

Structure and dynamics of proteins under force

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

“Doctor rerum naturalium”

der mathematisch-naturwissenschaftlichen Fakultäten

der Georg-August-Universität zu Göttingen

Vorgelegt von Christian Kappel aus Kassel

Göttingen, 2011

Thesis Committee:

Prof. Dr. Helmut Grubmüller (Reviewer)

Prof. Dr. Ralf Ficner (Reviewer)

Prof. Dr. Marcus Müller

Tag der mündlichen Prüfung:

31. März 2011

Preface

This thesis is a cumulative PhD thesis. It is based on the following three publications and manuscripts, respectively:

- Christian Kappel and Helmut Grubmüller, Velocity-Dependent Mechanical Unfolding of Bacteriorhodopsin Is Governed by a Dynamic Interaction Network, *Biophysical Journal*, 100:1109-1119 (2011)
- Christian Kappel, Ulrich Zachariae, Nicole Dölker, and Helmut Grubmüller, An Unusual Hydrophobic Core Confers Extreme Flexibility to HEAT Repeat Proteins *Biophysical Journal*, 99:1596-1603 (2010)
- Christian Kappel, Nicole Dölker, Mareike Zink, Ulrich Zachariae, and Helmut Grubmüller, Non-equilibrium elastic behavior of macromolecules, in preparation

Contents

1. Introduction	1
2. Molecular dynamics simulations	5
3. Background of the publications	9
3.1. Velocity-Dependent Unfolding of Bacteriorhodopsin	9
3.2. Extreme Reversible Elasticity of Importin- β	11
3.3. Non-equilibrium elastic behavior of macromolecules	15
4. Summary and Conclusion	35
4.1. General	35
4.2. Velocity-Dependent Unfolding of Bacteriorhodopsin	36
4.3. Extreme Reversible Elasticity of Importin- β	38
4.4. Non-equilibrium elastic behavior of macromolecules	41
5. Declaration of own contribution	43
6. Acknowledgements	45
Bibliography	47
A. Velocity-Dependent Unfolding of Bacteriorhodopsin	53
B. Extreme Reversible Elasticity of Importin-β	67

1. Introduction

Proteins are one of the fundamental pillars of all known life. They serve, e.g., mechanical stability, selective transport of ions and small molecules across membranes, and energy conversion. Each protein acquires its unique three-dimensional structure in a process called folding [3].

The interplay between the structure and the dynamics of each protein determines its function. Depending on inherent structural rigidity and flexibility, several examples can be distinguished. Proteins that serve to stabilize supramolecular structures are characterized by a very rigid structure. Notable examples are the hair building protein collagen [4] and proteins that are the building blocks of outer viral shells [5]. The function of other proteins requires them to change between different well-defined conformations, thus requiring both rigid and flexible parts. In this case, rigid parts of such proteins are allowed to rearrange with respect to each other due to the flexible parts that connect them. Examples are lysozyme, where two domains connected by a flexible helix move relative to each other [6], and rhodopsin, where light-induced isomerization of a cofactor leads to conformational changes [7]. The function of proteins can also solely arise from their global flexibility. Examples are intrinsically disordered proteins that lack any kind of well-defined structure, like FG repeats within nuclear pore complexes [8]. Also nuclear transport receptors, whose structures are required to undergo large conformational changes in order to transport cargo molecules, belong to this group [9].

The structure of proteins can be determined in atomic detail by, e.g., x-ray crystallography and nuclear magnetic resonance. Also, from both methods information about protein flexibility can be derived. However, to study atomistic dynamics of proteins, one needs to employ molecular dynamics (MD) simulations. This computational method allows the numerical calculation of the development of atomic motions within a macromolecular system, yielding trajectories of all atoms.

A new understanding of proteins from a mechanical point of view was introduced in the 1980s by the development of atomic force microscopy (AFM) [10] and optical tweezers (OT) [11, 12]. For the first time, these methods allowed the direct manipulation of single proteins under external forces.

A prominent application of these methods is mechanical unfolding of proteins. Such experiments quantify the structural stability of proteins. The results from these experiments raise an understanding of how proteins fold into their well-defined structure by identifying folding units and key residues that stabilize the global protein structure. The unfolding of a broad range of both soluble proteins such as titin

[13], green fluorescent protein [14], and lysozyme [15] as well as membrane bound proteins such as bacteriorhodopsin [16], aquaporin [17] and rhodopsin [18] have been studied.

Another application of both methods is the quantification of elastic properties of proteins. Recent experiments determined the stiffness of viral outer shells [19] or ankyrin repeats [20]. This knowledge is crucial since the function of these proteins depends on their elasticity. AFM experiments were also employed to enforce the unbinding of ligands from their respective receptor [21].

A connection between AFM and OT experiments and MD simulations is provided by force probe molecular dynamics (FPMD) simulations [22, 23]. In this method, the influence of an external force is mimicked by a moving harmonic potential also referred to as “virtual spring”. Like in AFM and OT experiments, forces that are acting on single proteins are monitored. In addition to the experimental methods, FPMD simulations provide direct insight into the atomic processes governing these forces, thus revealing atomic causes behind the measured forces. In this way, e.g., the unfolding of titin domains [24], the function of titin kinase [25], and the elasticity of ankyrin repeats [26] were studied.

The current thesis consists of three parts that cover the diverse behavior of proteins seen in FPMD simulations. In the first two parts, FPMD simulations were applied to study the unfolding of membrane proteins as well as the elastic response of flexible proteins. The third part investigates non-equilibrium behavior of macromolecules under force at the short time scales present in FPMD simulations.

In the first part of this work (chapter 3.1 and reference [1]), the mechanical unfolding of the membrane protein bacteriorhodopsin was studied. AFM unfolding experiments on this protein characterized it as very stable and revealed that its stability is governed by a small number of key residues [16, 27]. The aim of this part is to explain the origin of the mechanical stability of bacteriorhodopsin from an atomistic point of view. To this end, the unfolding pathway of the protein from the FPMD simulations was thoroughly analyzed. In this process, a new analysis method for the identification of the key residues that provide mechanical stability to bacteriorhodopsin was developed and applied. Subsequent analyses of the atomic interactions that stabilize the protein revealed that these are forming a dynamic network. This dynamic interaction network explained why unfolding pathways of bacteriorhodopsin are velocity-dependent.

The second part of this work (chapter 3.2 and reference [2]) investigated the flexibility of the nuclear transport receptor importin- β . The aim of this part is to explain and quantify its flexibility. Both from different x-ray structures as well as from MD simulations of this protein an unusually large degree of flexibility was observed [9, 28]. To characterize and to quantify the flexibility of the protein, stretching and relaxation simulations on importin- β were carried out. The results from these simulations quantified the amount of binding energy that is stored by deformation of the protein. To determine the structural and energetic origins of the flexibility of

importin- β , the dynamics of the hydrophobic core of the protein were analyzed and set in relation to more rigid proteins.

The observation of a previously not described non-equilibrium elastic behavior of importin- β and of the outer shell of the southern bean mosaic virus [29] was the starting point for the third part of this work (chapter 3.3). Both systems exhibited velocity-dependent spring constants during elastic deformation in FPMD simulations. The aim of this part was to explain and quantify this behavior, that is not explained by any established model [23, 30–33]. Thus, a new model for non-equilibrium elastic behavior of macromolecules, that extends the previous models, was developed. As a validation, analytical solutions of this model correctly predicted elastic forces from both both systems. Additionally, this model was extended to predict rupture forces obtained by FPMD simulations of the southern bean mosaic virus.

