest energy state of those states not excluded will be determined. Proceeding in this manner, with each newly determined state gap-free lists of the lowest energy states are subsequently generated. This can be continued either until a certain number of states is reached or a specific energy is exceeded. Since X-DEE allows to exclude any state or even any list of states from the search space, this algorithm can also be employed to fill the gaps in incomplete lists of states or verify the completeness of others.

The crucial component and the main challenge of the X-DEE algorithm is the creation of a so-called search basis that allows to exclude certain states from the search space. How this objective could be achieved was described in detail in Chapter 6. The properties and the behavior of the developed component and the computational performance of X-DEE were analyzed. Furthermore, the validity of X-DEE was tested using a small system with 2^{21} states. For systems of this size an exhaustive search of all states is feasible. Using X-DEE the states were generated one after another starting from the lowest energy state. The list of states generated by X-DEE was compared to all states calculated by the exhaustive search. It could be shown that X-DEE reproduced all states.

In the present work, X-DEE was implemented to generate gap-free lists of low energy protonation states. The protonation state energy of proteins can be formulated as pairwise interaction between sites and their intrinsic energies. Therefore, the preconditions of the X-DEE algorithm are fulfilled. In Chapter 7, X-DEE was applied to determine the lowest energy protonation states of two different bacteriorhodopsin structures. A gap-free list of lowest energy states was generated in a certain, accessible energy range. The generated states were compared to states sampled by Metropolis Monte Carlo. The energy distribution of the gap-free list of lowest energy states was analyzed. Furthermore, the energy spectrum of functionally relevant protonation states was presented. This information can help to solve the open questions with respect to the kinetics of the photocycle, the interconversion between the different intermediate states and also the chronological

occurrence of conformational changes and the proton transfer events. Answers can be provided using for example differential equations or a dynamical Monte Carlo approach.

The X-DEE algorithm may in future be applied to other proton and ion pumps such as the proteins of the respiratory chain, the photosynthetic reaction centers or also Sodium-Potassium pumps. Dead-end elimination was originally developed to predict protein side chain positions. In protein design, X-DEE may be implemented to determine not only the protein conformation of lowest energy, but several conformations of lowest energy. Since proteins are flexible systems and commonly populate more than one functionally relevant state, the knowledge of all accessible states is of great advantage for the understanding of protein function. In enzymes, the active center often contains protonatable residues which protonation behavior is of significant functional relevance. Thus, the combined search for the lowest energy protein side chain conformations and protonation states will be of particular interest in enzyme design.

BIBLIOGRAPHY

- [1] Tamar Schlick. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer, New York, 2002.
- [2] Andrew R Leach. *Molecular Modelling. Principles and Applications*. Prentice Hall, London, 2001.
- [3] Oren M Becker, Alexander D MacKerell, and Benoit Roux, editors. *Computational Biochemistry and Biophysics*. Marcel Dekker, New York, 2001.
- [4] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, San Diego, 2001.
- [5] Eberhard O Voit. *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*. Cambridge University Press, Cambridge, 2000.
- [6] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, T N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The Protein Data Bank. *Nucleic Acids Res*, 28:235–242, 2000.
- [7] Jade Li, Patricia C Edwards, Manfred Burghammer, Claudio Villa, and Gebhard F X Schertler. Structure of bovine rhodopsin in a trigonal crystal form. *J Mol Biol*, 343:1409–1438, 2004.
- [8] John L Spudich, Chii-Shen Yang, Kwang-Hwan Jung, and Elena N Spudich. Retinylidene proteins: structures and functions from archaea to humans. *Ann Rev Cell Dev Biol*, 16:365–392, 2000.
- [9] Yufeng Zhai, Wilbert H M Heijne, Douglas W Smith, and Milton H Saier Jr. Homologues of archaeal rhodopsins in plants, animals and fungi: structural and functional predications for a putative fungal chaperone protein. *Biochim Biophys Acta*, 1511:206–223, 2001.
- [10] Wolfgang Gärtner and Aba Losi. Crossing the borders: archaeal rhodopsins go bacterial. *Trends Microbiol*, 11:405–407, 2003.
- [11] Akihisa Terakita. The opsins. *Genome Biol*, 6:213, 2005.

- [12] Oded Bèjà, L Aravind, Eugene V Koonin, Marcelino T Suzuki, Andrew Hadd, Linh P Nguyen, Stevan B Jovanovich, Christian M Gates, Robert A Feldman, John L Spudich, Elena N Spudich, and Edward F DeLong. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*, 289:1902–1906, 2000.
- [13] Richard A Krebs, Ulrike Alexiev, Ranga Partha, Anne Marie DeVita, and Mark S Braiman. Detection of fast light-activated H⁺ release and M intermediate formation from proteorhodopsin. *BMC Physiol*, 2:5, 2002.
- [14] Thomas Friedrich, Sven Geibel, Rolf Kalmbach, Igor Chizhov, Kenichi Ataka, Joachim Heberle, Martin Engelhard, and Ernst Bamberg. Proteorhodopsin is a light-driven proton pump with variable vectoriality. *J Mol Biol*, 321:821–838, 2002.
- [15] Oded Bèjà, Elena N Spudich, John L Spudich, Marion Leclerc, and Edward F DeLong. Proteorhodopsin phototrophy in the ocean. *Nature*, 411:786–789, 2001.
- [16] Jose R de la Torre, Lynne M Christianson, Oded Bèjà, Marcelino T Suzuki, David M Karl, John Heidelberg, and Edward F DeLong. Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc Natl Acad Sci USA*, 100:12830–12835, 2003.
- [17] Sergei P Balashov, Eleonora S Imasheva, Vladimir A Boichenko, Josefa Antón, Jennifer M Wang, and Janos K Lanyi. Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. *Science*, 309:2061–2064, 2005.
- [18] Stephen A Waschuk, Arandi G Bezerra Jr., Lichi Shi, and Leonid S Brown. *Leptospaeria* rhodopsin: bacteriorhodopsin-like proton pump from a eukaryote. *Proc Natl Acad Sci USA*, 102:6879–6883, 2005.
- [19] Georg Nagel, Tanjef Szellas, Suneel Kateriya, Nona Adeishvili, Peter Hegemann, and Ernst Bamberg. Channelrhodopsins: directly light-gated cation channels. *Biochem Soc Trans*, 33:863–866, 2005.
- [20] Slawomir Filipek, Ronald E Stenkamp, David C Teller, and Krzysztof Palczewski. G protein-coupled receptor rhodopsin: a prospectus. *Annu Rev Physiol*, 65:851–879, 2003.
- [21] Rolf Herrmann, Martin Heck, Petra Henklein, Peter Henklein, Christiane Kleuss, Klaus Peter Hofmann, and Oliver P Ernst. Sequence of interactions in receptor-G protein coupling. *J Biol Chem*, 279:24283–24290, 2004.
- [22] Dimitrios Fotiadis, Beata Jastrzebska, Ansgar Philippsen, Daniel J Müller, Krzysztof Palczewski, and Andreas Engel. Structure of the rhodopsin dimer: a working model for G-protein-coupled receptors. *Curr Opin Struct Biol*, 16:252–259, 2006.
- [23] Arlene D Albert and Philip L Yeagle. Structural studies on rhodopsin. *Biochim Biophys Acta*, 1565:183–195, 2002.
- [24] Tetsuji Okada, Krzysztof Palczewski, Oliver P Ernst, and Klaus Peter Hofmann. Activation of rhodopsin: new insights from structural and biochemical studies. *Trends Biochem Sci*, 26:318–324, 2001.

