

Enforced Unfolding of Bacteriorhodopsin via Force Probe Molecular Dynamics

Diploma Thesis

by

Christian Kappel

Department for Theoretical and Computational Biophysics

Max Planck Institute for Biophysical Chemistry

Göttingen

2006

Contents

1	Introduction	3
2	Biological and Physical Fundamentals	6
2.1	Proteins	6
2.2	Biological Membranes and Membrane Proteins	9
2.3	Bacteriorhodopsin and the Purple Membrane	13
2.4	Atomic Force Microscopy	13
2.4.1	Principle	13
2.4.2	Enforced Unfolding of Bacteriorhodopsin via AFM . . .	16
2.5	Unbinding Forces	18
3	Molecular Dynamics Simulations	19
3.1	Principles	19
3.2	Computational Details	22
3.3	Force Probe Molecular Dynamics	24
4	Methods	26
4.1	System Set-up	26
4.2	Simulation Set-up	28
4.3	FPMD Simulations	29

4.4	Interrupted FPMD Simulations	31
4.5	Data Analysis	32
4.5.1	RMSD and dRMSD	32
4.5.2	Formulas	33
4.5.3	Determining the Position of Unfolding Barriers	34
5	Results	38
5.1	Forces	38
5.2	Unfolding Pathway	43
5.3	Positions of Unfolding Barriers	54
5.4	Interrupted Pulling	58
6	Summary	63

Chapter 1

Introduction

Proteins perform most of the essential processes in living organisms [1]. Each protein is suited for one special task, like energy conversion, stabilization of tissues, or sensory functions. The specific function of a protein is tightly linked to its three dimensional shape. Therefore, the understanding of the forces that help to create and maintain the shape of proteins will increase our knowledge of their ability to perform different tasks.

The three dimensional structure of proteins is established in a process called protein folding. A protein folds from a randomly shaped amino acid chain into its native structure in a milliseconds to seconds timescale by minimizing its free energy. Until now, it is not known how protein folding takes place.

A method to investigate the underlying folding mechanism is to selectively unfold proteins under an external force. By doing so, the energy landscape which governs folding dynamics is explored by inverting the folding process.

One method that allows the unfolding of single proteins is the use of atomic force microscopy (AFM). In the last years unfolding experiments on

many different proteins have been carried out. A special form of AFM unfolding experiments has been performed in the groups of Hermann Gaub at the University of Munich and in the group of Daniel Müller at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden. In their experiments, enforced unfolding was carried out on the membrane protein bacteriorhodopsin [2, 3, 4, 5, 6, 7, 8, 9], which was the first time that single membrane proteins were extracted from their anchoring membrane.

Bacteriorhodopsin [10] is a helical integral membrane protein from the archaea species *halobacterium salinarum*, that converts sunlight into a proton gradient. The protein has become a model system for the investigation of membrane proteins over the last years. Its natural environment are membrane patches known as purple membrane, a hexagonal structure formed of circular bacteriorhodopsin trimers and surrounding lipid molecules.

The unfolding experiments on bacteriorhodopsin showed that the single proteins are strongly anchored in the membrane. Furthermore, the order of unfolding events was found to be determined by the order of the helices in the peptide chain. Closer analysis revealed that single residues, referred to as unfolding barriers, govern the stability of the protein. These unfolding barriers provide hints to the internal stability of the protein.

In this work, force probe molecular dynamics (FPMD) simulations [11] that resemble these experiments are presented. Starting from a simulation system with an explicit solvent and lipid bilayer environment, unfolding simulations towards the cytoplasmic and towards the extracellular side using various pulling velocities were performed.

The aim of this work was to obtain unfolding force profiles, to investigate

the unfolding pathway, and to compare unfolding barriers derived from the AFM experiments with those derived from the simulations.

This work is structured as following. In chapter 2, the biological and physical principles necessary to understand the work are explained together with a short description of the AFM experiments. The method of molecular dynamics simulations and FPMD simulations is the topic of chapter 3.

The methods and steps necessary to perform this work are described in chapter 4. In this chapter also introduces newly developed methods.

In chapter 5, the results are presented and discussed. After an analysis of the unfolding forces, the peak forces from the AFM experiments and the simulations are compared. This is followed by an extensive investigation of the unfolding pathway extracted from the simulations, together with an assignment to the unfolding forces. Furthermore, a closer look is taken at the role of the interaction energies in the simulations. Next, the unfolding barriers resulting from the AFM experiments are compared with the unfolding barriers derived from the simulations. The last part of this chapter analyses the behaviour of the protein under a non-moving pulling potential, allowing to explore the effects of the non-equilibrium state created by the very fast spring movement.

The results are summarized and discussed in a wider context in chapter 6.

Chapter 2

Biological and Physical Fundamentals

2.1 Proteins

Proteins are macromolecules, present in all living organisms. Nearly all functions essential for living are performed by proteins. They serve for example as structural proteins to build various tissues, as ion transporters, as channels for the transport of small molecules, or as enzymes to catalyze chemical reactions.

The size of a protein is measured by its molecular weight, which is given in the unit Dalton (Da). One Dalton is nearly equal to the weight of one hydrogen atom ($1 \text{ Da} = 1.661 \cdot 10^{-27} \text{ kg}$). The weight of typical proteins is in the range of 50 kDa. The largest known protein in the human body, the muscle protein titin, has a weight of 3000 kDa.

Proteins are linear polymers consisting of 20 different amino acids. All amino acids have the same backbone, consisting of one nitrogen and two

carbon atoms. The chemical and physical properties of each amino acid are determined by its side chain, which is connected to the C_α atom. Two amino acids can be connected by a chemical bond, the so called *peptide bond*. Short amino acid oligomers are also called *peptides* (figure 2.1).

The structure of proteins is described by four different levels. The *primary structure* of a protein is its amino acid sequence, which is coded in DNA or RNA. Each protein is identified by its primary structure.

Short amino acid segments can arrange into recurring local structure elements, referred to as *secondary structure*. Secondary structure elements include for example α -helices, β -sheets, and random coils. Both α -helices and β -sheets are defined by stabilizing hydrogen bonds involving the protein backbone. Amino acid side chains do not contribute to secondary structure elements. α -helices (figure 2.2 A) are regular spiral staircase like structures. A hydrogen bond is formed between the $C_\beta=O$ group from residue n and the N-H group from residue $n + 4$. Each helix turn consists of 3.6 residues, leading to an average angle of 100° per residue. The side chain atoms are oriented towards the helix outside. β -sheets (figure 2.2 B) are similarly stabilized by hydrogen bonds between the $C_\beta=O$ and the N-H group. But unlike α -helices, β -sheets are formed by two adjacent peptide chains rather than one single chain. According to the orientation of the two strands relative to each other, parallel and anti-parallel β -sheets are distinguished.

Proteins arrange into a well-defined spatial shape termed *tertiary structure*, defined by their primary structure, in a process called *protein folding*. With only few exceptions, the folded protein has a minimal free energy G . Unlike the secondary structure, the tertiary structure is not mainly stabilized

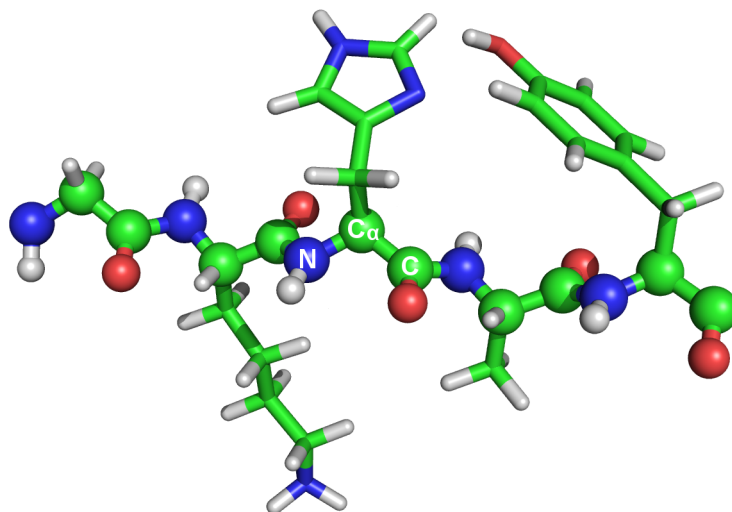


Figure 2.1: A sample peptide consisting of five different amino acids. Backbone atoms are drawn as spheres. The order of the residues from left to right is Glycine-Lysine-Histidine-Alanine-Tyrosine.

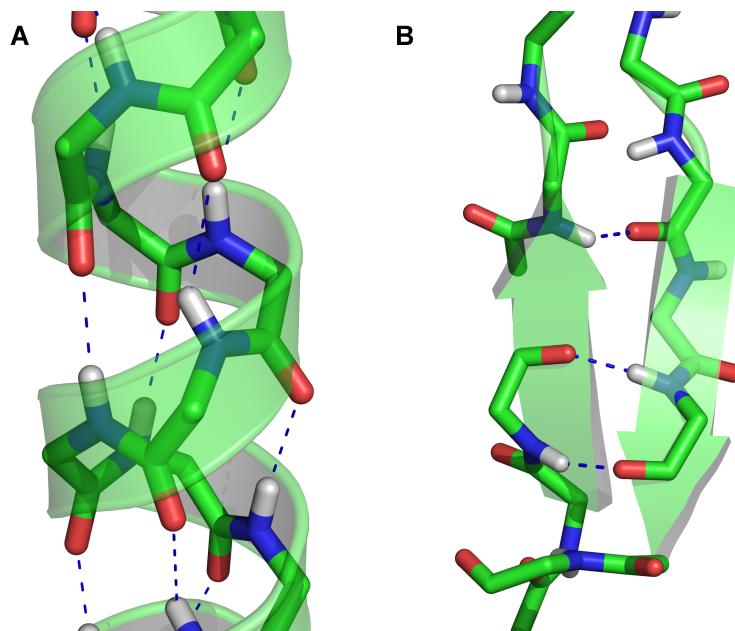


Figure 2.2: Secondary structure elements. Backbone atoms are shown in a stick representation. Hydrogen bonds appear as dotted blue lines. A: α -helix B: β -sheet

by hydrogen bonds inside the protein backbone, but also by van-der-Waals forces, by electrostatic forces, and by the hydrophobic effect. Additionally, side chain atoms are involved in the folding process. Even other molecules than amino acids or ions can contribute to the tertiary structure of a protein.

For functional purposes, folded proteins sometimes form larger assemblies. These structures of several folded proteins are called *quarternary structure*. The association forces are, as for the tertiary structure, van-der-Waals forces, electrostatic forces, and the hydrophobic effect.

2.2 Biological Membranes and Membrane Proteins

Biological membranes are barriers between a cell and the outside or between different compartments within a cell. The two main components of membranes are lipid molecules, which are arranged as a bilayer, and proteins that are incorporated within the lipid bilayer. The mass ratio of lipid molecules to proteins ranges between 1:4 and 4:1. The single molecules of a membrane are not covalently bound, but rather associate via van-der-Waals and electrostatic forces and under the influence of the hydrophobic effect. The two sides of a membrane always differ from each other, which allows a clear distinction between the inner and outer side.

Lipid molecules (figure 2.3 A) form the basis of every membrane. They consist of a polar head group and one or more apolar tails. In water, lipid molecules form compact structures. Due to the hydrophobic effect, the apolar tails aggregate such that the polar head groups face the polar water

environment. The simplest form of such a complex is a micelle (figure 2.3 B). This arrangement is unfavourable due to the ineffective packing of the rather thick tails. A common structure is a bilayer (figure 2.3 C), where lipid molecules form a thin film of 5 to 10 nm thickness.

The specific properties of a membrane is determined by its *membrane proteins*. Membrane proteins can be divided in two subclasses, *peripheral membrane proteins* and *integral membrane proteins*. Integral membrane proteins are fully incorporated in a membrane, whereas the main part of peripheral membrane proteins is outside the membrane and only the small anchor part is within the membrane.

The lipid bilayer and the membrane proteins act like a two-dimensional liquid. This behavior was first investigated by Singer and Nicolson and was called *fluid mosaic model* [12]. The model suggests that membrane proteins are free to diffuse in the membrane area, and tilting relative to the membrane normal is very unlikely and only possible under energy consumption.

An interesting type of integral membrane proteins are *helical membrane proteins*. These proteins consist of multiple membrane spanning α -helices, which are connected via loops. An easy way to detect membrane spanning helices is a *hydrophobicity plot* [13]. First, the difference in free energy that is necessary to move an amino acid from a lipid bilayer environment to water is determined. In the next step, the free energy changes are plotted over the residue number by grouping 20 amino acids together. The grouping is carried out by summing up the free energy changes of 20 amino acids (from n to $n + 20$). This sum is then plotted for each residue index n . Membrane spanning helices appear in the plot as values above +84 kJ/mol. A sample

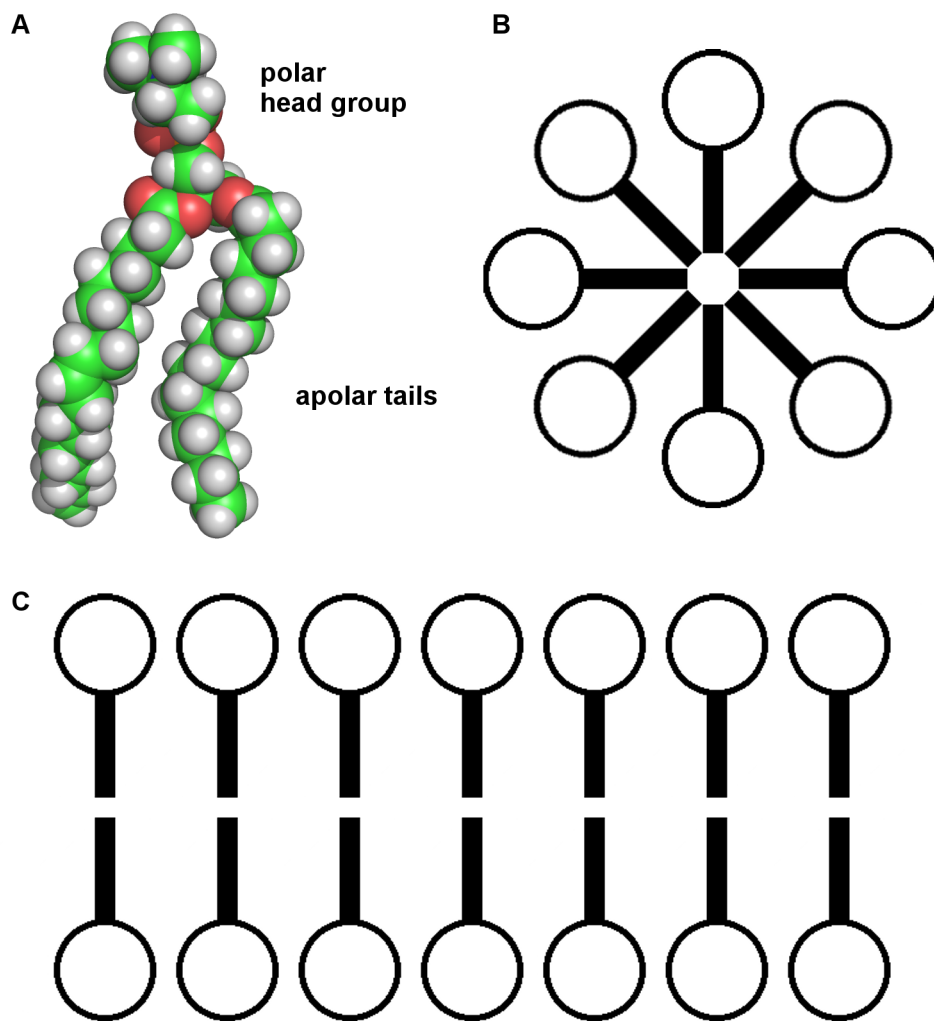


Figure 2.3: Structure and arrangements of lipid molecules. A: All-atom representation of a lipid molecules. Green atoms are carbon, white ones are hydrogen, and red ones are oxygen. B: Schematic representaion of a micelle. C: Schematic representation of a lipid bilayer.