2. Molecular dynamics simulations

Molecular dynamics (MD) simulations describe the atomic motions within a system of macromolecules of up to 10^7 atoms. In principle, such a system is described by the time-dependent Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \psi(\mathbf{R}, \mathbf{r}) = H\psi(\mathbf{R}, \mathbf{r}) \quad (2.1)$$

where $\hbar = \frac{h}{2\pi}$ is the reduced Planck constant, H the Hamiltonian of the system, ψ its wave function, and \mathbf{R} and \mathbf{r} denote the coordinates of the nuclei and the electrons, respectively. However, an analytical solution of the Schrödinger equation is already impossible for a helium atom and numerical solutions are only feasible for up to 10 atoms. To circumvent these restrictions, MD employs three approximations.

The first approximation is the Born-Oppenheimer approximation [34], which simplifies the description of the motions of nuclei and electrons. By utilizing the fact that electrons are ~ 2000 times lighter than nuclei and therefore move much faster, the electronic wave function can be assumed to instantaneously follow the motion of the nuclei. As a consequence, the wave function of the system is separated into two parts

$$\psi(\mathbf{R}, \mathbf{r}) = \psi_{\text{nuclei}}(\mathbf{R})\psi_{\text{electrons}}(\mathbf{R}, \mathbf{r}) \quad . \quad (2.2)$$

In this way, the coordinates of the nuclei are exclusively used to describe atomic motions.

The next approximation is a simplified description of the interactions between atoms by a so-called force field. In force fields, these interactions are approximated by simple analytic functions such as x^n , $1/x^n$, or $\cos(x)$ (figure 2.1). The parameters describing the single energy terms are derived from experiments or quantum mechanical calculations. Examples of commonly used force fields for protein MD simulations are OPLS [35], Amber [36], GROMOS [37], and Charmm [38].

As a last approximation, the motions of the nuclei positions are calculated by Newton's equation of motion,

$$m_i \frac{d^2 \mathbf{R}_i(t)}{dt^2} = -\nabla_i V(\mathbf{R}_i) \quad , \quad (2.3)$$

where m_i is the mass of the i^{th} atom and $V(\mathbf{R}_i)$ is the potential acting on it as given by the force field.

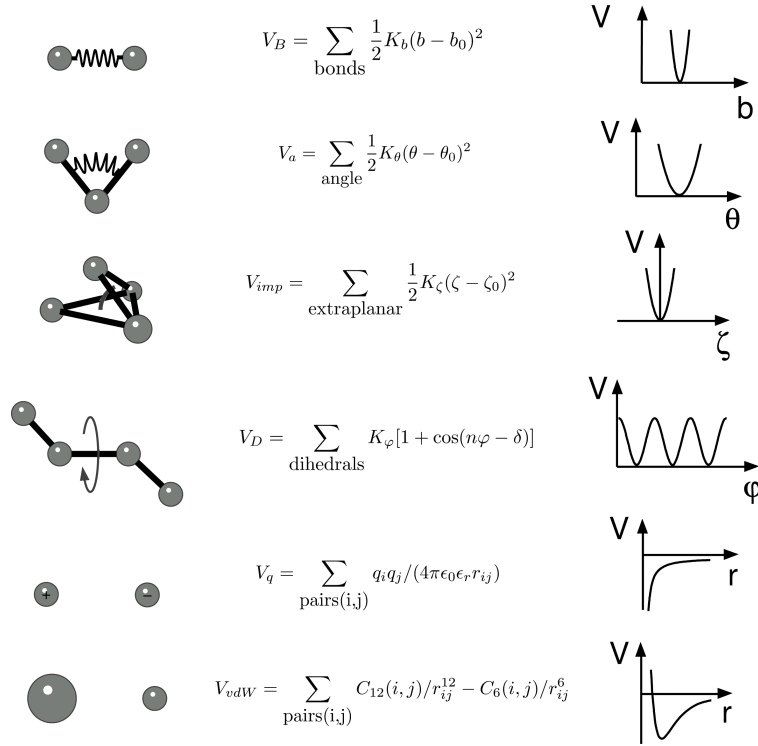


Figure 2.1.: Overview of the different energy terms of a force field. Picture provided by Gunnar Schröder.

Despite all approximations, it is still not possible to obtain an analytical time-dependent solution for atomic motions. Rather, new coordinates and velocities are calculated numerically. In this work, the leap-frog algorithm [39] is used to iteratively determine new velocities

$$\mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) = \mathbf{v}_i \left(t - \frac{\Delta t}{2} \right) + \frac{\mathbf{F}_i(t)}{m_i} \Delta t \quad (2.4)$$

and coordinates

$$\mathbf{R}_i(t + \Delta t) = \mathbf{R}_i(t) + \mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) \Delta t \quad (2.5)$$

at discrete times. Here, $\mathbf{F}_i(t) = -\nabla_i V(\mathbf{R})$ is the force acting on the i^{th} atom and Δt is the time step between each iteration.

Since simulation systems need to have a finite size, as indicated by a simulation box (figure 2.2, black lines), the treatment of system boundaries is crucial. To this end, periodic boundary conditions were employed in this work. Hence, a simulation system (green) is surrounded by virtual copies of itself that are translated according to the given lattice symmetry (gray).

As a consequence, e.g., a particle which leaves the box to the right side enters it again

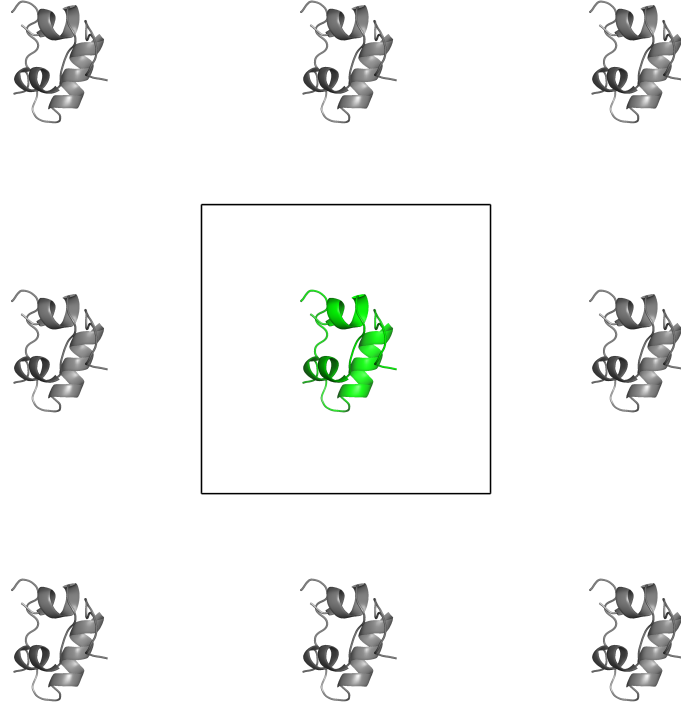


Figure 2.2.: Principle of periodic boundary conditions in two dimensions. A protein (green) in a simulation box (black) is surrounded by virtual copies of itself (gray).

from the left side. Furthermore, because no surface is present, surface artifacts do not occur. Artifacts arising from interactions between a molecule and one of its periodic images were avoided in this work by choosing a sufficiently large box so that these interactions become negligible. In addition, periodic boundary conditions are ideal for simulating systems forming a crystal lattice structure, such as bacteriorhodopsin arranged within the purple membrane.

To keep temperature and pressure close to reference values during a simulation, temperature and pressure coupling were employed. A reference temperature T_0 was maintained by employing a Berendsen thermostat [40], thus scaling the velocities of all atoms by

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right)} \quad , \quad (2.6)$$

where T is the current temperature and τ_T a coupling constant. Similarly, a Berendsen barostart [40] was used to keep pressure at a reference pressure p_0 by scaling the length vectors of the simulation box by

$$\mu = 1 - \frac{\Delta t}{3\tau_p} \kappa(p_0 - p) \quad , \quad (2.7)$$

with τ_p as the corresponding coupling constant, p as the current pressure, and κ as the compressibility of the solvent.