- [25] Krzysztof Palczewski. G protein-coupled receptor rhodopsin. *Annu Rev Biochem*, 75:743–767, 2006.
- [26] Kwang-Hwan Jung, Vishwa D Trivedi, and John L Spudich. Demonstration of a sensory rhodopsin in eubacteria. *Mol Microbiol*, 47:1513–1522, 2003.
- [27] Lutz Vogeley, Oleg A Sineshchekov, Vishwa D Trivedi, Jun Sasaki, John L Spudich, and Hartmut Luecke. Anabaena sensory rhodopsin: a photochromic color sensor at 2.0 Å. *Science*, 306:1390–1393, 2004.
- [28] Werner Deininger, Petra Kröger, Utah Hegemann, Friedrich Lottspeich, and Peter Hegemann. Chlamyrorhodopsin represents a new type of sensory photoreceptor. *EMBO J*, 14:5849–5858, 1995.
- [29] Tida Kumbalasingam and Ignacio Provencio. Melanopsin and other novel mammalian opsins. *Exp Eye Res*, 81:368–375, 2005.
- [30] John L Spudich. The multitasking microbial sensory rhodopsins. *Trends Microbiol*, 14:480–487, 2006.
- [31] Adrian K Sharma, John L Spudich, and W Ford Doolittle. Microbial rhodopsins: functional versatility and genetic mobility. *Trends Microbiol*, 14:463–469, 2006.
- [32] Michael T Madigan and John Martinko. *Brock Biology of Microorganisms*. Prentice Hall International, London, 2005.
- [33] Michael T Madigan and Aharon Oren. Thermophilic and halophilic extremophiles. *Curr Opin Microbiol*, 2:265–269, 1999.
- [34] Aharon Oren. Bioenergetic aspects of halophilism. *Microbiol Mol Biol Rev*, 63:334–348, 1999.
- [35] Günter Schäfer, Martin Engelhardt, and Volker Müller. Bioenergetics of the archaea. *Microbiol Mol Biol Rev*, 63:570–620, 1999.
- [36] Dieter Oesterhelt and Walter Stoeckenius. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nature*, 233:149–152, 1971.
- [37] Brigitte Schobert and Janos K Lanyi. Halorhodopsin is a light-driven chloride pump. *J Biol Chem*, 257:10306–10313, 1982.
- [38] Roberto A Bogomolni and John L Spudich. Identification of a third rhodopsin-like pigment in phototactic *Halobacterium halobium*. *Proc Natl Acad Sci USA*, 79:6250–6254, 1982.
- [39] Tetsuo Takahashi, Hiroaki Tomioka, Naoki Kamo, and Yonosuke Kobatake. A photosystem other than PS370 also mediates the negative phototaxis of *Halobacterium halobium*. *FEMS Microbiol Lett*, 28:161–164, 1985.
- [40] Elmar K Wolff, Roberto A Bogomolni, Peter Scherrer, Benno Hess, and Walter Stoeckenius. Color discrimination in halobacteria: spectroscopic characterization of a second sensory receptor covering the blue-green region of the spectrum. *Proc Natl Acad Sci USA*, 83:7272–7276, 1986.

- [41] Hassan Belrhali, Peter Nollert, Antoine Royant, Christoph Menzel, Jürg P Rosenbusch, Ehud M Landau, and Eva Pebay-Peyroula. Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution. *Structure*, 7:909–917, 1999.
- [42] Hartmut Luecke, Brigitte Schobert, Hans-Thomas Richter, Jean-Philippe Cartailler, and Janos K Lanyi. Structure of bacteriorhodopsin at 1.55 Å resolution. *J Mol Biol*, 291:899–911, 1999.
- [43] Marc T Facciotti, Sahab Rouhani, Frederick T Burkard, Felicia M Betancourt, Kenneth H Downing, Robert B Rose, Gerry McDermott, and Robert M Glaeser. Structure of an early intermediate in the M-state phase of the bacteriorhodopsin photocycle. *Biophys J*, 81:3442–3455, 2001.
- [44] Kazuki Takeda, Yasuhiro Matsui, Nobuo Kamiya, Shin-ichi Adachi, Hideo Okumura, and Tsutomu Kouyama. Crystal structure of the M intermediate of bacteriorhodopsin: allosteric structural changes mediated by sliding movement of a transmembrane helix. *J Mol Biol*, 341:1023–1037, 2004.
- [45] Brigitte Schobert, Leonid S Brown, and Jano K Lanyi. Crystallographic structures of the M and N intermediates of bacteriorhodopsin: assembly of a hydrogen-bonded chain of water molecules between Asp-96 and the retinal Schiff base. *J Mol Biol*, 330:553–570, 2003.
- [46] Michael Kolbe, Hüseyin Besir, Lars-Oliver Essen, and Dieter Oesterhelt. Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. *Science*, 288:1390–1396, 2000.
- [47] Walter Gmelin, Kornelius Zeth, Ruslan Efremov, Joachim Heberle, Jörg Tittor, and Dieter Oesterhelt. The crystal structure of the L1 intermediate of halorhodopsin at 1.9 Å resolution. *Photochem Photobiol*, 83:369–377, 2007.
- [48] Hartmut Luecke, Brigitte Schobert, Janos K Lanyi, Elena N Spudich, and John L Spudich. Crystal structure of sensory rhodopsin II at 2.4 ångströms: insights into color tuning and transducer interaction. *Science*, 293:1499–1502, 2001.
- [49] Antoine Royant, Peter Nollert, Karl Edman, Richard Neutze, Ehud M Landau, and Eva Pebay-Peyroula. X-ray structure of sensory rhodopsin II at 2.1 Å resolution. *Proc Natl Acad Sci USA*, 98:10131–10136, 2001.
- [50] Valentin I Gordeliy, Jörg Labahn, Rouslan Moukhametzianov, Rouslan Efremov, Joachim Granzin, Ramona Schlesinger, Georg Büldt, Tudor Savopol, Axel J Scheidig, Johann P Klare, and Martin Engelhard. Molecular basis of transmembrane signaling by sensory rhodopsin II-transducer complex. *Nature*, 419:484–487, 2002.
- [51] Yasuo Mukohata, Kunio Ihara, Takeshi Tamura, and Yasuo Sugiyama. Halobacterial rhodopsins. *J Biochem*, 125:649–657, 1999.
- [52] Li Tang, Thomas G Ebrey, and Shankar Subramaniam. Sequences and structures of retinal proteins. *Isr J Chem*, 35:193–209, 1995.