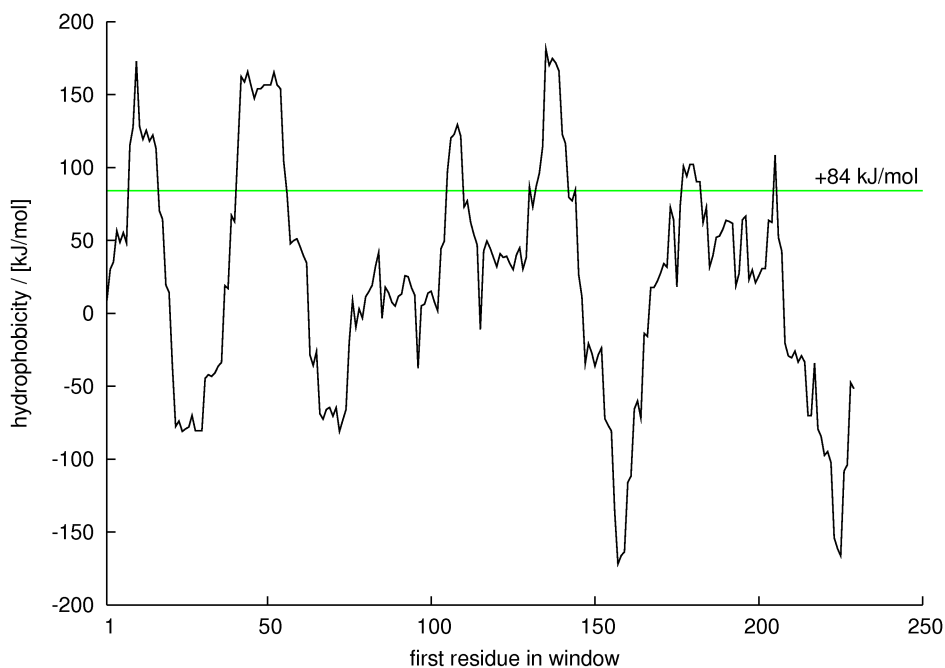


Figure 2.4: Hydrophobicity plot for bacteriorhodopsin. Values above the green line indicate trans-membrane α -helices.

plot for the helical membrane protein bacteriorhodopsin is given in figure 2.4.

The folding of helical membrane proteins can be described by the *two-stage model* proposed by Popot and Engelman [14]. This model suggests that in the first step the trans-membrane α -helices fold independently. The second step consists of the aggregation of the single helices, resulting in the final tertiary structure.

Both methods use a rather heuristic approach for the description of membrane proteins. It would be more interesting, however, to explore membrane protein behaviour by a physical point of view. This can be done by inverse folding via AFM and FPMD, as described in more detail in the results.

2.3 Bacteriorhodopsin and the Purple Membrane

The integral membrane protein bacteriorhodopsin is a light-driven proton pump, found in the archaea species *halobacterium salinarum*. The protein consists of seven membrane spanning α -helices, forming an arc (figure 2.5 A). In the protein core a chromophore, called *retinal*, is buried.

Bacteriorhodopsin converts green light (500 – 650 nm wavelength) into a proton gradient by light-driven proton transfer from the cytoplasmic side to the extracellular side. The proton transfer is initiated by a light induced isomerisation of the retinal. This isomerisation starts a reaction cascade, in which the protein undergoes several conformational changes.

Bacteriorhodopsin is located in membrane patches known as *purple membrane* (figure 2.6). The purple membrane is a 2D-crystal consisting of bacteriorhodopsin and lipid molecules in a weight ratio of 4 : 1. Three bacteriorhodopsin proteins build a cylindrical structure forming a hexagonal lattice with adjacent lipid molecules. The purple membrane is known to be remarkably stable.

2.4 Atomic Force Microscopy

2.4.1 Principle

Various techniques exist which allow the investigation of materials at the atomic level. These include x-ray diffraction, scanning tunnel microscopy (STM) or electron microscopy. Another very simple yet effective method is

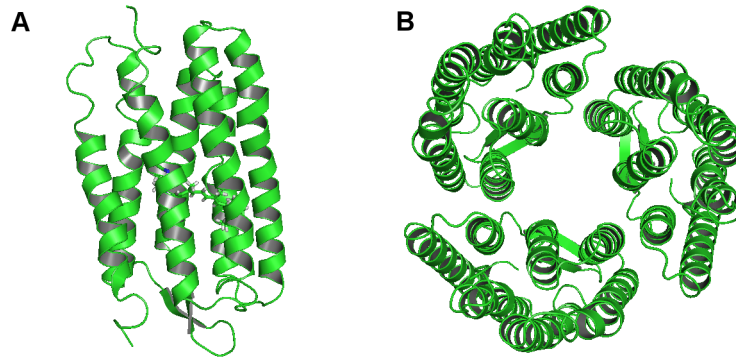


Figure 2.5: Ribbon diagram of bacteriorhodopsin. A: Protein monomer (side view). The retinal is drawn in a stick representation. B: Top view of a bacteriorhodopsin trimer from the cytoplasmic side.

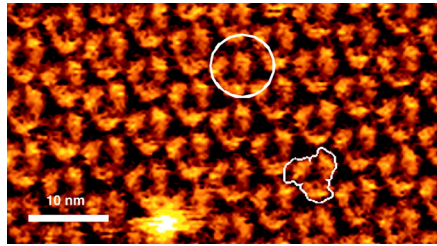


Figure 2.6: AFM picture of the purple membrane. The irregular shape outlines a trimer. In the center of the circle a single protein is denoted. The bar in the lower left shows a reference distance (10 nm). Taken from ref [2].

atomic force microscopy (AFM), which was invented by Binnig, Quate, and Gerber in 1986 (figure 2.7) [15].

The main component of an AFM is a cantilever with a very sharp tip which ends up in only a few atoms. Like in a STM, a sample is investigated by a scanning process, where the tip is moved across the sample. This movement is carried out by piezo crystals, which also allow vertical displacement of the cantilever.

The deflection Δx of the cantilever is induced by a force F which is given

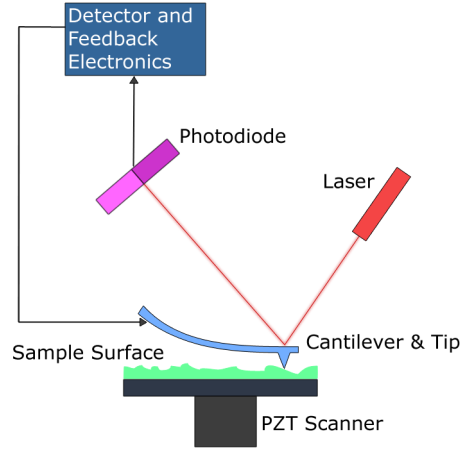


Figure 2.7: Schematic picture of an AFM. Green: sample, the tip of the cantilever scans the sample surface, the deflection of the cantilever is measured with the help of the reflected laser beam via a photo diode

by Hooke's law

$$F = -k \cdot \Delta x, \quad (2.1)$$

where k is the spring constant of the cantilever. The deflection is measured by a light beam that is reflected on the cantilever. The total height information is given by the sum of cantilever deflection and the vertical position from the piezo crystal elongation.

AFM images of a sample can be created in different ways. In *constant height mode* the distance between the cantilever and the sample is kept fixed during scanning. The surface information is inferred from the cantilever deflection. The advantage of this method is that a regulation of the cantilever-sample distance is not needed, but due to possible destruction of the sample, this method is only applied to flat or hard samples. In *constant force mode* the cantilever-sample distance is constantly adopted to achieve a constant force. This scanning mode is slower than constant height mode but it has

the advantage that the risk of damaging the sample is lower.

A different approach is used in *tapping mode*. Here, the cantilever is oscillated through an external force, where the frequency is slightly above the resonant frequency. The tip touches the sample at each oscillation maximum. Interactions between the tip and the sample result in a change in frequency and amplitude, which is measured. This method is useful for soft surfaces like biological samples.

For this work, the interesting application of AFM is the measurement of force-deflection curves. Here, the tip is linked to a molecule. By moving the cantilever, the molecule is exposed to a force which can be measured by the cantilever deflection. In this way, forces in the range of few pN can be applied and measured. With AFM biological samples can not only be imaged, but also manipulated. One application is the enforced unfolding of single proteins. In those experiments, a linker molecules is linked to the AFM tip and attached to a strand of DNA or a protein. By moving the cantilever with constant velocity, the biological sample is exposed to a force and unfolds. The forces during the enforced unfolding of the sample can be measured.

2.4.2 Enforced Unfolding of Bacteriorhodopsin via AFM

In recent experiments in the group of Hermann Gaub at the University of Munich, single bacteriorhodopsin proteins were extracted from the purple membrane using AFM [2, 8]. A pulling force of some 100 pN was applied to enforce unfolding. The force profiles from these experiments showed four characteristic peaks (figure 2.8). After extraction of one protein, a hole was

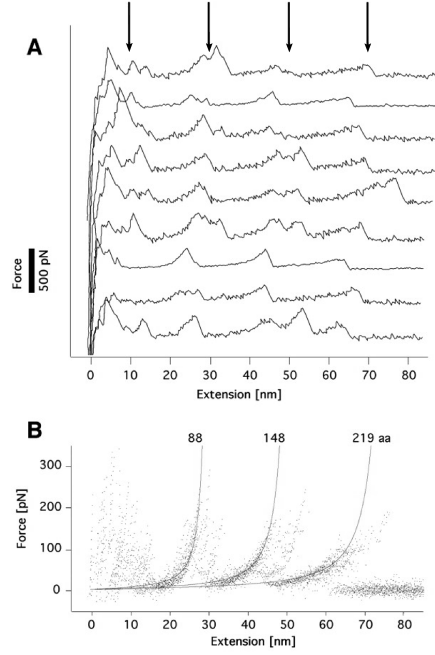


Figure 2.8: Results from the AFM experiments. A: Several single unfolding force profiles. The arrows on top point to the mean positions of the force peaks. B: Overlaid force profiles. Three peaks can be clearly identified. Taken from [2].

left in the purple membrane. The remaining proteins kept their initial positions in the protein/lipid matrix.

By fitting a worm-like-chain (WLC) force-extension curve to each peak (figure 2.8 B), the length of the unfolded peptide chain was determined, implying that the protein unfolds helix-wise in a linear order.

Further examination of the force profiles yielded residues that act as *unfolding barriers*. These residues show a particularly strong resistance against unfolding forces and therefore provide insight in the mechanical stability of the protein.

The aim of this work is to reproduce these experiments by FPMD simulations. This way, observables not accessible via AFM can be explored, like the

actual unfolding pathway or the atomic behaviour of the unfolding barriers. The atomic description of the unfolding process accessible with FPMD helps to increase the understanding of the AFM experiments.

2.5 Unbinding Forces

Although AFM experiments and FPMD simulations cover the same unfolding events, the unfolding velocities from both methods differ by six orders of magnitude. Accordingly, the unbinding forces differ by several orders of magnitude. To draw quantitative conclusions from comparing unbinding forces obtained from both methods, a method developed by Heymann and Grubmüller [16] is used in this work.

This method suggests that unfolding forces from different pulling velocities v can be extrapolated using a total force

$$F(v) = F_{frict}(v) + F_{act}(v), \quad (2.2)$$

where $F_{frict}(v)$ is a velocity dependent friction force and $F_{act}(v)$ is a velocity dependent force used for the crossing of a free energy barrier.

The final equation is given by

$$F(v) = \gamma v + \frac{k_B T}{L} \ln \left(\frac{v}{k_0 \Delta L} \right), \quad (2.3)$$

where γ is a friction coefficient, k_B the Boltzmann constant, T the temperature, L the unfolding length, k_0 the spontaneous dissociation rate, and ΔL the scatter of the unfolding length.

Chapter 3

Molecular Dynamics Simulations

3.1 Principles

Molecular dynamics (MD) simulations describe atomic motions in large molecular systems. This method calculates all atomic positions depending on inter- and intramolecular forces for a given time range. A interesting field of application for MD is the description of protein motions. Here, MD simulations are able to provide atomistic insight in biological processes, that are in many cases otherwise not accessible.