FPMD simulations incorporate a “virtual spring” (figure 2.3) into a MD simulation to mimic a moving AFM cantilever. This is achieved by employing a probe potential

$$V_p(z_p, t) = \frac{k_p}{2}(z_p - z_0 - vt)^2 \quad (2.8)$$

that is acting on a selected set of atoms located at z_p (rightmost atom of the molecule depicted as red spheres) and moves with a constant probe velocity v along a reaction coordinate z . Here, k_p is the spring constant of the probe potential and z_0 denotes an initial offset. In this way, a force

$$F_p(z_p, t) = -k_p(z_p - z_0 - vt) \quad (2.9)$$

is acting on the system. In the example given in figure 2.3, this enforces the release of a ligand (red) from a protein (green).

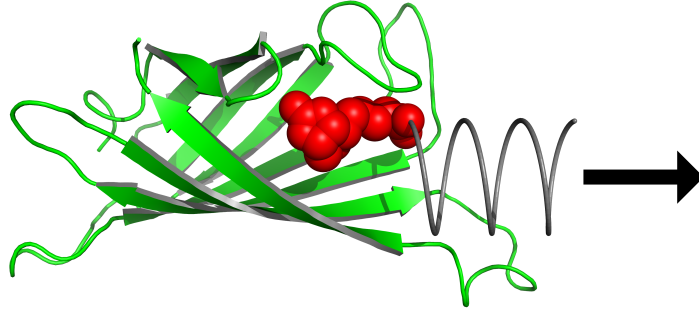


Figure 2.3.: Principle of FPMD simulations. A “virtual spring” is acting on a ligand (red), enforcing its release from a protein (green) in pulling direction (arrow).

3. Background of the publications

3.1. Velocity-Dependent Mechanical Unfolding of Bacteriorhodopsin Is Governed by a Dynamic Interaction Network

Besides soluble proteins, membrane proteins embedded in a lipid bilayer are crucial for living organisms. One special family of membrane proteins are seven α -helical transmembrane proteins. The α -helices of these proteins are oriented parallel to the membrane normal and are embedded within the lipid bilayer. It has been proposed that the single α -helices act as folding units of the protein [41].

One of the best studied α -helical transmembrane proteins is bacteriorhodopsin from halobacteria (figure 3.1A, green) [42, 43]. This protein converts light energy into a proton gradient. Light of a wavelength between 500 and 650 nm is absorbed by a chromophore called retinal (magenta), leading to its isomerization. This leads to a conformational change within the protein, which, in turn, causes protons to be pumped across the membrane. In its native environment, bacteriorhodopsin is arranged in trimers. These trimers form the unit cells of a hexagonal lattice (figure 3.1B, green). Lipid molecules (yellow) are present between the proteins. This very stable structure is called “purple membrane”.

In a pioneering work by Oesterhelt et al. [16], single bacteriorhodopsin molecules were unfolded and extracted from the purple membrane by AFM. Overlays of force curves obtained from the experiments revealed four distinct force peaks with decreasing heights. Fits of the forces to the worm-like chain polymer model [44] proposed an unfolding pathway according to which pairs of helices were extracted in a concerted manner from the membrane. Furthermore, the unusual decreasing force peak values suggested that the protein unfolded along its amino acid sequence. This is in contrast to results from unfolding experiments on soluble multi-domain proteins, where the weakest unit unfolds first, independent of its position along the amino acid sequence. Also, such an unfolding pattern is accompanied by increasing instead of decreasing force peak values. After extraction, a stable water-filled hole was left at the location of the extracted protein, underlining the stability of the purple membrane.

Later experiments systematically investigated the extraction and unfolding of bacteriorhodopsin under different extraction velocities [45]. These experiments revealed

3.1 Velocity-Dependent Unfolding of Bacteriorhodopsin

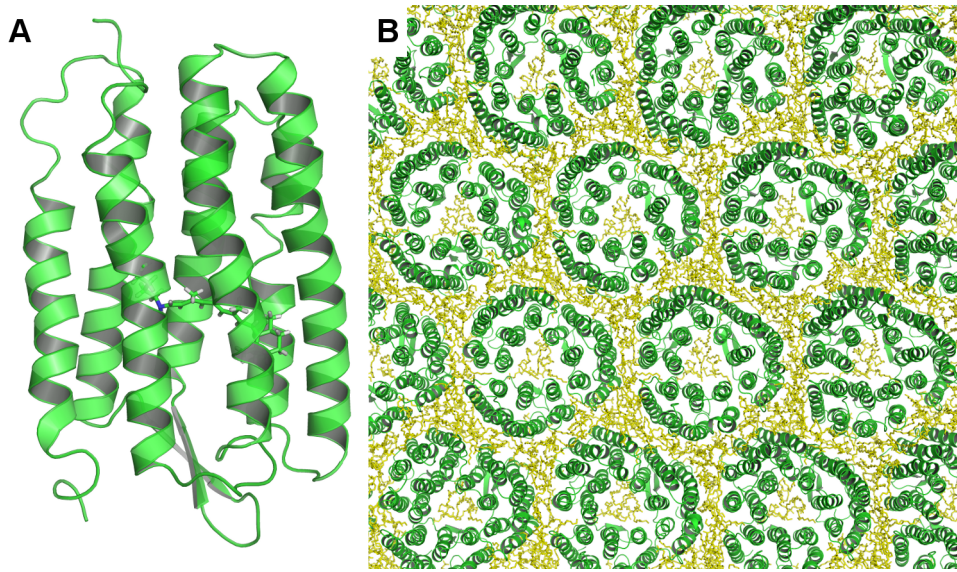


Figure 3.1.: A: Bacteriorhodopsin (cartoon representation) and the retinal (stick representation). B: Top view of the purple membrane. Proteins and lipid molecules are shown in green and yellow, respectively.

a velocity-dependent unfolding pathway, in which, instead of the unfolding of helix pairs, the unfolding of single helices became more likely at higher extraction velocities.

In another study, intermediate force peaks from single unfolding force curves were investigated in more detail [27]. Analysis of single force curves with the worm-like chain model revealed the location of a few key residues within bacteriorhodopsin that withstand larger forces and interrupt the mechanical unfolding. Thus, these residues, referred to as “anchor points”, are likely to play a key role in stabilizing the three-dimensional structure of the protein.

Although the AFM experiments provided a thorough quantification of the mechanical stability of bacteriorhodopsin, a number of open questions remain:

- How does the actual unfolding proceed at an atomistic level?
- What is the relation between the observed forces and the unfolding pathway?
- How are anchor points stabilized within the protein?
- Why are unfolding pathways velocity-dependent?

These questions are addressed in the study “Velocity-Dependent Mechanical Unfolding of Bacteriorhodopsin Is Governed by a Dynamic Interaction Network” [1].

3.2. An Unusual Hydrophobic Core Confers Extreme Flexibility to HEAT Repeat Proteins

The cell nucleus is an important part of all eukaryotic cells. It contains the genetic information in the form of DNA and is also the compartment in which gene expression is initiated. The nuclear envelope, a double membrane, encloses the nucleus and separates it from the cytoplasm. The function of the cell relies on the selective transport of molecules between the nucleus and the cytoplasm. This transport is mediated by nuclear pore complexes, large protein assemblies that span the nuclear envelope and have a diameter of ~ 100 nm [46]. Nuclear pore complexes are filled with a mesh-work of intrinsically disordered protein domains known as FG-rich nucleoporins [47], which guarantee a remarkable filter function. Ions and small molecules with a mass below 30 kDa pass through nuclear pores by diffusion whereas larger molecules are blocked. However, molecules with a mass above 30 kDa can pass through the pore once they are in complex with a transport factor [48, 49].