- [53] David L Nelson and Michael M Cox. *Lehninger Principles of Biochemistry*. W H Freeman, New York, 2004.
- [54] Richard A Mathies, Steven W Lin, James B Ames, and W Thomas Pollard. From femtoseconds to biology: mechanism of bacteriorhodopsin's light-driven proton pump. *Annu Rev Biophys Biophys Chem*, 20:491–518, 1991.
- [55] Janos K Lanyi. Proton translocation mechanism and energetics in the light-driven pump bacteriorhodopsin. *Biochim Biophys Acta*, 1183:241–261, 1993.
- [56] Ulrich Haupts, Jörg Tittor, and Dieter Oesterhelt. Closing in on bacteriorhodopsin: progress in understanding the molecule. *Annu Rev Biophys Biomol Struct*, 28:367–399, 1999.
- [57] Joachim Heberle, Jörg Fitter, Hans Jürgen Sass, and Georg Büldt. Bacteriorhodopsin: the functional details of a molecular machine are being resolved. *Biophys Chem*, 85:229–248, 2000.
- [58] Norbert A Dencher, Hans Jürgen Sass, and Georg Büldt. Water and bacteriorhodopsin: structure, dynamics, and function. *Biochim Biophys Acta*, 1460:192–203, 2002.
- [59] Janos K Lanyi and Hartmut Luecke. Bacteriorhodopsin. *Curr Opin Stru Biol*, 11: 415–419, 2001.
- [60] Richard Neutze, Eva Pebay-Peyroula, Karl Edman, Antoine Royant, Javier Navarro, and Ehud M Landau. Bacteriorhodopsin: a high-resolution structural view of vectorial proton transport. *Biochim Biophys Acta*, 1565:144–167, 2002.
- [61] Janos K Lanyi. Proton transfers in the bacteriorhodopsin photocycle. *Biochim Biophys Acta*, 1757:1012–1018, 2006.
- [62] Robin Rammelsberg, Gregor Huhn, Mathias Lübben, and Klaus Gerwert. Bacteriorhodopsin's intramolecular proton-release pathway consists of a hydrogen-bonded network. *Biochemistry*, 37:5001–5009, 1998.
- [63] Florian Garczarek, Leonid S Brown, Janos K Lanyi, and Klaus Gerwert. Proton binding within a membrane protein by a protonated water cluster. *Proc Natl Acad Sci USA*, 102:3633–3638, 2005.
- [64] Nicolas Calimet and G Matthias Ullmann. The influence of a transmembrane pH gradient on protonation probabilities of bacteriorhodopsin: the structural basis of the back-pressure effect. *J Mol Biol*, 339:571–589, 2004.
- [65] Elisa Bombarda, Torsten Becker, and G Matthias Ullmann. Influence of the membrane potential on the protonation of bacteriorhodopsin: insights from electrostatic calculations into the regulation of proton pumping. *J Am Chem Soc*, 128:12129–12139, 2006.
- [66] N Friedman, Y Gat, Mordechai Sheves, and Michael Ottolenghi. On the heterogeneity of the M population in the photocycle of bacteriorhodopsin. *Biochemistry*, 33: 14758–14767, 1994.

- [67] Wolf Einfeld, Thomas Althaus, and Manfred Stockburger. Evidence for parallel photocycles and implications for the proton pump in bacteriorhodopsin. *Biophys Chem*, 56:105–112, 1995.
- [68] Krisztina Ludmann, Csilla Gergely, and Gyögy Váró. Kinetic and thermodynamic study of the bacteriorhodopsin photocycle over a wide pH range. *Biophys J*, 75:3110–3119, 1998.
- [69] Richard W Hendler, Richard I Shrager, and Salil Bose. Theory and procedures for finding a correct kinetic model for the bacteriorhodopsin photocycle. *J Phys Chem B*, 105:3319–3328, 2001.
- [70] Richard W Hendler and Salil Bose. Interconversions among four M-intermediates in the bacteriorhodopsin photocycle. *Eur J Biochem*, 270:3518–2354, 2003.
- [71] A Lukács and Elmér Papp. Bacteriorhodopsin photocycle kinetics analyzed by the maximum entropy method. *J Photochem Photobiol B*, 77:1–16, 2004.
- [72] Richard W Hendler. An apparent general solution for the kinetic models of the bacteriorhodopsin photocycles. *J Phys Chem B*, 109:16515–16528, 2005.
- [73] Richard H Lozier and John F Nagle. Comment on "An apparent general solution for the kinetic models of the bacteriorhodopsin photocycles". *J Phys Chem B*, 110:15041–2, 2006.
- [74] Richard W Hendler. Reply to "comment on 'an apparent general solution for the kinetic models of the bacteriorhodopsin photocycle' ". *J Phys Chem B*, 110:15043–15044, 2006.
- [75] Jérôme Baudry, Serge Crouzy, Benoît Roux, and Jeremy C Smith. Simulation analysis of the retinal conformational equilibrium in dark-adapted bacteriorhodopsin. *Biophys J*, 76:1909–1917, 1999.
- [76] Jérôme Baudry, Emad Tajkhorshid, Ferenc Molnar, James Phillips, and Klaus Schulten. Molecular dynamics study of bacteriorhodopsin and the purple membrane. *J Phys Chem B*, 105:905–918, 2001.
- [77] Alexey Onufriev, Alexander Smondryev, and Donald Bashford. Proton affinity changes driving the unidirectional proton transport in the bacteriorhodopsin photocycle. *J Mol Biol*, 332:1183–1193, 2003.
- [78] Antonio M Ferreira and Donald 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.
- [79] Volker Müller and Aharon Oren. Metabolism of chloride in halophilic prokaryotes. *Extremophiles*, 7:261–266, 2003.
- [80] Dieter Oesterhelt. Structure and function of halorhodopsin. *Isr J Chem*, 35:475–494, 1995.