An exact description of the motions of atoms is given by the time-dependent Schrödinger equation. But already for a helium atom it is not possible to find an analytical solution. Numerical solutions are only feasible for up to 10 atoms. Therefore, to date an explicit quantum mechanical description for large molecules is not possible. For larger systems, molecular dynamics simulations can be employed. To reduce the computational effort, three approximations are made. The first approximation is the *Born-Oppenheimer approximation*. This approximation rests on the fact that the mass of elec-

trons and nuclei differs by several orders of magnitude. Accordingly, electrons move much faster than nuclei. Thus, the motion of electrons and nuclei can be described separately, allowing to consider the nuclei positions as quasistatic. This approximation results in a time-independent Schrödinger equation for the electron positions. Therefore, an effective time-dependent potential for the atom motions is given, which only depends on the position on the nuclei.

The second approximation is the use of a *force field* for the description of inter- and intramolecular energies (figure 3.1). A force field is composed of energy terms describing interactions between selected atoms. These energy terms include bond stretching, bond angles, extraplanar and dihedral angles, coulombic interactions, van-der-Waals interactions, and Pauli repulsion (figure 3.1). All these terms are described with mathematically simple expressions (x^n , $1/x^n$, or $\cos x$). The parameters (equilibrium constants, Pauli radii, partial charges, etc.) are derived from quantum mechanical calculations and experiments. The parameters sets are adjusted, such that they reproduce the relevant physical observables.

As a third approximation, the motions of all nuclei positions \mathbf{r}_i are described classically by *Newton's equation of motion*

$$m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} = -\nabla_i V(\mathbf{r}_1, \dots, \mathbf{r}_N), \quad (3.1)$$

where i is the atom index, N the number of atoms, t the time, m_i the mass of atom i , and V the force field. The use of a classical description instead of a quantum mechanical description is justified as long as quantum effects are not needed for the description of the system.

Because the resulting equations of motion are not analytically solvable, *numerical integration* is used to calculate the atom positions as a function

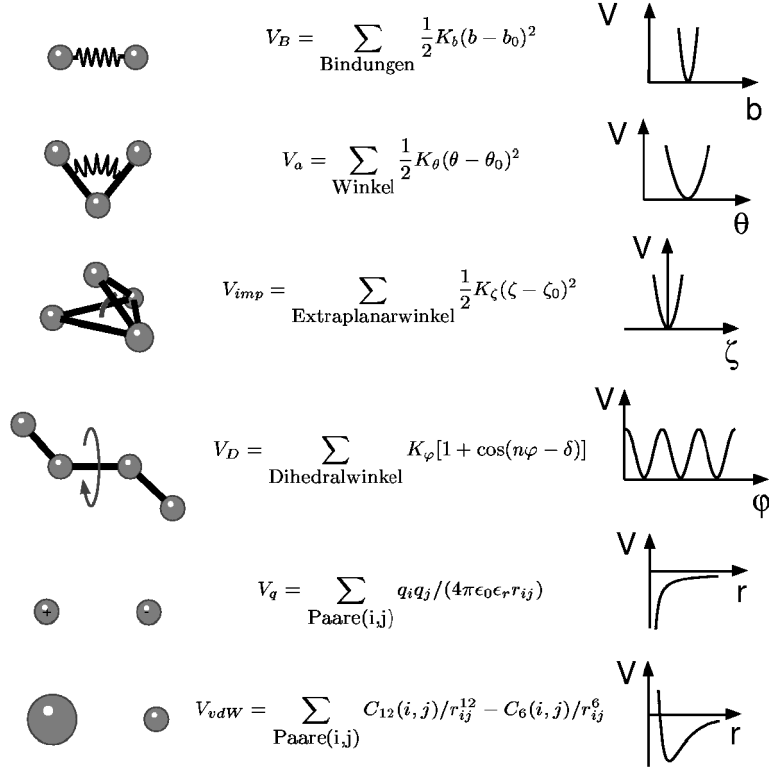


Figure 3.1: Single energy terms in a MD force field. Picture taken from reference [17].

of time. Accordingly, all atom positions are calculated at every integration time step Δt from previous positions and velocities. This time step has to be much smaller than one period of the fastest fluctuations in the system.

The calculation of the position \mathbf{r} and the velocity \mathbf{v} of an atom is calculated in this work using the leap-frog algorithm [18]:

$$\mathbf{v}\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}\left(t - \frac{\Delta t}{2}\right) + \frac{\mathbf{F}(t)}{m} \Delta t \quad \text{and} \quad (3.2)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}\left(t + \frac{\Delta t}{2}\right) \Delta t, \quad (3.3)$$

where \mathbf{F} is the force acting on the atom.

3.2 Computational Details

In an MD simulation only a limited number of atoms can be described. Therefore, the considered system size is limited and artefacts may occur.

To avoid artefacts due to the finite system size or surfaces, *periodic boundary conditions* were used in this work.

To this end, the atoms of the system are put into a space-filling box, which is surrounded by translated copies of itself (figure 3.2). Thus, surface effects were avoided, but artefacts may be induced when molecules interact with their copies. To avoid those interactions, a sufficiently large system size was chosen.

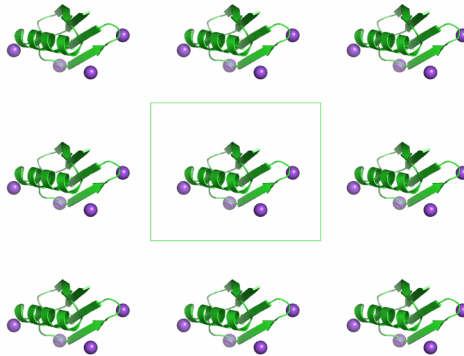


Figure 3.2: Principle of periodic boundaries conditions shown in two dimensions. The simulation system (the green box in the middle) is surrounded by copies of itself.

Because living organisms and therefore proteins exist in an environment with small changing temperature and pressure, these parameters have to be close to a constant value in order to reproduce the natural environment.

In the case of temperature, the velocities of the atoms were adjusted to

keep a reference temperature T_0 . To this end a *Berendsen thermostat* [19] was employed, a method which resembles a heat bath. In each time step all velocities were scaled by

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau} \left(\frac{T_0}{T} - 1 \right)}, \quad (3.4)$$

where τ is a time constant and T is the instantaneous kinetic temperature of the system.

Pressure coupling to a reference pressure P_0 was obtained by a *Berendsen barostat* [19]. Similar to the Berendsen thermostat, the lengths of the box vectors of the simulation system were rescaled by a factor

$$\mu = 1 - \frac{\Delta t}{3\tau} \kappa (P_0 - P) \quad (3.5)$$

with a time constant τ and a compressibility κ .

For a large system, direct calculation of interaction energies is computationally very expensive. To reduce the computational effort, interactions at long-range distances were treated differently from short-range distances. Short-range interactions below a *cut-off distance* of 1 nm were calculated directly via the given energy terms. For long ranges, Lennard-Jones interactions above the cut-off distance were not calculated in this work, and electrostatic interactions were calculated with *particle mesh Ewald summation* (PME) [20, 21].

3.3 Force Probe Molecular Dynamics

The main goal of this work was to simulate AFM unfolding experiments. This was done by *Force probe molecular dynamics* [11] (FPMD) simulations. FPMD simulations are a method to interfere with and control atomic motions in a MD simulation. An atom is selected and probed by applying a force F on it (figure 3.3).

A moving AFM cantilever is described by a pulling potential V_{pull} that acts on a selected atom. The pulling potential moves with constant velocity v in pulling direction. A description of the pulling potential is given by

$$V_{pull}(z, t) = \frac{k}{2} (z - z_0 - vt)^2, \quad (3.6)$$

where t is the time, k the width of the potential (equivalent to the spring constant of the cantilever), z the position of the subjected atom, and z_0 the initial spring position. The force F acting on the pulled atom is given by

$$F(z, t) = -k(z - z_0 - vt). \quad (3.7)$$

The resulting forces depend on the chosen spring constant as well as on the pulling velocity.

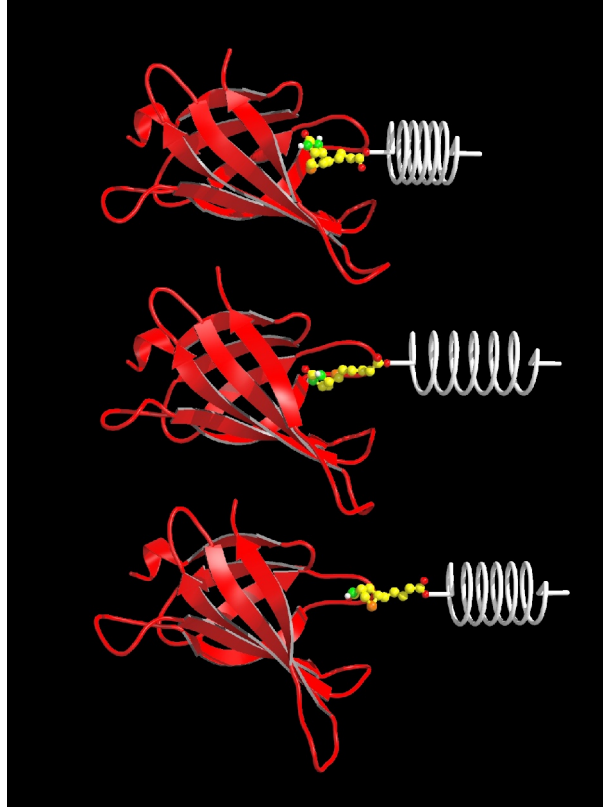


Figure 3.3: Schematic representation of the FPMD principle. The three images show a time development, starting from the upper image. A ligand molecule (drawn in yellow) is pulled out of the protein (drawn in red). The spring symbolizes the force acting on the ligand. A stretched spring indicates a high force. Taken from ref [11]

Chapter 4

Methods

4.1 System Set-up

As a starting structure, the bacteriorhodopsin x-ray structure (Protein Data Bank model 1QHJ) [22] was used. Three inner residues not resolved in the x-ray structure (MET163, ARG227, and GLU232) were modelled using Whatif [23]. Missing terminal residues (1-4 and 233-248) were not taken into account. The native trimer conformation was created by applying the symmetry operations given in reference [22]. A model for a POPC lipid membrane patch, kindly provided by Peter Tieleman (<http://moose.bio.ucalgary.ca/index.php?page=Downloads>) was placed to surround the proteins following the procedure described in reference [24]. Water molecules and sodium and chloride ions according to the physiological salt concentration of archaea (300 mM) were added using the programs genbox and genion from the GRO-MACS simulation suite [25]. This way a solvent layer of 5 nm thickness was placed on top of the system.

Subsequently, energy minimization of 150 steps was performed using

steepest descent. Three equilibration runs with increasingly relaxed harmonic position restraints were performed. First, for 200 ps restraints with a force constant of $k = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ were applied to all non-hydrogen atoms. Subsequently, for 1 ns only the protein restraints were kept. Finally, a free equilibration run of 3.5 ns was performed without any position restraints.

To leave sufficient room for the extracted and unfolded peptide, we added further 5 nm of water with 300 mM ions on top of the system, accumulating to a total water layer of 10 nm. This enlarged system was equilibrated for further 500 ps.

A 2×2 trimer system (figure 4.1) was then constructed by placing three copies of the trimer system according to the the crystal symmetry. The full simulation system thus contained 12 monomers. Despite the significantly increased computational effort, we considered such a large system essential to avoid self-contact of the trimers due to the periodic boundaries, which would likely cause artefacts for the extraction of a monomer.

This full system was equilibrated for another 500 ps and used as the starting structure for all subsequent cytoplasmic pulling simulations described below. For pulling towards the extracellular side, the 10 nm water layer was placed at the opposite side of the system as described above, and the whole system was rotated by 180° around the x-axis, such that the same pulling direction (positive z -direction) could be used for both systems. Both simulation systems comprised 236 124 atoms, with a box size of $12.16 \times 12.16 \times 15.32 \text{ nm}^3$.

Pulling simulations with various pulling velocities were carried out for each of the two systems as summarized in table 4.1.

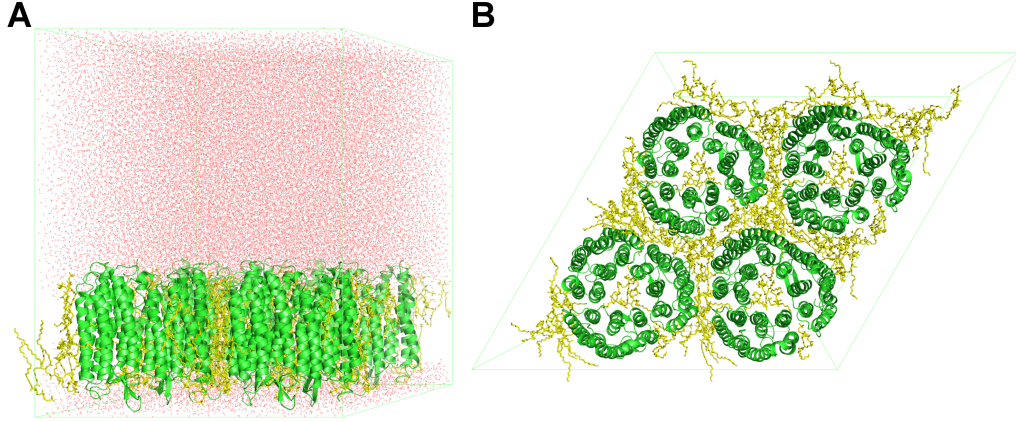


Figure 4.1: The simulation system. Proteins are shown in green, lipid molecules in yellow, and water molecules in red. The simulation box is drawn as green lines. A: side view of the system. B: top view of the system.

4.2 Simulation Set-up

All simulations were carried out using the software package GROMACS 3.3. Proteins, and the sodium and chloride ions were described with the OPLS all-atom force field [26]. For water the TIP4P water model [27] was employed. Lipid molecules were described with a unified atom model using parameters taken from reference [28] and modified to match the OPLS force field. Partial charges for the retinal were taken from reference [29]. All other force field parameters were converted into the OPLS force field according to reference [29].