These transport factors are called karyopherins. Karyopherin comprise of importins [50] and exportins [51, 52], which mediate nuclear import and export, respectively. All karyopherins are built of HEAT repeats, structural motifs consisting of two anti-parallel helices connected by a short loop (figure 3.2A) [53]. These HEAT repeats are stacked to form superhelical structures (figure 3.2B) [54].

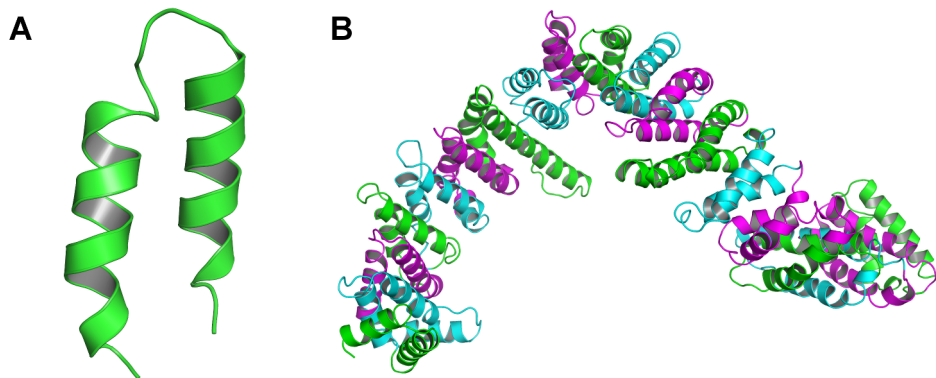


Figure 3.2.: HEAT repeats and global structure of karyopherins. A: Example of one HEAT repeat. B: The karyopherin kap95p (yeast importin- β). Different colors represent different HEAT repeats.

Importin- β (figure 3.2B) is the prototypical karyopherin. This protein (figure 3.3, green) transports cargo molecules (red) from the cytoplasm into the nucleus (left side). Inside the nucleus, cargo molecules are released once RanGTP (magenta) binds to importin- β . The importin- β -RanGTP complex is then transported back into the cytoplasm (right side), where it is disassembled after GTP is hydrolyzed to GDP. This transport cycle is mediated by a RanGTP/RanGDP gradient between

3.2 Extreme Reversible Elasticity of Importin- β

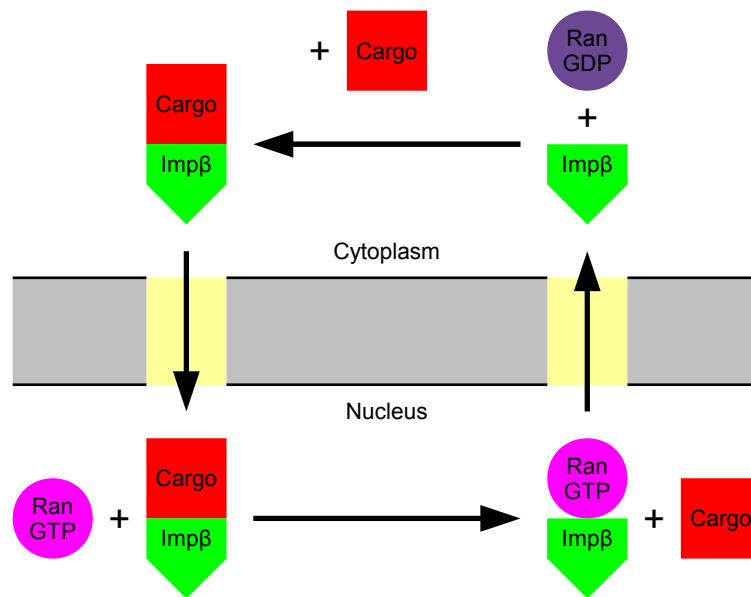


Figure 3.3.: Schematic importin- β transport cycle. Importin- β transports cargoes from the cytoplasm into the nucleus across the nuclear envelope (gray) through nuclear pore complexes (yellow).

the nucleus and the cytoplasm. RanGDP travels back into the nucleus with the help of the protein nuclear transport factor 2 (NTF2) [55, 56].

Unlike most receptor-ligand complexes, importin- β and its homologues are able to bind to a variety of cargo molecules that differ in structure and sequence [9]. In most cases, cargoes are not bound directly but via adapter proteins such as importin- α or snurportin that possess binding domains for both, cargo molecules and importin- β .

Crystal structures of importin- β in complex with different cargoes and adapters, respectively, revealed that importin- β adapts its shape to its different binding partners by changing its global conformation (figure 3.4) [9, 57–60]. In all known structures, importin- β (green) wraps around its respective binding partner (red and magenta). Because the protein adapts to its different binding partners, it has to possess an inherent flexibility. This flexibility could also explain the paradox of importin- β binding tightly to its binding partners while only needing small amounts of energy for dissociation. It has been suggested, that the binding energy of importin- β is partly stored by mechanical deformation of the protein [9].

One first explanation how importin- β changes its conformation was given by Zachariae and Grubmüller [28]. In this MD study, RanGTP was removed from a RanGTP-importin- β complex and, thus, the behavior of free importin- β was simulated. In these simulations, importin- β opened up by extending its end-to-end distance from 9 nm to 13 nm within a time of only ~ 30 ns. This work characterized importin- β as a spring-loaded molecule.

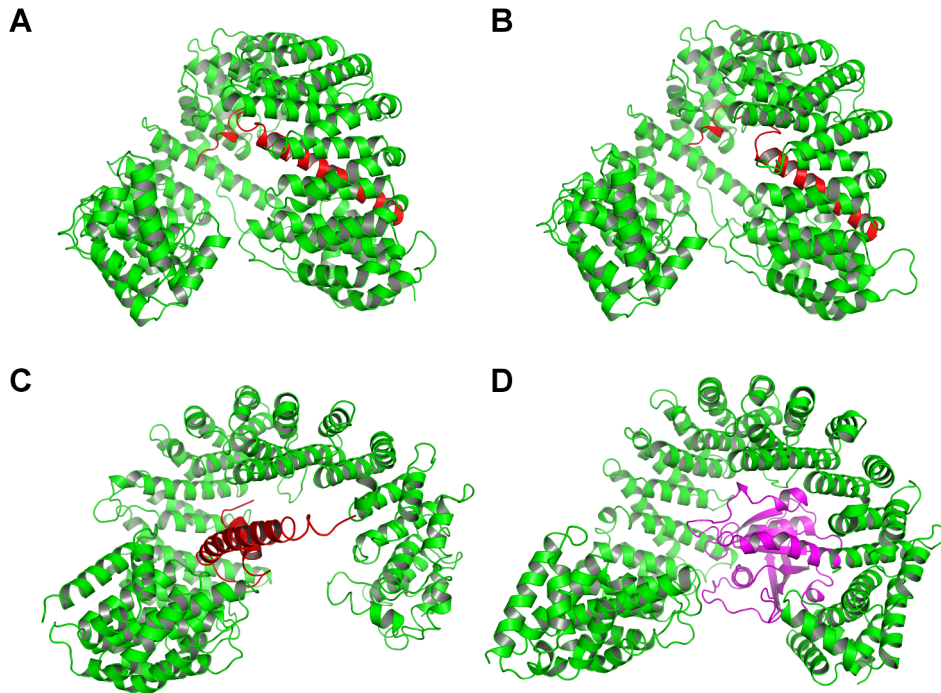


Figure 3.4.: Importin- β (green) with various adapter proteins (red) and RanGTP (magenta). Different complexes involving the following molecules are shown. Abbreviations in parenthesis denote the PDB identifier of the complex. A: Importin- α (1QGK [57]). B: Snurportin 1 (2P8Q [58]). C: SREBP-2 (1UKL [59]). D: RanGTP (2BKU [60]).