- [81] Lars-Oliver Essen. Halorhodopsin: light-driven ion pumping made simple? *Curr Opin Struct Biol*, 12:516–522, 2002.
- [82] Andreea D Gruia, Ana-Nicoleta Bondar, Jeremy C Smith, and Stefan Fischer. Mechanism of a molecular valve in the halorhodopsin chloride pump. *Structure*, 13: 617–627, 2005.
- [83] Ernst Bamberg, Jörg Tittor, and Dieter Oesterhelt. Light-driven proton or chloride pumping by halorhodopsin. *Proc Natl Acad Sci USA*, 90:639–643, 1993.
- [84] Gyögy Váró, Leonid S Brown, Richard Needleman, and Janos K Lanyi. Proton transport by halorhodopsin. *Biochemistry*, 35:6604–6611, 1996.
- [85] Jun Sasaki and John L Spudich. Proton transport by sensory rhodopsins and its modulation by transducer-binding. *Biochim Biophys Acta*, 1460:230–239, 2000.
- [86] Johann P Klare, Valentin I Gordeliy, Jörg Labahn, Georg Büldt, Heinz-Jürgen Steinhoff, and Martin Engelhard. The archaeal sensory rhodopsin II/transducer complex: a model for transmembrane signal transfer. *FEBS Lett*, 564:219–224, 2004.
- [87] Jun Hirayama, Yasushi Imamoto, Yoshinori Shichida, Naoki Kamo, Hiroaki Tomioka, and Tôru Yoshizawa. Photocycle of phoborhodopsin from haloalkaliphilic bacterium (*Natronobacterium pharaonis*) studied by low-temperature spectrophotometry. *Biochemistry*, 31:2093–2098, 1992.
- [88] Ulrich Haupts, Jörg Tittor, Ernst Bamberg, and Dieter Oesterhelt. General concept for ion translocation by halobacterial retinal proteins: the isomerization/switch/transfer (IST) model. *Biochemistry*, 36:2–7, 1997.
- [89] John L Spudich. Variations on a molecular switch: transport and sensory signaling by archaeal rhodopsins. *Mol Microbiol*, 28:1051–1058, 1998.
- [90] Dieter Oesterhelt, Jörg Tittor, and Ernst Bamberg. A unifying concept for ion translocation by retinal proteins. *J Bioenerg Biomembr*, 24:181–191, 1992.
- [91] Walter Stoeckenius. Bacterial rhodopsins: evolution of a mechanistic model for the ion pumps. *Prot Sci*, 8:447–459, 1999.
- [92] Gyögy Váró. Analogies between halorhodopsin and bacteriorhodopsin. *Biochim Biophys Acta*, 1460:220–229, 2000.
- [93] Jun Sasaki, Leonid S Brown, Young-Shin Chon, Hideki Kandori, Akio Maeda, Richard Needleman, and Janos K Lanyi. Conversion of bacteriorhodopsin into a chloride ion pump. *Science*, 269:73–75, 1995.
- [94] Jörg Titor, Ulrich Haupts, Christina Haupts, Dieter Oesterhelt, Anja Becker, and Ernst Bamberg. Chloride and proton transport in bacteriorhodopsin mutant D85T: different modes of ion translocation in a retinal protein. *J Mol Biol*, 271:405–416, 1997.

- [95] Jun Sasaki and John L Spudich. Proton circulation during the photocycle of sensory rhodopsin II. *Biophys J*, 77:2145–5212, 1999.
- [96] James O Erickson and Paul E Blatz. N-retinylidene-1-amino-2-propanol: a Schiff base analog for rhodopsin. *Vision Res*, 8:1367–1375, 1968.
- [97] Gerd G Kochendoerfer, Steven W Lin, Thomas P Sakmar, and Richard A Mathies. How color visual pigments are tuned. *Trends Biochem Sci*, 24:300–305, 1999.
- [98] Michael A van der Horst and Klaas J Hellinger. Photoreceptor proteins, "star actors of modern times": a review of the functional dynamics in the structure of representative members of six different photoreceptor families. *Acc Chem Res*, 37:13–20, 2004.
- [99] Koji Nakanishi, Valeria Balogh-Nair, Maria Arnaboldi, Kazuo Tsujimoto, and Barry Honig. An external point-charge model for bacteriorhodopsin to account for its purple color. *J Am Chem Soc*, 102:7945–7947, 1980.
- [100] Dikla Man, Weiwu Wang, Gazalah Sabehi, L Aravind, Anton F Post, Ramon Masana, Elena N Spudich, John L Spudich, and Oded Bèjà. Diversification and spectral tuning in marine proteorhodopsins. *EMBO J*, 22:1725–1731, 2003.
- [101] Chmouel R Goldschmidt, Michael Ottolenghi, and Rafi Korenstein. On the primary quantum yields in the bacteriorhodopsin photocycle. *Biophys J*, 16:839–843, 1976.
- [102] John L Spudich and Roberto A Bogomolni. Spectroscopic discrimination of the three rhodopsinlike *Halobacterium halobium* membranes. *Biophys J*, 43:243–246, 1983.
- [103] Lars H Andersen, Iben B Nielsen, Michael B Kristensen, Mohamed O A El Ghazaly, Stefan Haacke, Mogens Brndsted Nielsen, and Michael Åxman Petersen. Absorption of Schiff-base retinal chromophores *in vacuo*. *J Am Chem Soc*, 127:12347–12350, 2005.
- [104] Peter W Atkins and Julio de Paula. *Atkins' Physical Chemistry*. Oxford University Press, Oxford, 2006.
- [105] Jochen Kleinschmidt and Ferenc I Harosi. Anion sensitivity and spectral tuning of cone visual pigments *in situ*. *Proc Natl Acad Sci USA*, 89:9181–9185, 1992.
- [106] Steven W Lin, Gerd G Kochendoerfer, Kate S Carroll, Dorothy Wang, Richard A Mathies, and Thomas P Sakmar. Mechanisms of spectral tuning in the blue cone visual pigments. *J Biol Chem*, 273:24583–24591, 1998.
- [107] John L Spudich, Donald A McCain, Koji Nakanishi, Masami Okabe, Nobuko Shimizu, Hillary Rodman, Barry Honig, and Roberto A Bogomolni. Chromophore/protein interaction in bacterial sensory rhodopsin and bacteriorhodopsin. *Biophys J*, 49:479–483, 1986.
- [108] Tetsuo Takahashi, Bing Yan, Paul Mazur, Fadila Derguini, Koji Nakanishi, and John L Spudich. Color regulation in the archaeobacterial phototaxis receptor phoborhodopsin (sensory rhodopsin II). *Biochemistry*, 29:8467–8474, 1990.