Simulations were run in the NPT ensemble. Temperature coupling at $T = 300\text{ K}$ was performed using a Berendsen thermostat [19] with a relaxation time constant $\tau = 0.1\text{ ps}$. Pressure was kept constant at $p = 1\text{ bar}$ using semiisotropic pressure coupling. The box size was kept constant in x - and y -direction, i.e., within the membrane plane, preserving the area of

the membrane in these directions. In contrast, the box size in z -direction was free to adopt to pressure changes. A compressibility of $4.5 \cdot 10^{-5} \text{ bar}^{-1}$ and a relaxation time constant $\tau_p = 1.0 \text{ ps}$ were chosen. Long range electrostatic interactions beyond 1.0 nm were calculated using particle mesh Ewald summation [20, 21]. A grid dimension of 0.12 nm and fourth order b-spline interpolation was used. Short range electrostatic interactions were calculated by direct summation. For Lennard-Jones interactions a cut-off length of 1.0 nm was chosen. The lengths of bonds involving hydrogen atoms were constrained using LINCS [30]. An integration time step of 2 fs was used.

4.3 FPMD Simulations

Force probe simulations were carried out for extraction towards both sides, the cytoplasmic and the extracellular side. In each of these simulations the C_α carbon atom of the C- and N-termini (figure 4.2), respectively, was subjected to a harmonic pulling potential V_{pull} , which was moved with constant velocity in z -direction away from the membrane, parallel to the membrane normal,

$$V_{pull}(t) = \frac{1}{2}k(z_{C_\alpha}(t) - z_{\text{Spring},0} - vt)^2, \quad (4.1)$$

where $k = 500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ is the spring constant, $z_{C_\alpha}(t)$ is the z -position of the respective C_α atom, $z_{\text{Spring},0}$ the z -coordinate of the initial spring position, and v is the pulling velocity. Unlike the usual FPMD approach, here the spring position was calculated in absolute coordinates rather than relative to a reference atom. No position restraints were applied to the system.

Aiming at extracting and unfolding of a complete bacteriorhodopsin monomer, the fully extended polypeptide chain would be much too long to

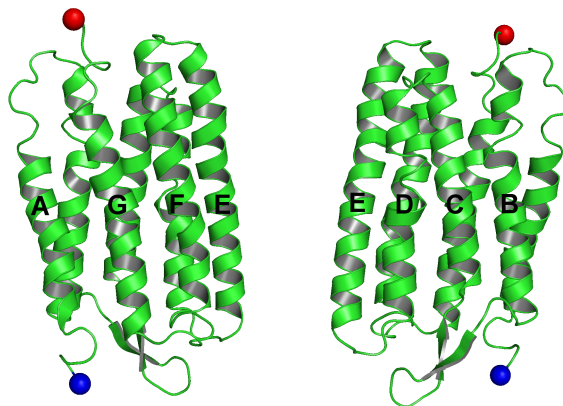


Figure 4.2: Cartoon representation of bacteriorhodopsin. Atoms which were subjected to a pulling potential are marked with spheres. The red sphere shows the C-terminus and the blue one the N-terminus. Helices A to G are indicated.

fit into a reasonably sized simulation box. To keep the simulation system computationally tractable, we repeatedly cut off those unfolded parts of the protein that had moved sufficiently far away from the membrane to render interaction with the membrane negligible. Accordingly, whenever the pulled C_α atom reached a distance of 1 nm to the system border, the unfolding simulation was interrupted and extracted residues that were more than 1 nm above the membrane in z -direction were removed.

The resulting solvent gaps were filled with water molecules, and ions were added or removed to maintain the system uncharged. Subsequently, new termini were built, the system was energy minimized, and equilibrated for 20 ps with position restraints on the heavy atoms of the protein and on all lipid atoms.

The force probe unfolding simulation was resumed with a force F_{new} equal to the force F_{old} applied at the time of interruption. From equation (4.1) and

$F = -\nabla V$, this was achieved by a suitably chosen new spring position,

$$k \cdot (z_{Spring,new} - z_{C_\alpha,new}) = k \cdot (z_{Spring,old} - z_{C_\alpha,old}), \quad (4.2)$$

$$\text{implying } z_{Spring,new} = z_{C_\alpha,new} + z_{Spring,old} - z_{C_\alpha,old}. \quad (4.3)$$

The spring constant was kept unchanged during the simulation. The whole procedure was iterated until the protein was completely extracted.

pulling velocity in nm/ps	length of pulling simulation in ns	
	cytoplasmic side	extracellular side
0.001	78.000	-
0.005	16.340	16.400
0.01	8.370	8.200
0.02	4.395	4.214
0.05	1.765	1.752

Table 4.1: Summary of pulling velocity and length of the force probe simulations carried out.

4.4 Interrupted FPMD Simulations

Compared to the millisecond experimental time scale, unfolding is enforced to proceed very fast in our simulations due to the limited time scale of MD. As a consequence, during pulling the system is not likely to be as close to equilibrium as in the experiment. To determine how far the protein is away from equilibrium and thus to better connect to experiment, we studied relaxation motions of the protein under the influence of a static pulling potential.

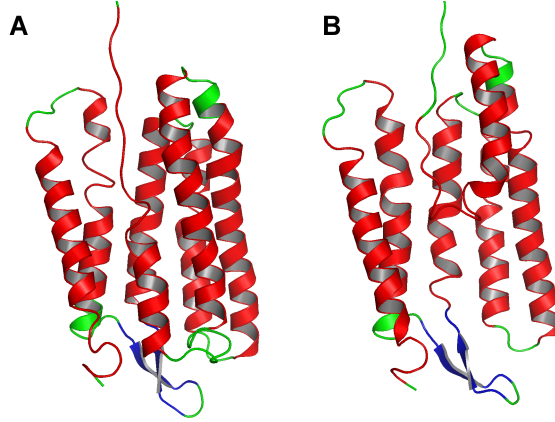


Figure 4.3: Snapshots of the partially unfolded proteins. A: 300 ps, B: 600 ps

Two snapshots of the trajectory from the pulling towards the cytoplasmic side with pulling velocity $v = 0.02 \text{ nm/ps}$ were used. The first one (figure 4.3A) was taken at 300 ps (6 nm spring elongation), where helix G was half unfolded. The second one (figure 4.3B) was taken at 800 ps (16 nm spring elongation), where helix F was half unfolded. For each of the cases a 20 ns FPMD simulation with $v = 0.0 \text{ nm/ps}$ with static spring position was resumed.

4.5 Data Analysis

4.5.1 RMSD and dRMSD

Beside the coordinates and velocities of the atoms, further observables are *root mean square deviation* (RMSD) and *root mean square deviation of atom distances* (dRMSD). In this work, these two observables are employed to examine structure deviations.

RMSD values show overall structural changes of a protein and are given

by

$$RMSD(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N [(x_i(t) - x_i(0))^2 + (y_i(t) - y_i(0))^2 + (z_i(t) - z_i(0))^2]}, \quad (4.4)$$

where t is the time, x_i , y_i , and z_i are the coordinates of the examined structure, and N is the number of atoms. In many cases, as a first step before calculating the RMSD, the structure at time t is translated and rotated to match best with the reference structure. The RMSD is sensitive to small structural changes and therefore a good indicator for beginning unfolding events. If the investigated structure is in a stable state, the RMSD stays stable.

The dRMSD indicates the changes of the distance $r_{ij}(t)$ between all possible pairs of atoms i and j at a given time t in reference to the starting structure, according to

$$dRMSD(t) = \sqrt{\frac{1}{2N(N-1)} \sum_{i \neq j} (r_{ij}(t) - r_{ij}(0))^2}. \quad (4.5)$$

dRMSD values of α -helices show a steady increase upon a loss of secondary structure. Stable states with half unfolded helices moving as a whole, would cause RMSD values to be unstable, while dRMSD curves remain stable.

4.5.2 Formulas

For each of the simulations, the pulling force exerted onto the respective atom subjected to the pulling potential,

$$-\nabla V = -k \cdot (z_{\text{Spring}} - z_{C\alpha}), \quad (4.6)$$

was calculated and recorded every picosecond. Hydrogen bond energies were estimated according to [31],

$$E_{HB} = -\frac{1}{2} \cdot (50 \cdot 10^3 \text{kJ/mol}) \cdot e^{-36\Delta s}. \quad (4.7)$$

Here, Δs denotes the donor-acceptor distance in nm.

4.5.3 Determining the Position of Unfolding Barriers

During the enforced unfolding certain residues are able to resist external force in a stronger way than others. We will refer to these residues, which permit the investigation of the stability of the protein, as unfolding barriers.

In the AFM experiments, the position of the unfolding barriers of bacteriorhodopsin were determined by fitting a worm-like-chain force curve to the force profiles and identifying those residues that cause the force peaks. In contrast, our approach is based on the trajectory of the unfolding protein. To this end, a snapshot of the unfolded protein was taken for every time frame. From these snapshots the deviations of the positions of the C_α -atoms from the starting structure in z -direction were determined and plotted over the residue index.

As an illustration, figure 4.4 shows a partly unfolded helix with a corresponding Δz -plot. As can be seen, the folded part remains near $\Delta z \approx 0$ nm, whereas the unfolded part is identified by a steep rise in Δz . The *transition point* between the folded and the unfolded part of the protein is assumed to be determined by the kink in the Δz -plot. *Unfolding barriers* are defined as transition points, which occurs over a longer time and therefore are more frequent.

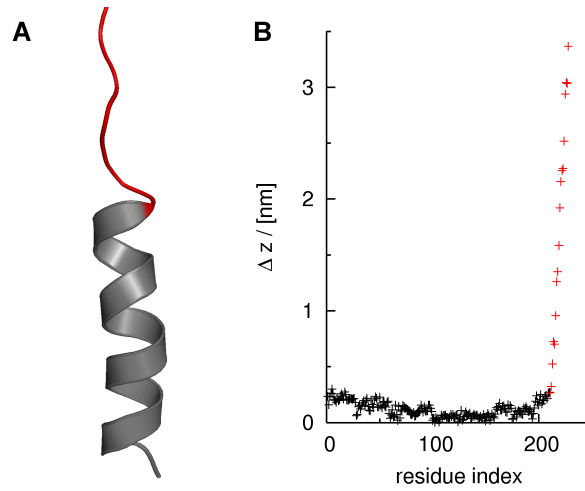


Figure 4.4: Explanation of the Δz -plot. A: Snapshot of a partially unfolded helix, B: corresponding Δz -plot for the whole protein. Grey regions in the protein picture correspond to black points in the plot. Red regions in both plots depict the unfolded part.

The analysis of these curves, described below, rests on a number of assumptions, which are fulfilled here for the unfolding of bacteriorhodopsin:

- Enforced unfolding takes place in one direction, here the z -direction
- Unfolding occurs in a sequential order, one helix after the other
- Remaining folded parts of the protein do not perform significant translational or rotational movements
- Unfolded parts form a stretched rod-like peptide chain
- The transition point between the folded and the unfolded part is sharp

The unfolded part corresponds to a linear increase in the Δz -plot, starting with the transition point. By combining these two facts, transition points

were determined in this work by identifying the linearly increasing parts in the Δz -plots through a line fitting algorithm.

To this end, first the complete data set with starting residue $N_0 = 1$ was taken. Next, a line fit was calculated via linear regression. Each line fit yielded a slope m and a slope error σ_m . The relative slope error σ_m/m together with the number of the first considered residue was stored. As a next step, the starting residue was discarded, and another line fit was calculated between the second residue and the last residue. This procedure was repeated until only three residues were left.

The line fit belonging to the minimal σ_m/m corresponds to the unfolded part of the protein, since only the linearly increasing part in the Δz -plot is taken into account and residues belonging to the folded part of the protein were not considered. Therefore, the residue belonging to the minimal σ_m/m denotes the transition point. A schematic illustration of the algorithm is presented in figure 4.5. The corresponding σ_m/m values are plotted in figure 4.6.

This algorithm yields a time-development of transition points, from which histograms were created. From these histograms, the most frequent transition points were defined as unfolding barriers.

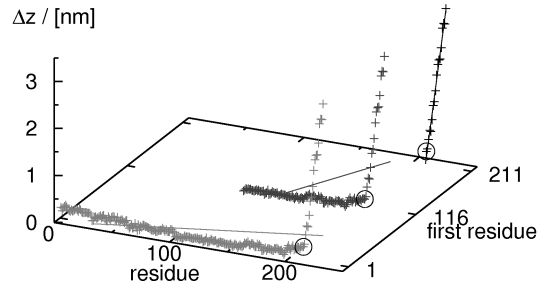


Figure 4.5: Illustration of the linear regression method. The circles depict the transition point. Different subsets are drawn as points in different colors. The x-axis denotes the residue index, the y-axis the first residue of the drawn subset, and the z-axis the Δz values. The lines are the regression lines for each subset.

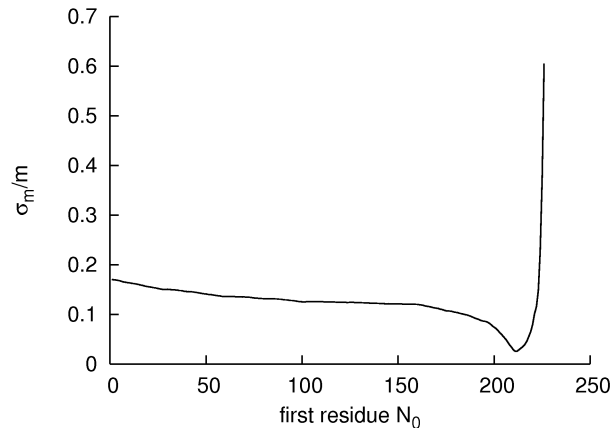


Figure 4.6: σ_m/m plotted as a function of the first residue. The data set is the same as in figure 4.5. The minimum position can be clearly seen.

Chapter 5

Results

Several unfolding simulations where a complete bacteriorhodopsin monomer was extracted and unfolded from the purple membrane were carried out. Simulations of both pulling directions, towards the cytoplasmic and towards the extracellular side, were conducted using different pulling velocities for each side (summarized in table 4.1). Unless otherwise mentioned, all presented results were obtained from the simulation with pulling velocity $v = 0.005 \text{ nm/ps}$ towards the cytoplasmic side.