Despite the different x-ray structures and its classification as a molecular spring by MD simulations, a number of open questions about the flexibility of importin- β remained:

- To which extent is importin- β flexible?
- Does importin- β have a discrete set of conformations or can the protein change continuously between different shapes?
- Are these changes permanent (plastic flexibility) or are they reversible (elastic flexibility)?
- Which amount of energy is needed to drive importin- β into a different conformation?
- What are the atomistic determinants that cause importin- β to be flexible?

These questions are addressed in the study “An Unusual Hydrophobic Core Confers Extreme Flexibility to HEAT Repeat Proteins” [2].

3.2 Extreme Reversible Elasticity of Importin- β

3.3. Non-equilibrium elastic behavior of macromolecules

Introduction

Molecular bonds in biological systems are typically modelled by a one-dimensional free energy landscape $G_0(x)$ along a reaction coordinate x . Reaction coordinates are, e.g., the distance between molecules or the length of a macromolecule. The bound state is modelled as the minimum of $G_0(x)$ that is separated from the unbound state by a barrier at a barrier location x_b (figure 3.5, black line). Unbinding events are described as a crossing of the barrier. In equilibrium, thermally activated crossing of the barrier is quantified by a rate constant r [61, 62].

External forces F , that are applied by, e.g., an AFM cantilever, lower the height of the barrier by Fx_b . In this way, thermally activated crossing is more likely to occur (figure 3.5, red line).

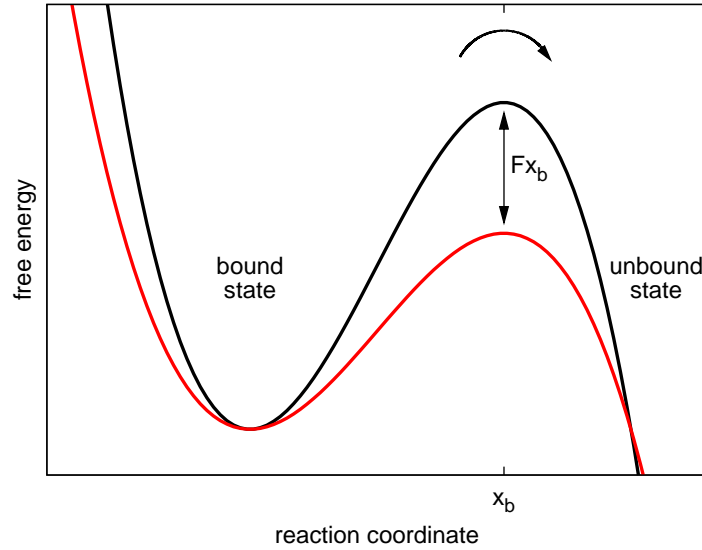


Figure 3.5.: Molecular bonds are modeled by a free energy landscape. Unbinding events are described as crossing of a barrier located at x_b . External forces F lower the height of the barrier by Fx_b and accelerate barrier crossing and thus unbinding in this way.

Rupture forces occurring upon unbinding events are typically found to depend on the loading rate dF/dt or on the probe velocity v . This behavior is explained by Bell's model [30], which predicts a force-dependent rupture rate

$$r(F) = \omega_0 e^{\beta F x_b} \quad (3.1)$$

3.3 Non-equilibrium elastic behavior of macromolecules

for barrier crossing, where ω_0 is an attempt frequency, and $\beta = 1/(k_B T)$ is the inverse temperature of the system ($k_B = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant). Bell's model is based on the assumption that an underlying one-dimensional free energy landscape is sufficient to describe macromolecular systems. Furthermore, this model predicts that other elastic properties of macromolecules besides rupture forces are, to first order, independent of the probe velocity.

However, recent FPMD simulations of importin- β [2] and of the outer shell of the southern bean mosaic virus [29] revealed a velocity dependence of spring constants upon elastic deformation at high probe velocities. Yet, the velocity dependence of rupture forces remained as predicted by Bell's model. This phenomenon is neither explained by Bell's model nor by any later models that extend Bell's model [23, 31–33].

To that aim, I present a general model for non-equilibrium elasticity of macromolecules that is based on a multi-dimensional free energy landscape. The following questions were addressed:

- Does this general model describe the observed phenomena for both examples?
- Does this model also correctly predict rupture forces?
- Is an analytical solution of the forces predicted by this model possible?

The study “Non-equilibrium elastic behavior of macromolecules” presented in this chapter describes this model and answers the above questions.

Model for non-equilibrium elasticity

Previous models for the elastic behavior of macromolecules employ a one-dimensional free energy landscape to describe the length of the molecule. Accordingly, the length is used as a reaction coordinate (figure 3.6, red arrows).

This one-dimensional description implies that molecular motions along the length of a macromolecule solely account for elastic behavior. However, in large macromolecules also motions along other degrees of freedom play a role (figure 3.6, green arrows). These motions will be referred to as “perpendicular modes”.

Here, I present a new model for the elastic behavior of macromolecules that not only considers motions along their length but also their perpendicular modes. In detail, two assumptions about perpendicular modes are made:

1. Perpendicular modes of macromolecules are assumed to be coupled to motions along the length of a macromolecule. Consequently, length changes, e.g. due to deformation in AFM experiments or FPMD simulations, also enforce changes in perpendicular modes.
2. The dynamics of perpendicular modes are assumed not to be faster than the dynamics along the length of a macromolecule. Thus, perpendicular modes are

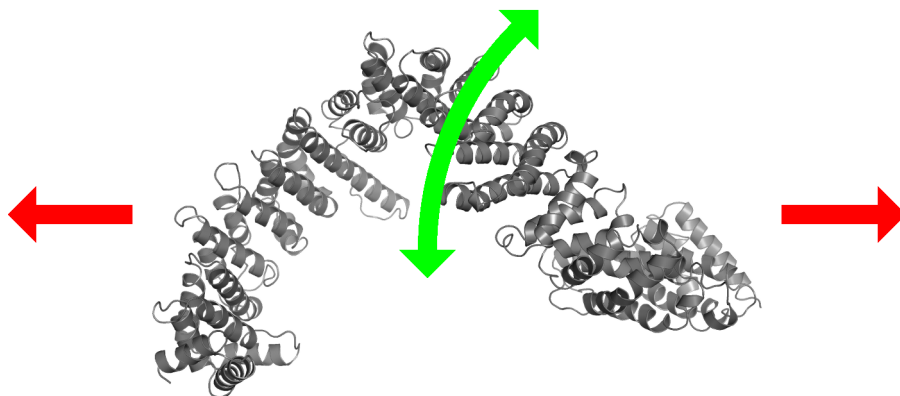


Figure 3.6.: Schematic representation of molecular motions along the reaction coordinate (red arrows) as well as perpendicular modes (green arrows).

assumed to not instantaneously follow length changes, but after a relaxation phase.

The extent of relaxation of perpendicular modes upon length increase depends on the applied time scales. Upon an enforced elongation that is slow in comparison to the dynamics of the perpendicular modes, relaxation of perpendicular modes to their respective minima takes place. This behavior is indistinguishable from the behavior proposed by previous models.

However, upon a fast enforced elongation, this relaxation does not take place. As a consequence, an additional mechanical resistance of the macromolecule against enforced elongation is present. This results in a stiffening of the molecule and, conversely, in the observation of higher spring constants upon elastic deformation.

Modelling thermal fluctuations of macromolecules

For the mathematical description of the model for non-equilibrium elasticity, a two-dimensional free energy landscape was used. There, the state of a macromolecule $\mathbf{x} = (x, y)^T$ describes a reaction coordinate x and a perpendicular coordinate y . The length of the macromolecule is taken as the reaction coordinate x . The perpendicular coordinate y represents all perpendicular modes of the macromolecule. This simplification is based on the assumption that the perpendicular modes behave collectively.