- [109] Kazumi Shimono, Masayuki Iwamoto, Masato Sumi, and Naoki Kamo. Effects of three characteristic amino acid residues of *Pharaonis* phoborhodopsin on the absorption maximum. *Photochem Photobiol*, 72:141–145, 2000.
- [110] Ramkumar Rajamani and Jiali Gao. Combined QM/MM study of the opsin shift in bacteriorhodopsin. *J Comp Chem*, 23:96–105, 2002.
- [111] Lei Ren, Charles H Martin, Kevin J Wise, Nathan B Gillespie, Hartmut Luecke, Janos K Lanyi, John L Spudich, and Robert R Birge. Molecular mechanism of spectral tuning in sensory rhodopsin II. *Biochemistry*, 40:13906–13914, 2001.
- [112] Shigehiko Hayashi, Emad Tajkhorshid, Eva Pebay-Peyroula, Antoine Royant, Ehud M Landau, Javier Navarro, and Klaus Schulten. Structural determinants of spectral tuning in retinal proteins – bacteriorhodopsin *vs.* sensory rhodopsin II. *J Phys Chem B*, 105:10124–10131, 2001.
- [113] David G Nicholls and Stuart J Ferguson. *Bioenergetics 3*. Academic Press, London, 2002.
- [114] Kim A Sharp and Barry Honig. Electrostatic interactions in macromolecules: theory and applications. *Annu Rev Biophys Biophys Chem*, 19:301–332, 1990.
- [115] Barry Honig and Anthony Nicholls. Classical electrostatics in biology and chemistry. *Science*, 268:1144–1149, 1995.
- [116] Haruki Nakamura. Roles of electrostatic interaction in proteins. *Q Rev Biophys*, 29:1–90, 1996.
- [117] Arieh Warshel and Arno Papazyan. Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Curr Opin Stru Biol*, 8:211–217, 1998.
- [118] Patrice Koehl. Electrostatics calculations: latest methodological advances. *Curr Opin Stru Biol*, 16:142–151, 2006.
- [119] Durba Sengupta, Lars Meinhold, Dieter Langosch, G Matthias Ullmann, and Jeremy C Smith. Energetics of helical-peptide orientations in membranes. *Proteins*, 58:913 – 922, 2005.
- [120] Federico Fogolari, Alessandro Brigo, and Henriette Molinari. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J Mol Recognit*, 15:377–392, 2002.
- [121] John David Jackson. *Classical Electrodynamics*. John Wiley & Sons Inc, New York, 1998.
- [122] Stephen C Harvey. Treatment of electrostatic effects in macromolecular modeling. *Proteins*, 5:78–92, 1989.
- [123] Jan Antosiewicz, James M Briggs, Adrian H Elcock, Michael K Gilson, and J Andrew McCammon. Computing ionization states of protein with a detailed charge model. *J Comput Chem*, 17:1633–1644, 1996.

- [124] Federico Fogolari, Pierfrancesco Zuccato, Gennaro Esposito, and Paolo Viglino. Biomolecular electrostatics with the linearized Poisson-Boltzmann equation. *Biophys J*, 76:1–16, 1999.
- [125] Donald Bashford and Klaus Gerwert. Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. *J Mol Biol*, 224:473–486, 1992.
- [126] G Matthias Ullmann and Ernst-Walter Knapp. Electrostatic models for computing protonation and redox equilibria in proteins. *Eur Biophys*, 28:533–551, 1999.
- [127] G Matthias Ullmann. Relations between protonation constants and titration curves in polyprotic acids: a critical view. *J Phys Chem B*, 107:1263–1271, 2003.
- [128] Charles Tanford. The interpretation of hydrogen ion titration curves of proteins. *Advan Prot Chem*, 17:69–165, 1962.
- [129] Jan Antosiewicz, J Andrew McCammon, and Michael K Gilson. Prediction of pH-dependent properties of proteins. *J Mol Biol*, 238:415–436, 1994.
- [130] William H Richardson, C-Y Peng, Donald Bashford, Louis Noodleman, and David A Case. Incorporating solvation effects into density functional theory: calculation of absolute acidities. *Int J Quant Chem*, 61:207–217, 1997.
- [131] Ilja N Bronstein and Konstantin A Semendjajew. *Momente einer mehrdimensionalen Zufallsgröße*. In Taschenbuch der Mathematik. Teubner, Stuttgart/Leipzig, 1991, p 671.
- [132] Nicholas Metropolis and Stanislaw Ulam. The Monte Carlo method. *J Amer Stat Assoc*, 44:335–341, 1949.
- [133] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing. *J Phys Chem*, 21:1087–1092, 1953.
- [134] Michael Shirts and Vijay S Pande. Screen savers of the world unite! *Science*, 290:1903–1904, 2000.
- [135] Vijay S Pande and Stanford University. Folding@home distributed computing, <http://folding.stanford.edu/English/Main>, 2000.
- [136] Bojan Zagrovic, Eric J Sorin, and Vijay Pande. β -hairpin folding simulations in atomistic detail using an implicit solvent model. *J Mol Biol*, 313:151–169, 2001.
- [137] Lillian T Chonga, William C Swope, Jed W Piterab, and Vijay S Pande. Kinetic Computational Alanine Scanning: Application to p53 Oligomerization. *J Mol Biol*, 357:1039–1049, 2006.
- [138] David Baker and the University of Washington. Rosetta@home protein folding, design, and docking, <http://boinc.bakerlab.org/rosetta/>, 2005.
- [139] Brian Kuhlman, Gautam Dantas, Gregory C Ireton, Gabriele Varani, Barry L Stoddard, and David Baker. Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302:1364–1368, 2003.