5.1 Forces

Figure 5.1 shows force profiles from pulling simulations towards the cytoplasmic and towards the extracellular side, respectively, at different speeds. Both plots show the applied unfolding force as a function of spring position. Each force curve depicts a different pulling speed and starts with $F = 0 \text{ pN}$. As can be seen, all force profiles from one pulling direction are relatively similar, independent of the pulling velocity. In particular each curve exhibits four

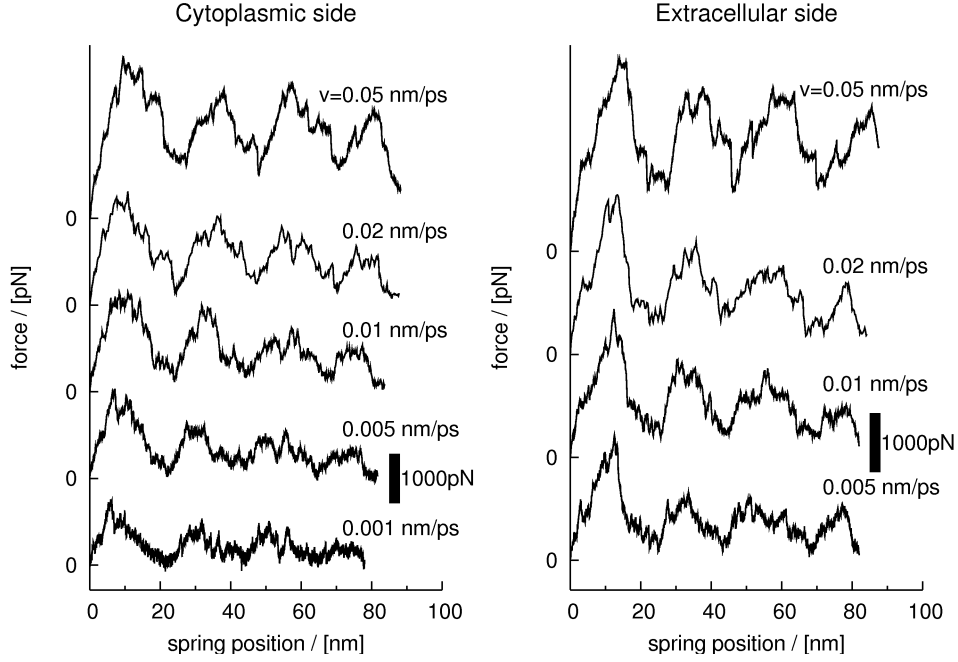


Figure 5.1: Force profiles from the FPMD simulations. All recorded force profiles from the pulling towards the cytoplasmic side (left plot) and towards the extracellular side (right plot) are shown. In the upper right corner of each curve the pulling velocity is given. All force profiles start with $F = 0$ pN.

separate force peaks. Some of these peaks split up into two subpeaks, like, e.g., the first and the third peak when pulling towards the cytoplasmic side. As must be expected, the height of the force peaks increased with pulling velocity, whereas the positions of the peaks remained similar for all pulling velocities, indicating that the unfolding pathways are also similar.

For comparison, figure 5.2 shows several measured AFM force profiles (data from Gaub and co-workers [8]). The forces were recorded using a pulling velocity of $1.4 \cdot 10^{-6} \text{ m/s}$, six orders of magnitude slower than in the simulations. Like in the MD force profiles, several distinct peaks can be seen for each curve. Upon overlaying all available force curves, Gaub and co-

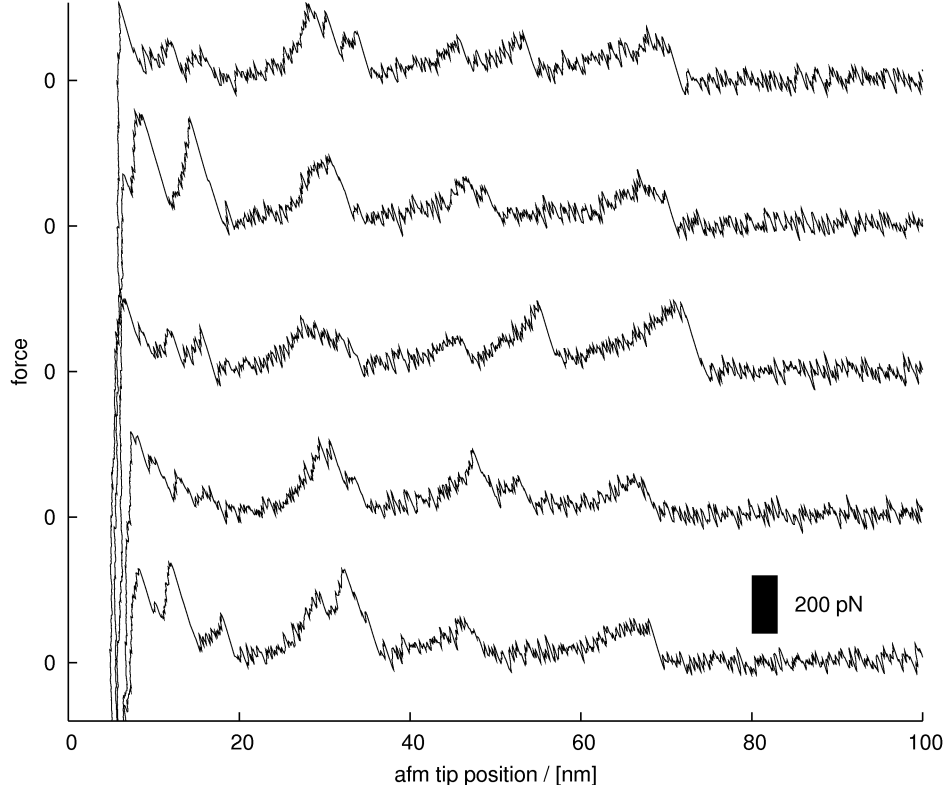


Figure 5.2: Several example force curves from the AFM experiments. The zero force values of each curve are marked on the y -axis.

workers [2] obtained a superposition of all force profiles which also revealed four force peaks, similar to the MD force profiles. The main difference between the AFM and the FPMD force profiles were the heights of the single force peaks, which were due to the higher pulling velocities in the FPMD simulations.

For a more quantitative comparison of the forces we plotted the four peak forces from the AFM experiments and the simulations over the applied pulling velocities and fitted according to

$$F_{max}(v) = \gamma v + \text{const} \ln(v/v_0) \quad (5.1)$$

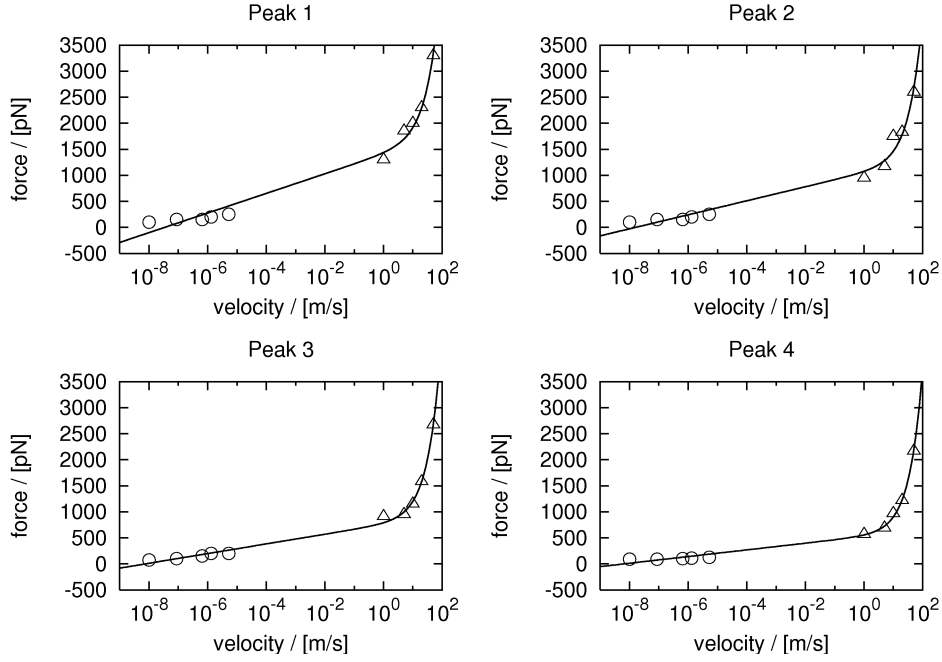


Figure 5.3: Comparison of velocity dependent peak forces from AFM experiments and FPMD simulations. AFM peak forces are plotted as circles and FPMD peak forces are plotted as triangles, respectively. The black line is the force fit according to equation 5.1.

[16], where γ is a friction coefficient (figure 5.3).

The fit was in good agreement with the different peak forces. However, the slope of the fitted curve at velocities the AFM regime was too steep. There are two possible explanations for the negative forces. First, the model proposed by Heymann and Grubmüller may not be suitable to describe the unfolding forces occurring during the extraction of one bacteriorhodopsin monomer. This model deals with the crossing of one free energy barrier, which is appropriate for single activation events, like the unbinding of a ligand from a protein. Bacteriorhodopsin, however, is a closely packed helical structure and it is likely, that the unfolding of the protein implies the crossing of sev-

eral free energy barriers. Second, the unfolding pathway of the protein – and therefore also the occurring forces – in the AFM experiments may be different from that observed in the simulations. The only way to prove this would be FPMD simulations with even slower pulling velocities.

5.2 Unfolding Pathway

In the AFM experiments one single protein was extracted from the purple membrane. After extraction of that protein, the remaining purple membrane stayed intact without any distortions. To test to which extent the extraction affected the adjacent monomers, figure 5.4 shows the root mean square deviations (RMSD) of the backbone atoms of each of the eleven non-extracted proteins in our simulation system.

All RMSD curves did not show strong deviations and stayed below 0.2 nm, which is in the range of normal fluctuations for proteins. Distorted proteins would have had higher a RMSD. These values indicate that the non-extracted proteins were not altered directly by the force applied to the extrated protein, which is remarkable, because no position restraints were applied to the system. It is a good indicator that our model system resembles the original purple membrane, because in the AFM experiments the purple membrane also stayed intact during extraction.

To investigate the unfolding pathway, we examined the structural changes of the single helices during extraction. To this end, the root mean square deviations of atom distances (dRMSD) of the backbone atoms of each helix were determined separately (figure 5.5).

dRMSD values near 0 nm correspond to an intact helical structure, whereas the steep increase seen in each of the curves indicates a loss of secondary structure. For each helix a sudden change from zero values to increased values was seen in the dRMSD curves, thus marking the onset of unfolding. The successive single unfolding events were separated from each other with no visible overlap. Hence, each helix unfolded independently, and sequently

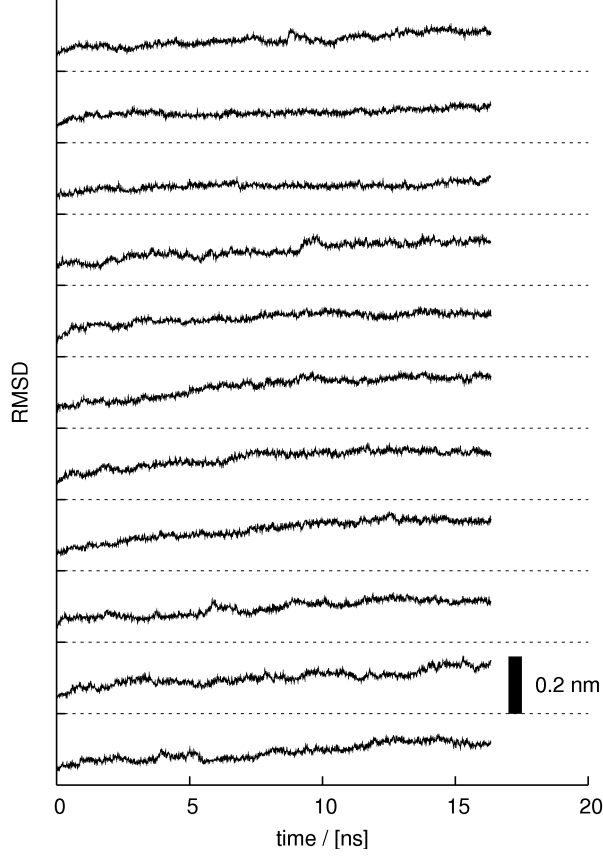


Figure 5.4: RMSD values of the backbone atoms of each of the eleven remaining proteins. The values were taken from the FPMD simulation with pulling velocity $v = 0.05 \text{ nm/ps}$ towards the cytoplasmic side.

(Helix G \rightarrow Helix F $\rightarrow \dots \rightarrow$ Helix A).

Our results suggest that at least at the fast MD timescale, the seven trans-membrane helices of bacteriorhodopsin behave as independent folding units. This is in accordance with the two-stage model proposed by Popot and Engelman [14], where trans-membrane α -helices are described as autonomous stable folding domains. The model describes the folding process of membrane proteins consisting of trans-membrane α -helices in two independent steps.

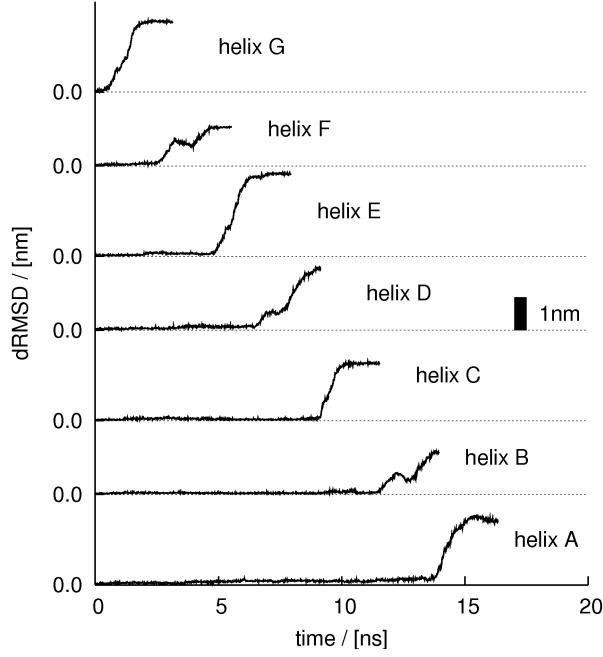


Figure 5.5: dRMSD values of the backbone atoms of each of the seven trans-membrane α -helices during unfolding. The values were taken from the FPMD simulation with pulling velocity $v = 0.05$ nm/ps towards the cytoplasmic side.