To completely describe the dynamics of the state $\mathbf{x} = (x, y)^T$ of the macromolecule, thermal fluctuations that result in random changes of \mathbf{x} have to be taken into account.

3.3 Non-equilibrium elastic behavior of macromolecules

A mathematical formalism for molecular motions within a free energy landscape $G(\mathbf{x}, t)$ that includes thermal fluctuations is given by the Langevin equation [63],

$$\frac{d}{dt}\mathbf{v} = -\gamma\mathbf{v} + \mu(t) - \frac{1}{m}\nabla G(\mathbf{x}, t) \quad . \quad (3.2)$$

Here, $\mathbf{v} = \frac{d}{dt}\mathbf{x}$ denotes changes of the state \mathbf{x} and γ is a friction coefficient. Thermal fluctuations are accounted for by a noise term $\mu(t)$. Solutions of the Langevin equation yield probability distributions $p(\mathbf{x}, t)$ of finding a macromolecule at a state \mathbf{x} for a given time t .

An equivalent formalism to the Langevin equation is given by the Smoluchowski equation [63, 64],

$$\frac{\partial}{\partial t}p(\mathbf{x}, t) = \nabla D \{ \nabla + \beta[\nabla G(\mathbf{x}, t)] \} p(\mathbf{x}, t) \quad . \quad (3.3)$$

Here, D denotes a diffusion coefficient and, as before, $G(\mathbf{x}, t)$ is a free energy landscape. Because solutions of the Smoluchowski equation are easier to obtain, in the following, the Smoluchowski equation will be used.

In the following, the goal is to find an analytical expression for the solution of the Smoluchowski equation for the model for non-equilibrium elasticity. Also, probe processes as exerted by AFM experiments and FPMD simulations will be included in the model. The two-dimensional expression for this probability distribution $p(x, y, t)$ is the generalization of the probability distribution $p(x, t)$ for elastic behavior as described by a one-dimensional free energy landscape.

Hence, as a first step, solutions to the Smoluchowski equation for the elastic behavior described by a one-dimensional free energy landscape are derived. As a next step, the enforced elongation of elastic macromolecules within this model will be discussed. Then, this solution is generalized to two dimensions to describe the elastic response according to the free energy landscape of the model for non-equilibrium elasticity.

One-dimensional models for elasticity (Bell's model)

The free energy landscape of an elastic macromolecule in one dimension is modeled by

$$G_0(x) = \frac{k_0}{2}x^2 \quad (3.4)$$

(figure 3.7A, black lines). The spring constant k_0 characterizes the stiffness of the macromolecule. The length of the macromolecule is taken as a reaction coordinate x , that is chosen such that the equilibrium length is $x = 0$.

Solutions $p(x, t)$ of the Smoluchowski equation for this free energy landscape are now used to provide the description of the relaxation of an elongated elastic macromolecule to its equilibrium length. We assume that, in the beginning, the length of the macromolecule has a value of $x_0 \neq 0$ and the probability distribution is located sharply at $x = x_0$. The probability distribution that solves the Smoluchowski

equation is a Gaussian normal distribution [65]

$$p_0(x, t) = \frac{1}{\sqrt{2\pi}\sigma_0(t)} \exp\left(-\frac{(x - \langle x_0(t) \rangle)^2}{2\sigma_0(t)^2}\right) \quad (3.5)$$

(blue lines) whose time-dependent center (figure 3.7B) and width (figure 3.7C) are given by

$$\langle x_0(t) \rangle = x_0 e^{-\beta k_0 D t} \quad (3.6)$$

and

$$\sigma_0(t) = \sqrt{\frac{1 - e^{-2\beta k_0 D t}}{\beta k_0}} \quad (3.7)$$

respectively.

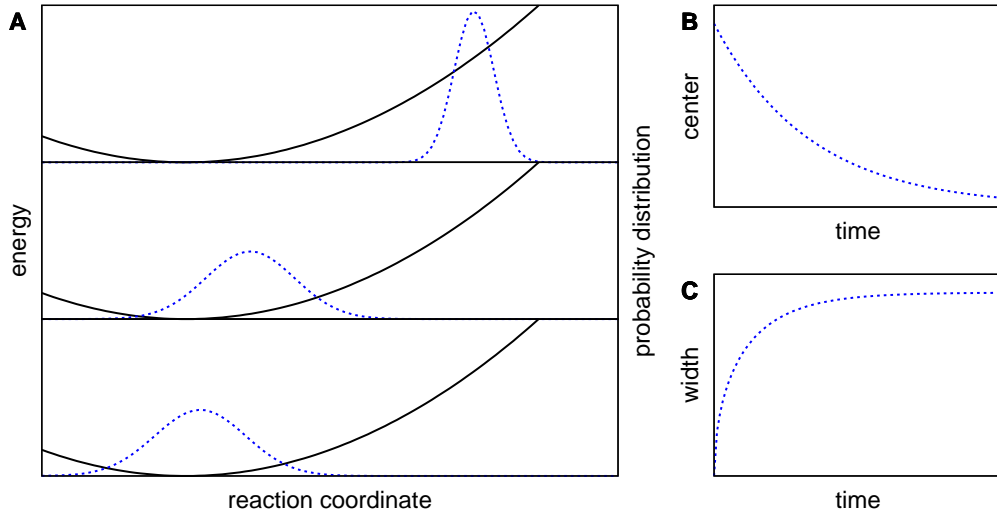


Figure 3.7.: Probability distributions according to solutions to the Smoluchowski equation for a static free energy landscape $G_0(x)$. A: Development of a probability distribution (blue) and the free energy landscape (black). Time advances from top to bottom. B: Center of the probability distribution over time. C: Width of the probability distribution over time.

As a next step, the enforced elongation of a macromolecule will be described. Hence, a solution of the Smoluchowski for a time-dependent free energy landscape

$$G_{\text{moving}}(x, t) = G_0(x - vt) = \frac{k_0}{2}(x - vt)^2 \quad (3.8)$$

that moves with a constant velocity v is needed. This time-dependent free energy landscape serves as a model for a moving AFM cantilever.

To this aim, the probability distribution $p_0(x, t)$ belonging to a static free energy

3.3 Non-equilibrium elastic behavior of macromolecules

landscape $G_0(x)$ was taken and the ansatz

$$p_{\text{moving}}(x, t) = p_0(x + x_{\text{off}}(t) - vt, t) \quad (3.9)$$

for the probability distribution belonging to a time-dependent free energy landscape $G_{\text{moving}}(x, t)$ was made. This ansatz introduces a time-dependent offset $x_{\text{off}}(t)$. The solution of the Smoluchowski equation yields a value of

$$x_{\text{off}}(t) = \frac{v}{\beta k D} (1 - e^{-\beta k D t}) \quad (3.10)$$

As a consequence, the center of the probability distribution is given by

$$\langle x_{\text{moving}}(t) \rangle = vt - \frac{v}{\beta k D} (1 - e^{-\beta k D t}) \quad (3.11)$$

The width of the probability distribution $\sigma_{\text{moving}}(t)$ remains identical to $\sigma_0(t)$.