- [140] Krzyszto Ginalski. Comparative modeling for protein structure prediction. *Curr Opin Struct Biol*, 16:172–177, 2006.
- [141] Judith C Madden and Mark T D Cronin. Structure-based methods for the prediction of drug metabolism. *Expert Opin Drug Metab Toxicol*, 2:545–557, 2006.
- [142] Kurt Wüthrich. *NMR of Proteins and Nucleic Acids*. Wiley, New York, 1986.
- [143] John C Kendrew, G Bodo, Howard M Dintzis, R G Parrish, H Wyckoff, and D C Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181:662–666, 1958.
- [144] Max F Perutz, M G Rossmann, A F Cullis, H Muirhead, G Will, and A C T North. Structure of haemoglobin: a three-dimensional fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature*, 185:416–422, 1960.
- [145] Carl Branden and John Tooze. *Introduction to Protein Structure*. Garland Publishing, New York, 1999.
- [146] Johann Deisenhofer, Otto Epp, K Miki, Robert Huber, and Hartmut Michel. X-ray structure analysis of a membrane protein complex. Electron density map at 3 Å resolution and a model of the chromophores of the photosynthetic reaction center from *Rhodospseudomonas viridis*. *J Mol Biol*, 180:385–398, 1984.
- [147] Ehud M Landau and Jürg P Rosenbusch. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc Natl Acad Sci USA*, 93:14532–14535, 1996.
- [148] Mark L Chiu, Peter Nollert, Michèle C Loewen, Hassan Belrhali, Eva Pebay-Peyroula, Jürg P Rosenbuscha, and Ehud M Landau. Crystallization *in cubo*: general applicability to membrane proteins. *Acta Crystallogr D Biol Crystallogr*, 56:781–784, 2000.
- [149] Yasser Qutub, Ilya Reviakine, Carrie Maxwell, Javier Navarro, Ehud M Landau, and Peter G Vekilov. Crystallization of transmembrane proteins *in cubo*: mechanisms of crystal growth and defect formation. *J Mol Biol*, 343:1243–1254, 2004.
- [150] Hartmut Luecke, Brigitte Schobert, Jean-Philippe Cartailler, Hans-Thomas Richter, Anja Rosengarth, Richard Needleman, and Janos K Lanyi. Coupling photoisomerization of retinal to directional transport in bacteriorhodopsin. *J Mol Biol*, 300:1237–1255, 2000.
- [151] Yasuhiro Matsui, Keisuke Sakai, Midori Murakami, Yoshitsugo Shiro, Shin-ichi Adachi, Hideo Okumura, and Tsutomu Kouyama. Specific damage induced by X-ray radiation and structural changes in the primary photoreaction of bacteriorhodopsin. *J Mol Biol*, 324:469–481, 2002.
- [152] Brigitte Schobert, Jill Cupp-Vickery, Victor Hornak, Steven O Smith, and Janos K Lanyi. Crystallographic structure of the K intermediate of bacteriorhodopsin: conservation of free energy after photoisomerization of the retinal. *J Mol Biol*, 321:715–726, 2002.

- [153] Karl Edman, Peter Nollert, Antoine Royant, Hassan Belrhali, Eva Pebay-Peyroula, Jonas Hajdu, Richard Neutze, and Ehud M Landau. High-resolution X-ray structure of an early intermediate in the bacteriorhodopsin photocycle. *Nature*, 401: 822–826, 1999.
- [154] Antoine Royant, Karl Edman, Thomas Ursby, Eva Pebay-Peyroula, Ehud M Landau, and Richard Neutze. Helix deformation is coupled to vectorial proton transport in the photocycle of bacteriorhodopsin. *Nature*, 406:645–648, 2000.
- [155] Janos K Lanyi and Brigitte Schobert. Mechanism of proton transport in bacteriorhodopsin from crystallographic structures of the K, L, M₁, M₂, and M₂' intermediates of the photocycle. *J Mol Biol*, 328:439–450, 2003.
- [156] Karl Edman, Antoine Royant, Gisela Larsson, Frida Jacobson, Tom Taylor, David van der Spoel, Ehud M Landau, Eva Pebay-Peyroula, and Richard Neutze. Deformation of helix C in the low temperature L-intermediate of bacteriorhodopsin. *J Biol Chem*, 279:2147–2158, 2004.
- [157] Hans Jürgen Sass, Georg Büldt, Ralf Gessenich, Dominic Hehn, Dirk Neff, Ramona Schlesinger, Joel Berendzen, and Par Ormos. Structural alterations for proton translocation in the M state of wild-type bacteriorhodopsin. *Nature*, 406:649–653, 2000.
- [158] Janos K Lanyi and Brigitte Schobert. Crystallographic structure of the retinal and the protein after deprotonation of the Schiff base: the switch in the bacteriorhodopsin photocycle. *J Mol Biol*, 321:727–727, 2002.
- [159] Sahab Rouhani, Jean-Philippe Cartailier, Marc T Faciotti, Peter Walian, Richard Needleman, Janos K Lanyi, Robert M Glaeser, and Hartmut Luecke. Crystal structure of the D85S mutant of bacteriorhodopsin: model of an o-like photocycle intermediate. *J Mol Biol*, 313:615–628, 2001.
- [160] Robert M Glaeser, J S Jubb, and Richard Henderson. Structural comparison of native and deoxycholate-treated purple membrane. *Biophys J*, 48:775–780, 1985.
- [161] Nikolaus Grigorieff, Thomas A Ceska, Kenneth H Downing, Joyce M Baldwin, and Richard Henderson. Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J Mol Biol*, 259:393–421, 1996.
- [162] Bernard R Brooks, Robert E Bruccoleri, Berry D Olafson, David J States, S Swaminathan, and Martin Karplus. CHARMM: a program for macromolecular energy, minimization, and dynamics calculation. *J Comput Chem*, 4:187–217, 1983.
- [163] A D MacKerell, D Bashford, M Bellott, R L Dunbrack, J D Evanseck, M J Field, S Fischer, J Gao, H Guo, S Ha, D Joseph-McCarthy, L Kuchnir, K Kuczera, F T K Lau, C Mattos, S Michnick, T Ngo, D T Nguyen, B Prodhom, W E Reiher, B Roux, M Schlenkrich, J C Smith, R Stote, J Straub, M Watanabe, J Wiorcikiewicz-Kuczera, D Yin, and M Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*, 102:3586–3616, 1998.