The first stage consists of the formation of single membrane-spanning α -helices, which are not yet connected to each other. The second stage of the process is the aggregation of these helices. This process is moderated by peptide links connecting helices, van-der-Waals forces between helices, van-der-Waals forces between helices and lipid molecules, polar interactions between helices, or hydrogen bonds between helices.

To gain further insight into the unfolding and extraction of the helices, the Coulombic and van-der-Waals energies of interaction between the seven helices and selected parts of the system were calculated. These interactions included (i) interactions of the atoms within a helix, (ii) interactions between one helix and the other helices of the protein, (iii) interactions between each

helix and the lipid molecules in the membrane, and (iv) interactions between each helix and the non-extracted proteins.

First, we examined the differences of the interaction energy at the starting time and the energy at the end of the simulation for the four types of interaction (i-iv). Each column in figure 5.6 shows the sum of the respective energy differences over all helices. The energy of the interaction between one helix and the other helices within the protein was calculated according to

$$E_{total} = \frac{1}{2} \sum_{i \neq j} E(\text{helix } i, \text{helix } j), \quad (5.2)$$

where $E(\text{helix } i, \text{helix } j)$ denotes the energies of interaction between helix i and helix j .

The largest individual energy contribution to the change in energy was caused by the interactions of the atoms within each helix. Nevertheless, the sum of the other three contributions was larger. Hence, in the purple membrane it is energetically preferable for the protein to unfold its helices one by one rather than unfolding several helices at the same time. Considering solely the coulombic energy changes, most interestingly, the largest contribution was provided by the interaction within the helices.

These energy calculations can only provide a rough estimate of the energetics during extraction. First, the role of the water environment was neglected in our estimate. Furthermore, a more qualitative analysis would also require the calculation of free energy changes.

As a next step, we determined the time development of the overall energy changes for all seven helices (figure 5.7). Again, the differences between the start and the end value were determined. The interaction between the considered helix and the lipid molecules as well as the non-extracted proteins were

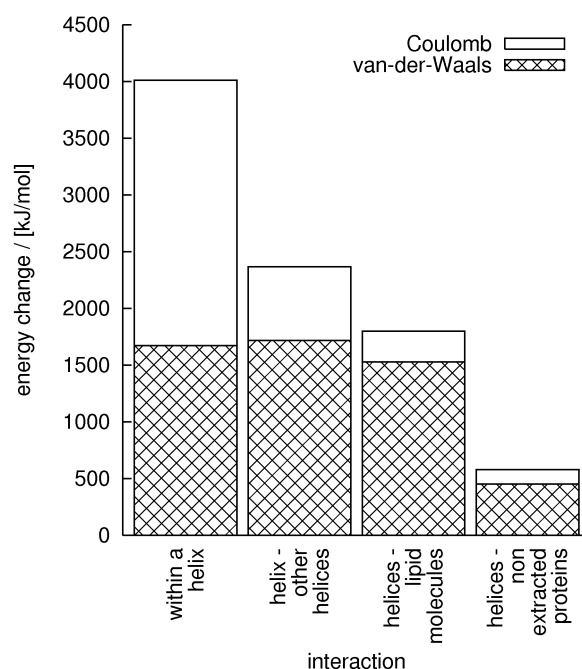


Figure 5.6: Different interaction energies regarding the seven trans-membrane α -helices. Columns show the difference of the interaction energy at the starting time and at the end of the simulation for the considered types of interactions.

summed up (figure 5.7, black areas named "purple membrane"). Grey areas show the interaction between the considered helix and the non-extracted helices, and hatched areas show the interactions of the atoms within each helix.

The interaction between one helix and the other helices was calculated different from the previous approach employed for figure 5.6. Here, only helices that remain in the membrane during extraction of the considered helix were taken into account. For example, for helix D the interactions with helices A, B, and C were calculated, but not interactions with helices E, F, and G. This is also the reason why for helix A interactions with other helices were not calculated.

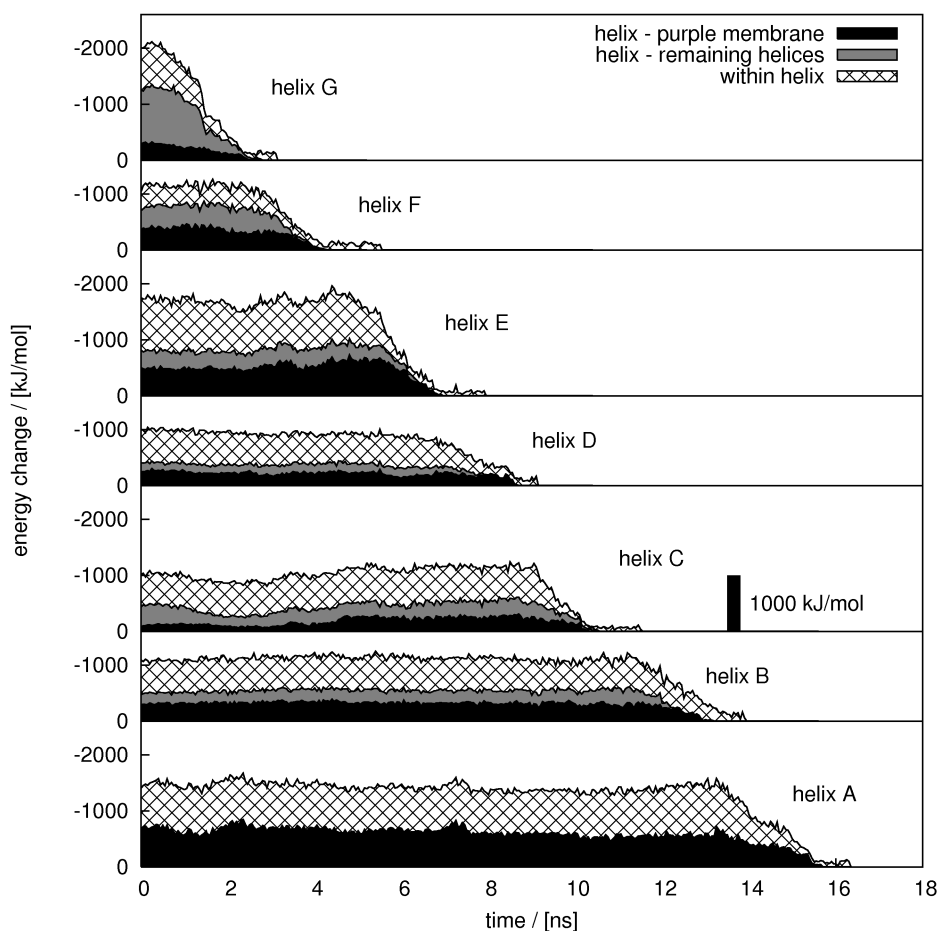


Figure 5.7: Time development of energy regarding the seven trans-membrane α -helices. All curves show the sum of the changes in interaction energies for the considered helix and are split up according to the contribution of the single interaction types.

In general, the total energy changes for helices resulting in a force minimum during extraction (B, D, and F) in the force profiles were smaller than for helices resulting in a force maximum (A, C, E, and G). The only exception was helix C, where the energy change was of the order of the values from the helices resulting in a force minima.

All energy curves had a similar shape. Before the unfolding of a helix, its energy values stayed stable. During extraction the energy increased until it reached its maximal value. The unfolding of a helix did not lead to energy changes for the remaining helices, which further corroborates the helix-wise unfolding of the protein independent of the state of the other helices.

To study the unfolding of single helices, the snapshots presented in figure 5.8 cover the unfolding of helices G and F during the pulling towards the cytoplasmic side. The corresponding forces are shown in figure 5.9. Unfolding of these helices serves as a representative example for the unfolding of all seven helices of the protein. All other helices unfolded in a similar fashion.

The unfolding snapshots reveal that the helices unfold in a spiral-like fashion. Helix G unfolded from top to bottom (figure 5.8 A to C). During unfolding the force increased and reached its maximal value once the helix was completely unfolded (figure 5.9, first three circles). Helix F unfolded in the opposite direction (bottom to top, see figure 5.8 D to F) and accordingly the force dropped during the unfolding and reached its minimal value when the helix had lost all secondary structure. The last part of helix F was kinked and moved out into the bulk water where it further unfolded (figure 5.8 E).

The first three helix pairs (GF, ED, and BC) all unfolded in this fashion and pairwise induced a force maximum followed by a minimum during un-

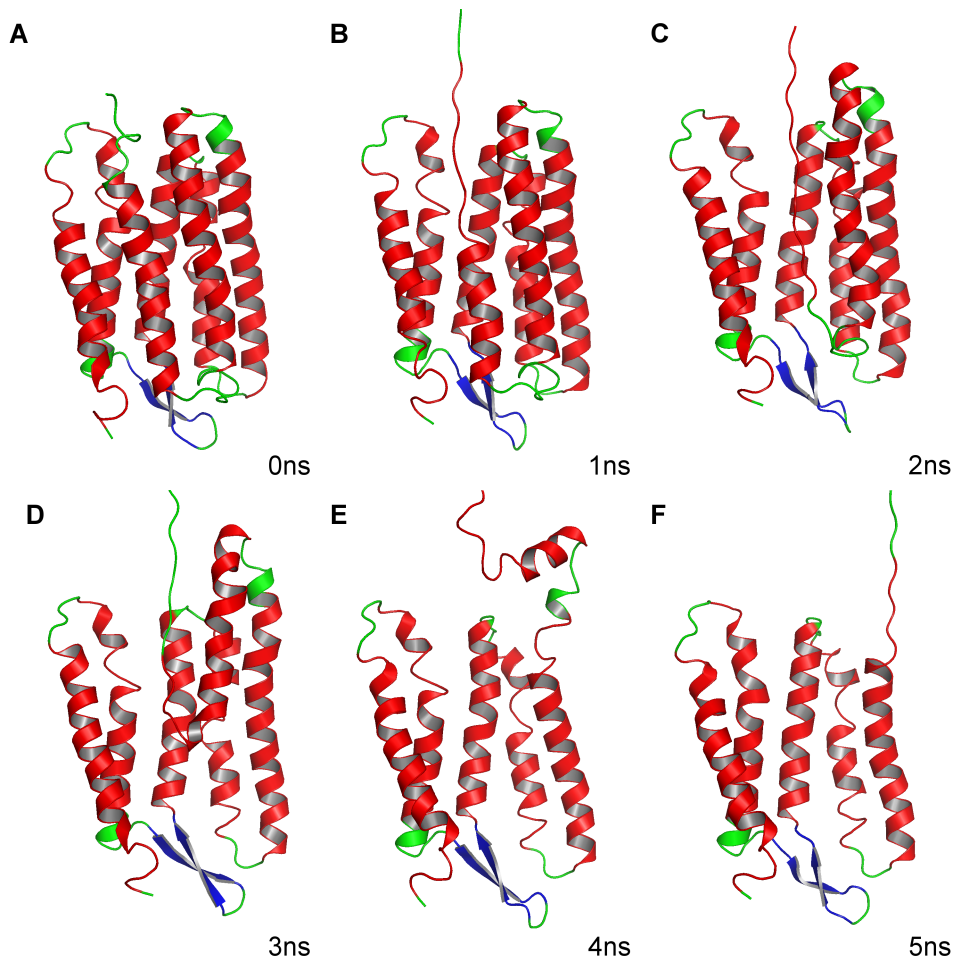


Figure 5.8: Unfolding and extraction of helices G and F. Snapshots during unfolding. Snapshot times are given in the lower right corner of each picture. The colors represent secondary structure elements of the crystal structure [22]. Red corresponds to α -helices, blue to β -sheets, and green to loops.

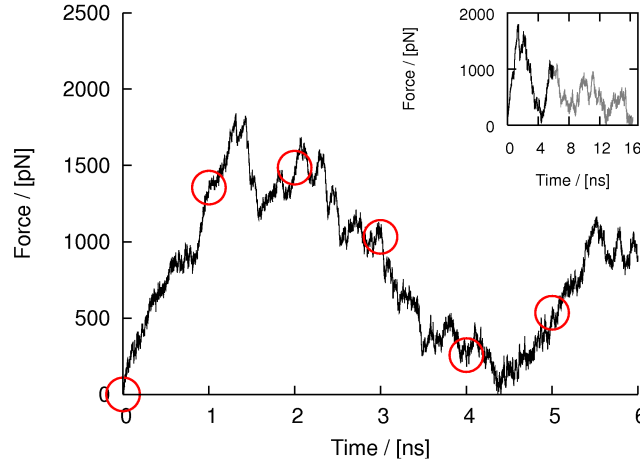


Figure 5.9: The main plot shows the force profile covering the time of snapshots in figure (5.8 A to F). The red circles denote the time of the snapshots. The inset shows the complete force profile where the black part corresponds to the main plot.

folding. The last helix (helix A) induced a force maximum on its own. This behaviour was observed in all unfolding simulations. For the pulling towards the extracellular side the process was similar, except that the order of the helix pairs was reversed (AB, CD, EF, and G as single peak).

The two subpeaks in the force profiles resulting from pulling towards the cytoplasmic side (see figure 5.1, first and third maximum) can also be interpreted at atomic level. The first subpeak is correlated to the extraction of the retinal out of the protein core caused by the unfolding of helix G (figure 5.10). The second subpeak arised when the unfolded β -sheets were moved through the lipid bilayer after the unfolding of helix C (figure 5.11).