As a next step, the stretching of an elastic macromolecule described by $G_0(x)$ (figure 3.8A, black lines) by a probe potential

$$V_p(x, t) = \frac{k_p}{2} (x - vt)^2 \quad (3.12)$$

(red lines) will be modeled. This process is modelled by a time-dependent free energy landscape $G(x, t)$, which is the sum of $G_0(x)$ and $V_p(x, t)$,

$$G(x, t) = \frac{k_0}{2} x^2 + \frac{k_p}{2} (x - vt)^2 \quad (3.13)$$

To determine a solution to the Smoluchowski equation for this free energy landscape, $G(x, t)$ is rewritten as

$$G(x, t) = \frac{k_{\text{eff}}}{2} (x - v_{\text{eff}} t)^2 + \frac{k_0 k_p}{k_0 + k_p} v^2 t^2 \quad (3.14)$$

with an effective spring constant

$$k_{\text{eff}} = k_0 + k_p \quad (3.15)$$

and an effective velocity

$$v_{\text{eff}} = \frac{k_p}{k_0 + k_p} v \quad (3.16)$$

In this form, the derivation of a solution to the Smoluchowski equation for $G(x, t)$ is analogue to the case of $G_{\text{moving}}(x, t)$.

Hence, the center of the probability distribution is given by

$$\langle x_{\text{eff}}(t) \rangle = v_{\text{eff}} \left(t - \frac{1 - e^{-\beta k_{\text{eff}} D t}}{\beta k_{\text{eff}} D} \right) \quad (3.17)$$

(figure 3.8B, blue line). With this result, a mean force $\langle F_{\text{eff}}(t) \rangle = k_p(vt - \langle x_{\text{eff}}(t) \rangle)$

is determined from the displacement between the center $\langle x_{\text{eff}}(t) \rangle$ of the probability distribution and of the center vt of the probe potential (red line), yielding

$$\langle F_{\text{eff}}(t) \rangle = k_p \left(vt - \frac{k_p}{k_0 + k_p} vt + \frac{k_p v}{\beta(k_0 + k_p)^2 D} (1 - e^{-\beta(k_0 + k_p)Dt}) \right) \quad (3.18)$$

(figure 3.8C).

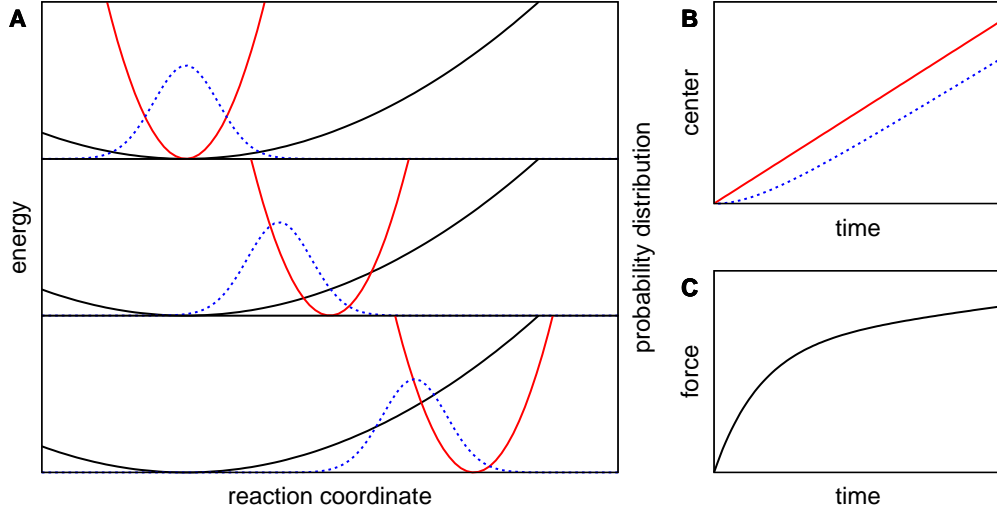


Figure 3.8.: Probability distribution for the enforced elongation of an elastic macromolecule by a probe potential. A: Development of the probability distribution (blue) and of the probe potential (red) and the free energy landscape of the elastic macromolecule (black). Time advances from top to bottom. B: Center of the probability distribution (blue) and of the probe potential (red) over time. C: Forces exerted by the probe potential on the elastic macromolecule.

Now, the mathematical description of the elastic behavior of macromolecules according to a one-dimensional free energy landscape, as proposed by Bell's model, is present. Fits of theoretical forces $F(t)$ given by equation (3.18) to forces from FPMD simulations or AFM experiments are used to determine spring constants upon elastic deformation. With this approach, Bell's model is validated with respect to elastic behavior.

Modeling non-equilibrium elasticity

In the following, a free energy landscape $G_0(x, y)$ that describes the model for non-equilibrium elasticity will be constructed. A spring constant k_{\parallel} along the reaction coordinate x and a spring constant k_{\perp} along the perpendicular coordinate y serve

3.3 Non-equilibrium elastic behavior of macromolecules

as parameters. Both are coupled via a coupling constant α , which is the third parameter. By construction, this free energy landscape is required to revert to Bell's model for the case of complete relaxation of perpendicular modes.

The two-dimensional harmonic free energy landscape

$$G_0(x, y) = \frac{k_{\parallel}}{2}x^2 - k_{\perp}x\alpha y + \frac{k_{\perp}}{2}\alpha^2 y^2 \quad , \quad (3.19)$$

whose principal components are rotated with respect to the Cartesian coordinates, fulfills these requirements (figure 3.9). In the case of a fully relaxed perpendicular coordinate, $dG_0(x, y)/dy \stackrel{!}{=} 0$, this model reverts to a one-dimensional free energy landscape

$$G_{0,\text{Bell}}(x) = \frac{k_{\parallel} - k_{\perp}}{2}x^2 \quad (3.20)$$

in accordance to Bell's model with an equilibrium spring constant $k_{eq} = k_{\parallel} - k_{\perp}$.

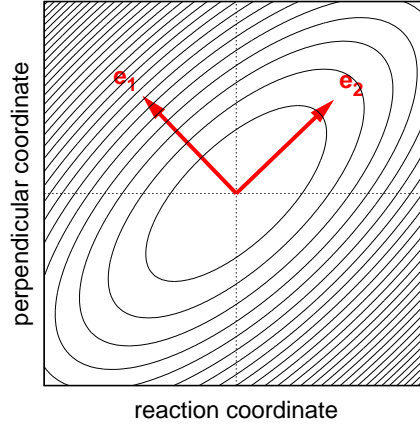


Figure 3.9.: Two-dimensional harmonic free energy landscape that is rotated with respect to the reaction coordinate (black contour lines) and its eigenvectors \mathbf{e}_1 and \mathbf{e}_2 (red arrows).

To describe the enforced elongation along the reaction coordinate, as done in AFM experiments or FPMD simulations, a probe potential

$$V_p(x, t) = \frac{k_p}{2}(x - vt)^2 \quad (3.21)$$

that acts exclusively along the reaction coordinate is added to $G_0(x, y)$. This results in a time-dependent free energy landscape

$$G(x, y, t) = \frac{k_{\parallel}}{2}x^2 - k_{\perp}x\alpha y + \frac{k_{\perp}}{2}\alpha^2 y^2 + \frac{k_p}{2}(x - vt)^2 \quad , \quad (3.22)$$

which can be written in a more convenient vector form

$$G(\mathbf{x}, t) = \frac{1}{2}(\mathbf{x} - \mathbf{v}_{\text{eff}}t)^T \mathbf{C}(\mathbf{x} - \mathbf{v}_{\text{eff}}t) + G_{\text{off}}(t) \quad . \quad (3.23)$$

Here,

$$\mathbf{C} = \begin{pmatrix} k_{\parallel} + k_p & -\alpha k_{\perp} \\ -\alpha k_{\perp} & \alpha^2 k_{\perp} \end{pmatrix} \quad (3.24)$$

is a symmetric matrix and

$$\mathbf{v}_{\text{eff}} = \frac{k_p v}{k_{\parallel} - k_{\perp} + k_p} \begin{pmatrix} 1 \\ 1/\alpha \end{pmatrix} \quad (3.25)$$

denotes the effective velocity of the minimum of $G_0(\mathbf{x}, t)$. $G_{\text{off}}(t)$ is a time-dependent offset.