- [164] Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Cryst A*, 32:922–923, 1976.
- [165] Yoshiaki Kimura, Dmitry G Vassylyev, Atsuo Miyazawa, Akinori Kidera, Masaaki Matsushima, Kaoru Mitsuoka, Kazuyoshi Murata, Teruhisa Hirai, and YYoshinori Fujiyoshi. Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature*, 389:206–211, 1997.
- [166] Velin Z Spassov, Hartmut Luecke, Klaus Gerwert, and Donald Bashford. pK_a calculations suggest storage of an excess proton in a hydrogen-bonded water network in bacteriorhodopsin. *J Mol Biol*, 312:203–219, 2001.
- [167] A Bondi. Van der Waals volumes and radii. *J Phys Chem*, 68:441–451, 1964.
- [168] Donald Bashford. *An object-oriented programming suite for electrostatic effects in biological molecules*. In Ishikawa, Y, Oldehoeft, R R, Reynders J V W, Tholburn, M, (Eds.), *Scientific Computing in Object-Oriented Parallel Environments*. Springer, Berlin, 1997, pp 233-240.
- [169] J Schwemer and H Langer. Insect visual pigments. *Meth Enzymol*, 81:182–190, 1982.
- [170] Arieh Warshel and Michael Ottolenghi. Kinetic and spectroscopic effects of protein-chromophore electrostatic interactions in bacteriorhodopsin. *Photochem Photobiol*, 30:291–293, 1979.
- [171] Hirohiko Houjou, Yoshio Inoue, and Minoru Sakurai. Physical origin of the opsin shift of bacteriorhodopsin. Comprehensive analysis based on medium effect theory of absorption spectra. *J Am Chem Soc*, 120:4459–4470, 1998.
- [172] Hirohiko Houjou, Yoshio Inoue, and Minoru Sakurai. Study of the opsin shift of bacteriorhodopsin: insight from QM/MM calculations with electronic polarization effects of the protein environment. *J Phys Chem B*, 2001:867–879, 2001.
- [173] Arieh Warshel and Z T Chu. Nature of the surface crossing process in bacteriorhodopsin: computer simulations of the quantum dynamics of the primary photochemical event. *J Phys Chem B*, 105:9857–9871, 2001.
- [174] Minoru Sugihara, Volker Buß, Peter Entel, Marcus Elstner, and Thomas Frauenheim. 11-*cis*-retinal protonated Schiff base: influence of the protein environment on the geometry of the rhodopsin chromophore. *Biochemistry*, 41:15259–15266, 2002.
- [175] Minoru Sakurai, Keiko Sakata, Shino Saito, Sawako Nakajima, and Yoshio Inoue. Decisive role of electronic polarization of the protein environment in determining the absorption maximum of halorhodopsin. *J Am Chem Soc*, 125:3108–3112, 2003.
- [176] Marko Schreiber, Minoru Sugihara, and Volker Buß. Exploring the opsin shift with ab initio methods: geometry and counterion effects on the electronic spectrum of retinal. *J Chem Phys*, 119:12045–12048, 2003.

- [177] Marius Wanko, Michael Hoffmann, Paul Strodel, Walter Thiel, Frank Neese, Thomas Frauenheim, and Marcus Elstner. Calculating absorption shifts for retinal proteins: computational challenges. *J Phys Chem B*, 109:3606–3615, 2005.
- [178] Michael Hoffmann, Marius Wanko, Paul Strodel, Peter H König, Thomas Frauenheim, Klaus Schulten, Walter Thiel, Emad Tajkhorshid, and Marcus Elstner. Color tuning in rhodopsins: the mechanism for the spectral shift between bacteriorhodopsin and sensory rhodopsin II. *J Am Chem Soc*, 128:10808–10818, 2006.
- [179] Jingui Hu, Robert G Griffin, and Judith Herzfeld. Synergy in the spectral tuning of retinal pigments: complete accounting of the opsin shift in bacteriorhodopsin. *Proc Natl Acad Sci USA*, 91:8880–8884, 1994.
- [180] Kazumi Shimono, Yukako Ikeura, Yuki Sudo, Masayuki Iwamoto, and Naoki Kamo. Environment around the chromophore in *pharaonis* phoborhodopsin: mutation analysis of the retinal binding site. *Biochim Biophys Acta*, 1515:92–100, 2001.
- [181] Juliet W Parry, Subathra Poopalasundaram, James K Bowmaker, and David M Hunt. A novel amino acid substitution is responsible for spectral tuning in a rodent violet-sensitive visual pigment. *Biochemistry*, 43:8014–8020, 2004.
- [182] Lars-Oliver Essen, R Siegert, Wolf D Lehmann, and Dieter Oesterhelt. Lipid patches in membrane protein oligomers: crystal structure of the bacteriorhodopsin-lipid complex. *Proc Natl Acad Sci USA*, 95:11673–11678, 1998.
- [183] Salem Faham and James U Bowie. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J Mol Biol*, 316:1–6, 2002.
- [184] Yasuhiro Matsui, Keisuke Sakai, Midori Murakami, Yoshitsugu Shiro, Shin-ichi Adachi, Hideo Okumura, and Tsutomu Kouyama. Specific damage induced by x-ray radiation and structural changes in the primary photoreaction of bacteriorhodopsin. *J Mol Biol*, 324:469–481, 2002.
- [185] Gerard S Harbison, Steven O Smith, Johannes A Pardoën, Jacques M L Courtin, Johan Lugtenburg, Judith Herzfeld, Richard A Mathies, and Robert G Griffin. Solid-state ^{13}C NMR detection of a perturbed 6-*s-trans* chromophore in bacteriorhodopsin. *Biochemistry*, 24:6955–6962, 1985.
- [186] David R Baselt, Stephen P A Fodor, Rob van der Steen, Johan Lugtenburg, Roberto A Bogomolni, and Richard A Mathies. Halorhodopsin and sensory rhodopsin contain a C6-C7 *s-trans* retinal chromophore. *Biophys J*, 55:193–196, 1989.
- [187] Masaru Sakamoto, Akimori Wada, Akiko Akai, Masayoshi Ito, Takehiko Goshima, and Tetsuo Takahashi. Evidence for the archaeobacterial-type conformation about the bond between the small β -ionone ring and the polyene chain of the chromophore retinal in chlamyrodopsin. *FEBS Lett*, 434:335–338, 1998.