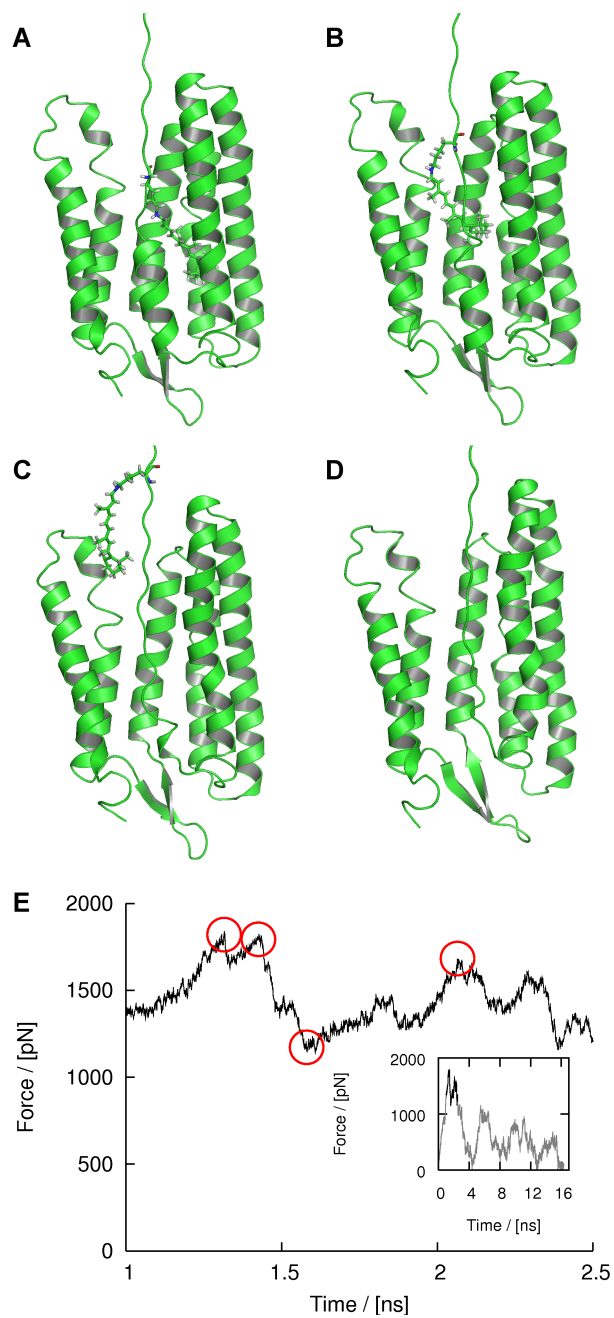


Figure 5.10: First sub peak from the pulling towards the cytoplasmic side. (A-D) Snapshots during unfolding. The retinal is drawn in a stick representation. (E) The main plot shows the force profile covering the time of pictures (A-D). Rest like in figure 5.8.

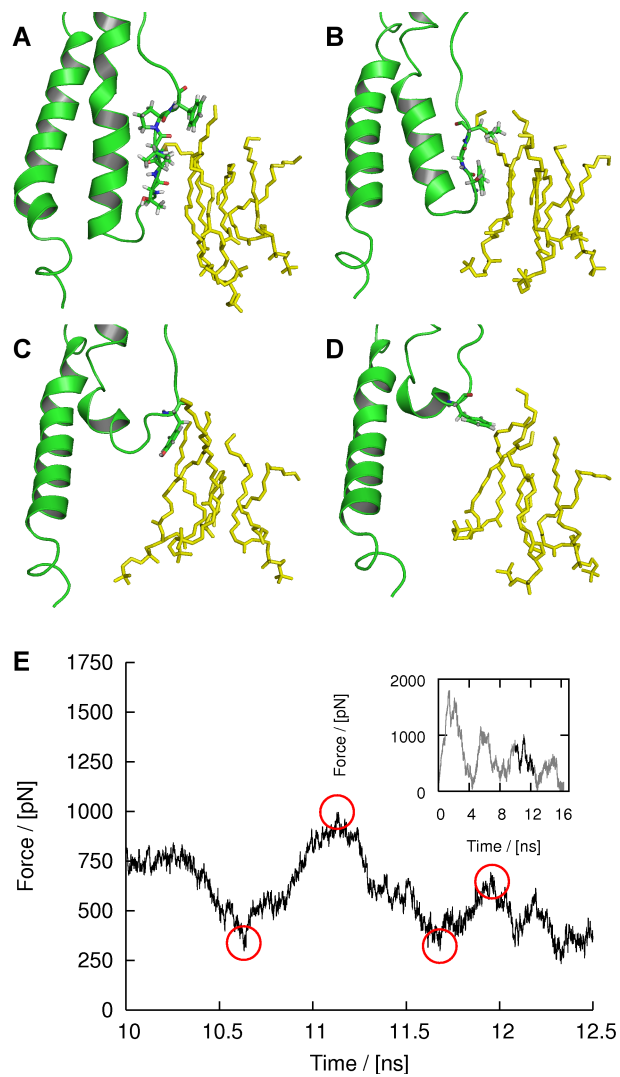


Figure 5.11: Second sub peak from the pulling towards the cytoplasmic side. (A-D) Snapshots during unfolding. Lipid molecules are drawn in a stick representation and in yellow. Residues interacting with lipid molecules are also drawn in a stick representation. (E) The main plot shows the force profile covering the time of pictures (A-D). Rest like in figure 5.8.

5.3 Positions of Unfolding Barriers

Apart from the large scale structural changes visible in the snapshots, investigation of transition points together with a more detailed examination of the unfolding process also allows insights into the mechanical stability of the protein. Transition points were determined for the pulling simulations with pulling velocities 0.005 nm/ps, 0.01 nm/ps, 0.02 nm/ps, and 0.05 nm/ps for both pulling directions, summing up to 8 data sets. Each histogram of transition points from each data set was normalized to 1. Then, the four normalized data sets from each pulling direction were summed up and the resulting data set was also normalized to 1 (figure 5.12 top and bottom). The results from both pulling directions revealed a few dominant peaks, indicating residues that resist unfolding with a high probability. These dominant peaks were defined as unfolding barriers (plotted in black in figure 5.12 for a better visualisation).

As can be seen, the unfolding barriers from the pulling towards the cytoplasmic side are preferentially located at the borders of the helices. Unfolding barriers within the helices correspond to a kink where helices were moved partially out of the membrane (see figure 5.8E).

Using a different approach, unfolding barriers were also determined from the AFM experiments. There, unfolding barriers were determined by fitting a worm-like-chain model force-extension curve

$$F_{wlc} = \frac{k_B T}{p} \left(\frac{1}{4(1 - x/L)^2} - \frac{1}{4} + \frac{x}{L} \right) \quad (5.3)$$

[32] to the superposition of the force profiles with force F_{wlc} , Boltzmann constant k_B , temperature T , persistence length p , extension x , and length of the peptide chain L . The only free parameter is the length L of the peptide

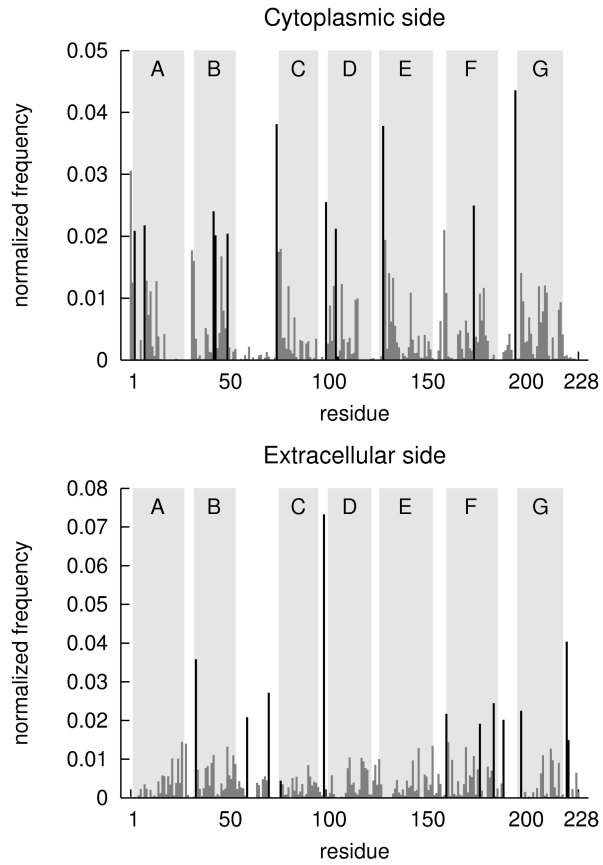


Figure 5.12: Histogram of the occurrence of transition points. The upper plot shows the distribution from the pulling towards the cytoplasmic and the lower plot shows the distribution from the pulling towards the extracellular side. Dominant peaks are plotted black. Helix regions for helices A-G are marked with grey bars.

chain. It is possible to infer the position of an unfolding barrier in the peptide chain from the length L .

To test whether the location of the unfolding barriers determined in the AFM experiments and the simulations agree with each other, figure 5.13 shows the positions of the unfolding barriers resulting from both methods. Unfolding barriers derived from both AFM experiments and simulations are plotted as filled symbols.

For the pulling towards the cytoplasmic side, three matches were found: Asp104, Tyr133, and Pro200. For the extracellular pulling direction five matches are visible: Gln75, Ala81, Ala103, Pro165, and Trp189. The other unfolding barriers from both pulling directions show no correlation.

Despite the large difference in pulling velocities, in both AFM experiments and FPMD simulations matching unfolding barriers can be found. The matching unfolding barriers therefore represent important regions in the energy landscape of the protein.

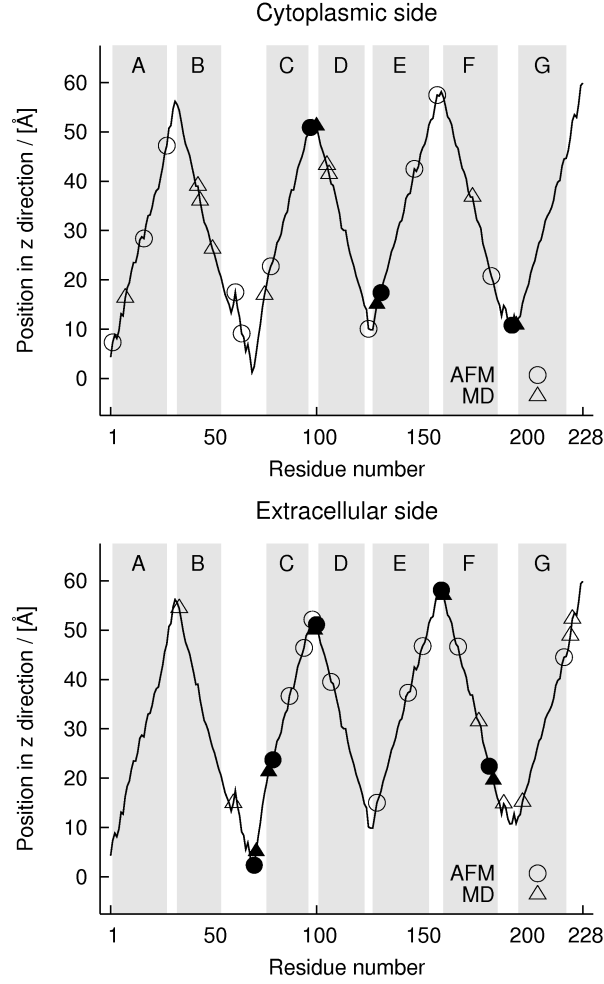


Figure 5.13: Comparison between unfolding barriers from AFM experiments and FPMD simulations. The black line connects the z -positions of the C_{α} -atoms of all residues. Unfolding barriers are plotted as circles (AFM experiments [8]) or triangles (FPMD simulations). Filled symbols indicate matching unfolding barriers. The grey bars denote helix regions.

5.4 Interrupted Pulling

The results from the AFM experiments suggested that two helices unfold simultaneously during extraction of one bacteriorhodopsin monomer. This unfolding pathway is not observed in the MD simulations. One reason for this observation may be the very slow AFM pulling velocities. On an MD timescale, these movements are quasi-equilibrium.

Although the previous results elucidate the dynamics of the unfolding process, it is impossible to draw conclusions about how far the protein was away from equilibrium during extraction. Further insight into the protein behaviour is provided by simulations with an interrupted spring movement. In such a situation, the relaxation of the protein towards equilibrium can be studied.

The main results from both simulations are shown in figure 5.14 (6 nm spring elongation, helix G partially unfolded) and figure 5.15 (16 nm spring elongation, helix F partially unfolded). In both figures the forces (figure 5.14 A, C and 5.15A, C) and the dRMSD values of the partially unfolded helices (figure 5.14 B, D and 5.15B, D) are shown. In both cases, different phases can be distinguished in the time development of these observables.

During the first phase (start until beginning of grey region) the forces and dRMSD values were subjected to strong and very fast changes. Subsequently, a short relaxation period was visible (grey area in figure 5.14 and 5.15), which can be described as a drift period. In this period, the forces and dRMSD values showed a damping behaviour. In the next phase (end of grey region until end of plots) no changes in the protein conformation were visible and the forces fluctuated around a constant value.

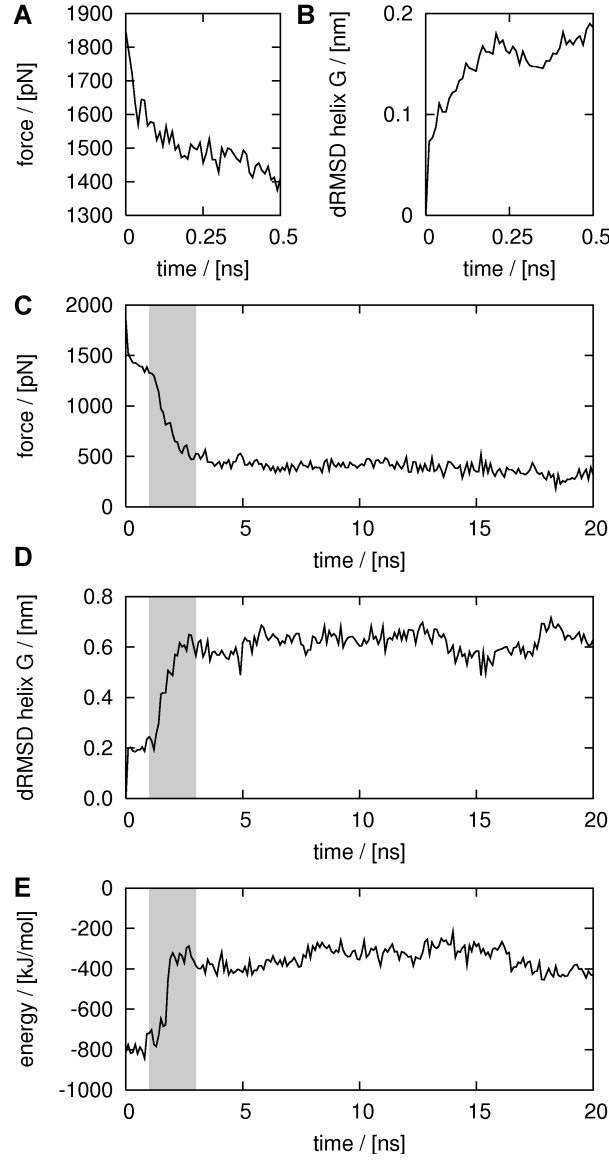


Figure 5.14: Results for the simulation with interrupted spring movement at 6 nm spring position. The grey bars denote the region of major changes. (A, C) Force profile. (B, D) dRMSD values of the backbone atoms of helix G. (E) Interaction energy (Coulomb and van-der-Waals) between helix G and the other six helices.