- [188] Hideki Kandori, Kazumi Shimono, Yuki Sudo, Masayuku Iwamoto, Yoshinori Shichida, and Naoki Kamo. Structural changes of *pharaonis* phoborhodopsin upon photoisomerization of the retinal chromophore: infrared spectral comparison with bacteriorhodopsin. *Biochemistry*, 40:9238–9246, 2001.
- [189] Christian Zscherp, Ramona Schlesinger, Jörg Tittor, Dieter Oesterhelt, and Joachim Heberle. *In situ* determination of transient pKa changes of internal amino acids of bacteriorhodopsin by using time-resolved attenuated total reflection Fourier-transform infrared spectroscopy. *Proc Natl Acad Sci USA*, 96:5498–5503, 1999.
- [190] Masayuki Iwamoto, Yuji Furutani, Yuki Sudo, Kazumi Shimono, Hideki Kandori, and Naoki Kamo. Role of Asp193 in chromophore-protein interaction of *pharaonis* phoborhodopsin (sensory rhodopsin II). *Biophysical J*, 83:1130–1135, 2002.
- [191] Masayuki Iwamoto, Yuji Furutani, Naoki Kamo, and Hideki Kandori. Proton transfer reactions in the F86D and F86E mutants of *pharaonis* phoborhodopsin (sensory rhodopsin II). *Biochemistry*, 42:2790–2796, 2003.
- [192] Yifan Song, Junjun Mao, and Marylin R Gunner. Calculation of proton transfers in Bacteriorhodopsin bR and M intermediates. *Biochemistry*, 42:9875–9888, 2003.
- [193] Michel F Sanner, Arthur J Olson, and Jean-Claude Spehner. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, 38:315–320, 1996.
- [194] Richard A Mathies and Lubert Stryer. Retinal has a highly dipolar vertically excited singlet state: implications for vision. *Proc Natl Acad Sci USA*, 73:2169–2173, 1976.
- [195] Paul E Blatz, Jonathan H Mohler, and Himanshoo V Navangul. Anion-induced wavelength regulation of absorption maxima of Schiff bases of retinal. *Biochemistry*, 11:848–855, 1972.
- [196] Timor Bassov and Mordechai Sheves. Model compounds for the study of spectroscopic properties of visual pigments and bacteriorhodopsin. *J Am Chem Soc*, 107:7524–7533, 1985.
- [197] Jingui G Hu, Robert G Griffin, and Judith Herzfeld. Interactions between the protonated Schiff base and its counterion in the photointermediates of bacteriorhodopsin. *J Am Chem Soc*, 119:9495–9498, 1997.
- [198] John Banisaukas, Jan Szczepanski, John Eyler, and Martin Vala. Vibrational and electronic spectroscopy of acenaphthylene and its cation. *J Phys Chem A*, 107:782–793, 2003.
- [199] Thomas M Halasinski, Jennifer L Weisman, Richard Ruitkamp, Timothy J Lee, Farid Salama, and Martin Head-Gordon. Electronic absorption spectra of neutral perylene (C₂₀H₁₂), terrylene (C₃₀H₁₆), and quaterrylene (C₄₀H₂₀) and their positive and negative ions: Ne matrix-isolation spectroscopy and time-dependent density functional theory calculations. *J Phys Chem A*, 107:3660–3669, 2003.

- [200] Ronald L Christensen, Elizabeth A Barney, Richard D Broene, Mary Grace I Galinatos, and Harry A Frank. Linear polyenes: models for the spectroscopy and photo-physics of carotenoids. *Arch Biochem Biophys*, 430:30–36, 2004.
- [201] A T Amos and B L Burrows. Solvent-shift effects on electronic spectra and excited-state dipole moments and polarizabilities. *Adv Quantum Chem*, 7:289–313, 1973.
- [202] Arieh Warshel. Calculation of chemical processes in solution. *J Phys Chem*, 83:1640–1652, 1979.
- [203] Jiali Gao. Monte Carlo quantum mechanical-configuration interaction and molecular mechanics simulation of solvent effects on the $n \rightarrow \pi^*$ blue shift of acetone. *J Am Chem Soc*, 116:9324–9328, 1994.
- [204] V Luzhkov and Arieh Warshel. Microscopic calculation of solvent effects on absorption spectra of conjugated molecules. *J Am Chem Soc*, 113:4491–4499, 1991.
- [205] Ignacio Tinoco, Kenneth Sauer, James C Wang, and Joseph D Puglisi. *Physical Chemistry: Principles and Applications in Biological Sciences*. Prentice Hall, New Jersey, 2001.
- [206] C J D Theodor de Grotthuss. Sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Ann Chim*, 58:54–73, 1806.
- [207] Klaus Gerwert, Benno Hess, Jörg Soppa, and Dieter Oesterhelt. Role of aspartate-96 in proton translocation by bacteriorhodopsin. *Proc Natl Acad Sci USA*, 86:4943–4947, 1989.
- [208] Benedikt Heßling, Georg Souvignier, and Klaus Gerwert. A model-independent approach to assigning bacteriorhodopsin's intramolecular reactions to photocycle intermediates. *Biophys J*, 65:1929–1941, 1993.
- [209] Johan Desmet, Marc de Maeyer, Bart Hazes, and Ignace Lasters. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*, 356:539–542, 1992.
- [210] Robert F Goldstein. Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys J*, 66:1335–1340, 1994.
- [211] Andrew R Leach and Andrew P Lemon. Exploring the conformational space of protein side chains using dead-end elimination and the A* algorithm. *Proteins*, 33:227–239, 1998.
- [212] Niles A Pierce, J A Spriet, J Desmet, and Steven L Mayo. Conformational splitting: a more powerful criterion for dead-end elimination. *J Comput Chem*, 21:999–1009, 2000.
- [213] Mario Hellings, Marc de Maeyer, Stefan Verheyden, Qiang Hao, Els J M van Damme, Willy J Peumans, and Yves Engelborghs. The dead-end elimination method, tryptophan rotamers, and fluorescence lifetimes. *Biophys J*, 85:1894–1902, 2003.

- [214] Stewart A Adcock. Peptide backbone reconstruction using dead-end elimination and a knowledge-based forcefield. *J Comput Chem*, 25:16–17, 2004.
- [215] Martin T Stiebritz and Yves A Muller. MUMBO: a protein-design approach to crystallographic model building and refinement. *Acta Crystallogr D Biol Crystallogr*, 62: 648–658, 2006.
- [216] Loren L Looger and Homme W Hellinga. Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. *J Mol Biol*, 307:429–445, 2001.
- [217] Premal S Shah, Geoffrey K Hom, and Steven L Mayo. Preprocessing of rotamers for protein design calculations. *J Comput Chem*, 25:1797–1800, 2004.
- [218] Ivelin Georgiev, Ryan H Lilien, and Bruce R Donald. Improved pruning algorithms and divide-and-conquer strategies for dead-end elimination, with application to protein design. *Bioinformatics*, 22:e174–e183, 2006.
- [219] Alexander V Lukashin and Joseph J Rosa. Local multiple sequence alignment using dead-end elimination. *Bioinformatics*, 11:947–953, 1999.
- [220] Naigong Zhang, Chen Zeng, and Ned S Wingreen. Fast accurate evaluation of protein solvent exposure. *Proteins*, 57:565–576, 2004.
- [221] I-Peng Chen, Paul Mathis, Juergen Koepke, and Hartmut Michel. Uphill electron transfer in the tetraheme cytochrome subunit of the *rhodospseudomonas viridis* photosynthetic reaction center: evidence from site-directed mutagenesis. *Biochemistry*, 39:3502–3602, 2000.
- [222] Torsten Becker, R Thomas Ullmann, and G Matthias 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*, in press, 2007.
- [223] Armen Y Mulkidjanian. Conformationally controlled pK-switching in membrane proteins: one more mechanism specific to the enzyme catalysis? *FEBS Lett*, 463: 199–204, 1999.
- [224] Marilyn R Gunner, Junjun Mao, Yifan Song, and Jinrang Kim. Factors influencing the energetics of electron and proton transfers in proteins. What can be learned from calculations? *Biochim Biophys Acta*, 1757:942–968, 2006.
- [225] David P Landau and Kurt Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, Cambridge, 2005.