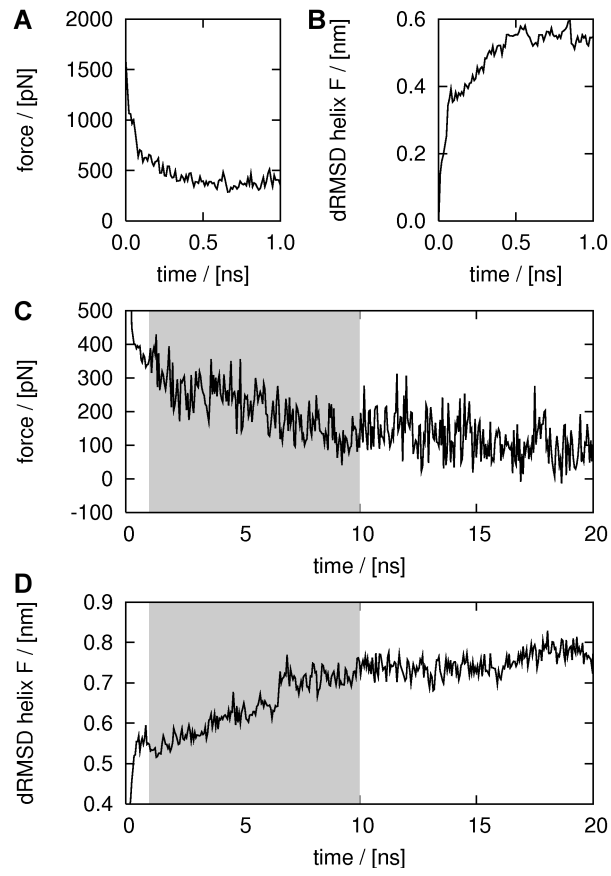


Figure 5.15: Results for the simulation with interrupted spring movement at 16 nm spring position.

In the first simulation (helix G partially unfolded, spring position 6 nm, figure 5.14) the drift phase lasted 2 ns. Here, the force change was followed by sudden energy changes (see grey area in figure 5.14 E). The retinal was moved out of the protein core. Subsequently, helix G was shifted in pulling direction and in this way lost contact to the other helices (snapshots are shown in figure 5.16). As a consequence, the interaction energy between helix G and the other helices increased. The drift time in the second simulation lasted 9 ns (figure 5.15), and no key event during unfolding was visible.

The forces acting on the protein after 20 ns simulation were significant smaller than the forces at the beginning of the simulations. For these large force differences, two reason are possible.

The first reason may be a large friction force occuring due the fast movement of the peptide chain. An instantaneous missing of a friction force would explain the very fast force changes in the first 100 ps in both simulations.

The second reason may be that the applied force is too strong to be captured through conformational changes within the protein. This situation can be compared to a driven oscillator after switching off the external force. In that case, the oscillating system starts to perform damped motions, like the ones visible during the drift phase.

The forces acting on the protein at the end of the 20 ns simulations are necessary to keep the protein in its enforced conformation. The unfolding pathway did not differ from that with a moving spring. Like in the normal FPMD simulations, folded helices remained stable and only the already partially unfolded helices were further unfolded.

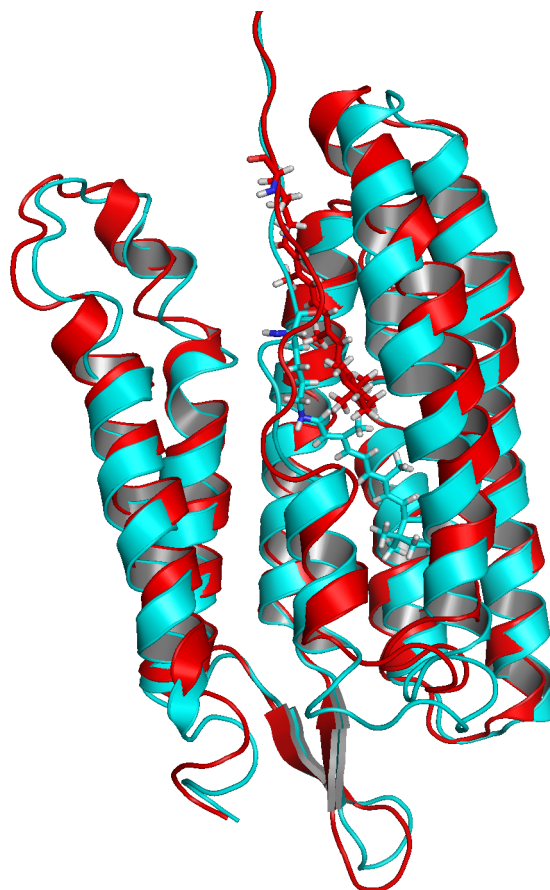


Figure 5.16: Snapshots from the simulation with interrupted spring movement at 6 nm spring position. The cyan snapshot was taken at 1 ns, and the red snapshot was taken at 3 ns. The retinal is drawn in a stick representation.

Chapter 6

Summary

In this work, force probe molecular dynamics simulations of the enforced unfolding of the membrane protein bacteriorhodopsin are presented. The project was inspired by AFM unfolding experiments of this protein, which were recently carried out in the groups of Hermann Gaub in Munich and Daniel Müller in Dresden.

The goal of this work was to simulate AFM measurements using MD techniques, to compare the results with the experiments, and to gain further insight into the enforced unfolding process at an atomic level. The realisation and analysis of this work required the development and implementation of several new methods.

Several unfolding simulations were carried out on bacteriorhodopsin. The protein was unfolded from both ends, using different pulling velocities. A new method was developed to allow the complete unfolding of the large protein, while at the same time keeping the computational effort tractable. During the extraction, the protein was truncated repeatedly and the pulling process was continued with the shorter peptide chain. All simulations resulted in the

complete extraction of the protein.

In our simulations the anchoring membrane stayed stable during extraction, as observed in the AFM experiments. The corresponding force profiles showed four significant peaks and also resembled the results obtained in the AFM measurements. Thus, the simulations were able to describe the AFM experiments.

From the simulation trajectories, it was possible to determine the unfolding pathway of the protein. In all FPMD simulations the protein unfolded helix-wise. This finding differs from the unfolding pathway that was inferred from the AFM experiments [2, 3, 4, 5, 6, 7, 8, 9]. The different pathway could be due to the much faster pulling velocities in the simulations. However, an estimation of the energies of interaction within the membrane system further corroborated a helix-wise unfolding. The unfolding pathway, resulting from the simulations, was correlated with the force profiles, allowing additionally to interpret subpeaks in the force profiles at an atomistic level.

The locations of unfolding barriers, stable residues within the protein, were determined using a new method developed in this work. Despite large differences in pulling velocities, some locations of the unfolding barriers from the AFM experiments were reproduced in the simulations. These residues were Asp104, Tyr133, and Pro200 when pulling towards the cytoplasmic side, and residues Gln75, Ala81, Ala103, Pro165, and Trp189 when pulling towards the extracellular side.

To explore how far the unfolded protein had moved away from equilibrium, simulations with a static spring position were carried out. In this way, the slow movement of an AFM cantilever was imitated. The results

showed that the protein was far away from an equilibrium state during the unfolding simulations. Nevertheless it is still possible to deduce equilibrium information from a sufficiently large set of non-equilibrium data, like shown by Jarzynski [33].

Alltogether it was shown in this work, that AFM experiments enforcing the unfolding of bacteriorhodopsin could be successfully reproduced by FPMD simulations. This method can not only be used to reproduce AFM experiments, but also allows further insight into the unfolding pathway of the protein under an externally applied force at an atomic level. Snapshots taken from the trajectories give information about intermediate unfolding states. The stability of the protein can be further examined by the location of unfolding barriers, which were determined and compared to the locations of unfolding barriers from the AFM measurements.

Thanks

Many thanks to Helmut Grubmüller for the overwhelming support, my whole department for help, when help was needed, Ira Tremmel for helping to make this diploma thesis from a sketch book to a masterpiece, Ulrich Zachariae for steady help during my work, Hermann Gaub and Max Kessler for the cooperation, and Tim Salditt for being the main corrector of this work.

Bibliography

- [1] Jeremy M. Berg, John K. Tymoczko, and Lubert Stryer. *Biochemie*. Spektrum Akademischer Verlag, 2003.
- [2] F. Oesterhelt, D. Oesterhelt, M. Pfeiffer, A. Engel, H. E. Gaub, and D. J. Müller. Unfolding pathways of individual bacteriorhodopsins. *Science*, 288:143–146, 2000.
- [3] Daniel J. Müller, Max Kessler, Filipp Oesterhelt, Clemens Müller, Dieter Oesterhelt, and Hermann Gaub. Stability of bacteriorhodopsin α -helices and loops analyzed by single-molecule force spectroscopy. *Biophysical Journal*, 83:3578–3588, 2002.
- [4] Harald Janovjak, Max Kessler, Dieter Oesterhelt, Hermann Gaub, and Daniel J. Müller. Unfolding pathway of native bacteriorhodopsin depend on temperature. *The EMBO Journal*, 22:5520–5229, 2003.
- [5] Harald Janovjak, Jens Struckmeir, Maurice Hubain, Alexej Kedrov, Max Kessler, and Daniel J. Müller. Probing the energy landscape of the membrane protein bacteriorhodopsin. *Structure*, 12:871–879, 2004.

- [6] David A. Cisneros, Dieter Oesterhelt, and Daniel J. Müller. Probing origins of molecular interactions stabilizing the membrane proteins halorhodopsin and bacteriorhodopsin. *Structure*, 13:235–242, 2005.
- [7] Harald Janovjak, Daniel J. Müller, and Andrew D. L. Humphris. Molecular force modulation spectroscopy revealing the dynamic response of single bacteriorhodopsins. *Biophysical Journal*, 88:1423–1431, 2005.
- [8] Max Kessler and Hermann Gaub. Unfolding barriers in bacteriorhodopsin probed from the cytoplasmic and the extracellular side by AFM. *Structure*, 14:521–527, 2006.
- [9] Max Kessler, Kay E. Gottschalk, Harald Janovjak, Daniel J. Müller, and Hermann E. Gaub. Bacteriorhodopsin folds into the membrane against an external force. *Journal of Molecular Biology*, 357:644–654, 2006.
- [10] Ulrich Haupts, Jörg Tittor, and Dieter Oesterhelt. Closing in on bacteriorhodopsin: Progress in understanding the molecule. *Annual Review of Biophysics and Biomolecular Structure*, 28:367–399, 1999.
- [11] Helmut Grubmüller, Berthold Heymann, and Paul Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271:997–999, 1996.
- [12] S. J. Singer and Garth L. Nicolson. The fluid mosaic model of the structure of cell membranes. *Science*, 175:720–731, 1972.
- [13] D. M. Engelman, T. A. Steitz, and A. Goldman. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins.

- Annual Review of Biophysics and Biophysical Chemistry*, 15:321–353, 1986.
- [14] J.-L. Popot and D. M. Engelman. Membrane protein folding and oligomerization: The two-stage model. *Biochemistry*, 29:4031–4037, 1990.
 - [15] G. Binnig, C. F. Quate, and Ch. Gerber. Atomic force microscopy. *Physical Review Letters*, 56:930–933, 1986.
 - [16] Berthold Heymann and Helmut Grubmüller. AN02/DNP-hapten unbinding forces studied by molecular dynamics atomic force microscopy simulations. *Chemical Physics Letters*, 303:1–9, 1999.
 - [17] Gunnar Schröder. Molekulardynamiksimulation der Flexibilität und Fluoreszenzanisotropie eines an ein Protein gebundenen Farbstoffs. Master’s thesis, MPIbpc Göttingen, 2000.
 - [18] R. W. Hockney and S. P. Goel. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics*, 14:148–158, 1972.
 - [19] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics*, 81:3684–3690, 1984.
 - [20] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: an $n \cdot \log(n)$ method for ewald sums in large systems. *Journal of Chemical Physics*, 98:10089–10092, 1993.

- [21] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, , and L. G. Pedersen. A smooth particle mesh ewald potential. *Journal of Chemical Physics*, 103:8577–8592, 1995.
- [22] Hassan Belrhali, Peter Nollert, Antoine Royant, Christoph Menzel, Jurg P Rosenbusch, Ehud M Landau, and Eva Pebay-Peyroula. Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular vire of the purple membrane at 1.9Å resolution. *Structure*, 7:909–917, 1999.
- [23] G. Vriend. What if: a molecular modeling and drug design program. *Journal of Molecular Graphics and Modelling*, 8:52–56, 1990.
- [24] José D. Faraldo-Gómez, Graham R. Smith, and Mark S. P. Sansom. Setting up and optimization of membrane protein simulations. *European Biophysics Journal*, 31:217–227, 2002.
- [25] Erik Lindahl, Berk Hess, and David van der Spoel. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 7:306–317, 2001.
- [26] W. L. Jorgensen and J. Tirado-Rives. The OPLS potential functions for proteins - energy minimizations for crystals of cyclic-peptides and cambin. *Journal of the American Chemical Society*, 110:1657–1666, 1988.
- [27] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *Journal of Computational Physics*, 79:926–935, 1983.

- [28] O. Berger, O. Edholm, and F. Jähnig. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical Journal*, 72:2002–2013, 1997.
- [29] Christian Kandt, Jürgen Schlitter, and Klaus Gerwert. Dynamics of water molecules in the bacteriorhodopsin trimer in explicit lipid/water environment. *Biophysical Journal*, 86:705–717, 2004.
- [30] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18:1463–1172, 1997.
- [31] E. Espinosa, E. Molins, and C. Lecomte. Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. *Chemical Physics Letters*, 285:170–173, 1998.
- [32] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of λ -phage DNA. *Science*, 265:1599–1600, 1994.
- [33] C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78:2690–2693, 1997.