For the solution of the Smoluchowski equation to $G(\mathbf{x}, t)$, the ansatz given by equation (3.9) is generalized to two dimensions. To this end, it is employed that the two-dimensional probability distribution $p(x, y, t)$ that solves the Smoluchowski equation for this energy landscape is the product of two uncoupled probability distributions along the eigenvectors \mathbf{e}_1 and \mathbf{e}_2 of \mathbf{C} (figure 3.10 and red arrows in figure 3.9). The eigenvalues λ_1 and λ_2 of \mathbf{C} are spring constants along the eigenvectors. This probability distribution is given by

$$p(\mathbf{x}, t) = \prod_{i=1,2} p_i(\mathbf{x}, t) = \prod_{i=1,2} \frac{1}{\sqrt{2\pi}\sigma_i(t)} \exp \left\{ -\frac{[(\mathbf{x} - \langle \mathbf{x}(t) \rangle) \cdot \mathbf{e}_i]^2}{2\sigma_i^2(t)} \right\} \quad (3.26)$$

with

$$\sigma_i(t) = \sqrt{\frac{1 - e^{-2\beta\lambda_i D t}}{\beta\lambda_i}} \quad (3.27)$$

as the width of each process along the eigenvectors (figure 3.10, center and right panel). The center of this two-dimensional probability distribution moves along

$$\langle \mathbf{x}(t) \rangle = \sum_{i=1,2} (\mathbf{v}_{\text{eff}} \cdot \mathbf{e}_i) \cdot \left(t - \frac{1 - e^{-\beta\lambda_i D t}}{\beta\lambda_i D} \right) \cdot \mathbf{e}_i, \quad (3.28)$$

which is a generalization of $\langle x_{\text{eff}}(t) \rangle$ given by equation (3.17) to two dimensions. The mean force acting along the reaction coordinate is given by

$$\langle F(t) \rangle = k_p \left(vt - \langle \mathbf{x}(t) \rangle \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right) \quad (3.29)$$

3.3 Non-equilibrium elastic behavior of macromolecules

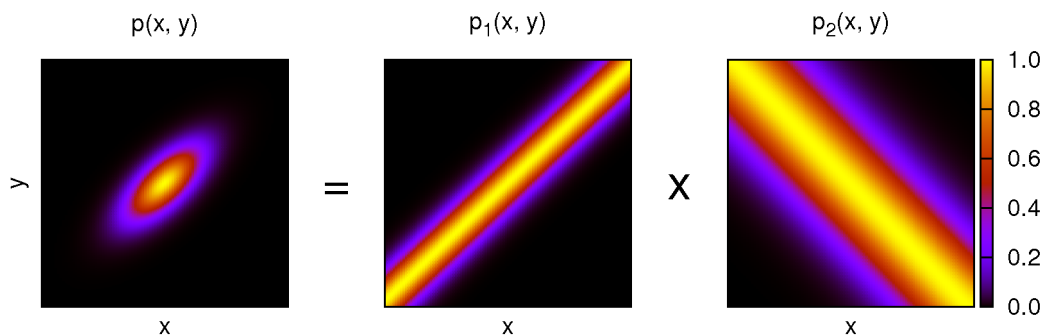


Figure 3.10.: A two-dimensional probability distribution $p(x, y)$ (left panel) is the product of two uncoupled probability distributions $p_1(x, y)$ (center panel) and $p_2(x, y)$ (right panel).

Testing the model

In the following, the model for non-equilibrium elasticity will be tested for the fulfillment of the initial requirements:

- Trajectories within the free energy landscape of the model are velocity-dependent.
- As a consequence, elastic properties become also velocity-dependent.

Example trajectories of the state \mathbf{x} of a macromolecule according to the probability distribution for the enforced elongation by a probe potential given by equation (3.26) were calculated for different probe velocities (figure 3.11A, symbols). These results confirm that trajectories within the free energy landscape given by equation (3.19) are, indeed, velocity-dependent. As predicted, slow probe velocities lead to trajectories that move along lower energies (blue symbols) whereas fast probe velocities lead to trajectories along higher energies (red symbols). For extremely fast and slow probe velocities, respectively, limiting cases for the mean of the trajectories are predicted (dark red and dark blue lines, respectively).

Projections of the energies onto the reaction coordinate closely resemble a harmonic potential $V_{\text{eff}} = (k_{\text{eff}}/2)x^2$ for all probe velocities (figure 3.11B, symbols). As predicted, the curvature of the energies along the reaction coordinate, that defines the spring constant k_{eff} , varies with the applied probe velocity. As a consequence, spring constants measured by a one-dimensional approach, such as from force-fits according to equation (3.18), are velocity-dependent.

The forces acting along the reaction coordinate exhibit similarities to forces predicted by elasticity models based on one-dimensional free energy landscapes, despite the difference in the underlying free energy landscapes (figure 3.11C, symbols).

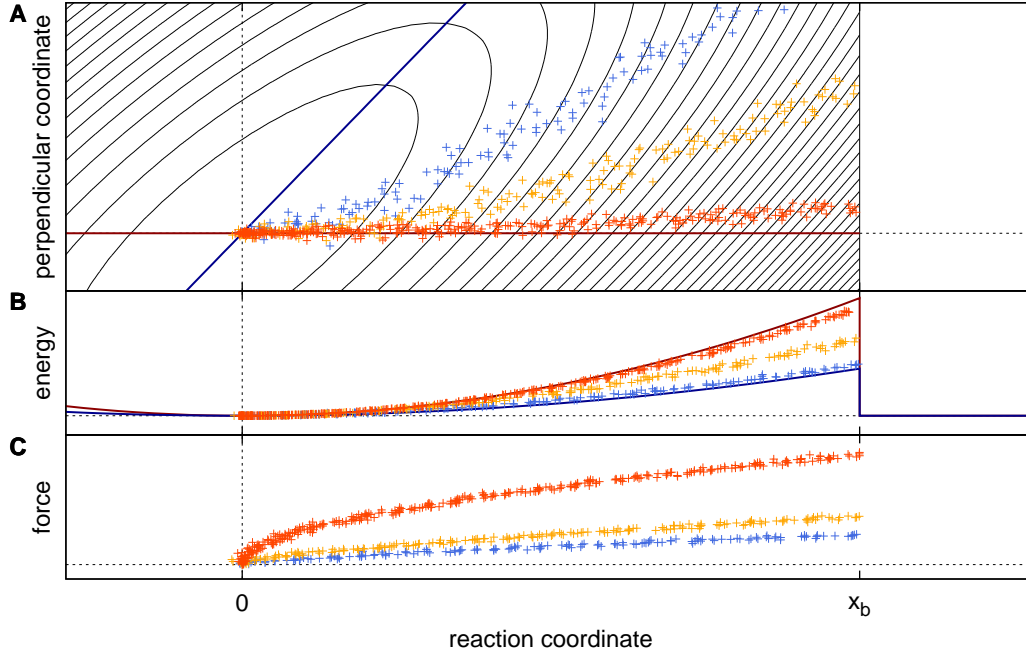


Figure 3.11.: The model for non-equilibrium elasticity and the elastic behavior of macromolecules under different probe velocities. Symbols represent trajectories under the influence of probe potentials with different probe velocities (blue: slow, orange: medium, red: fast). Dark blue and dark red lines depict limiting cases for slow and fast probe velocities, respectively. A: The two-dimensional free energy landscape of the model (iso-surface lines). Symbols depict trajectories within the free energy landscape. A barrier is located at x_b . B: Energy along the reaction coordinate. C: Forces exerted by the probe potential.

Rupture forces

As stated earlier, the rupture of a molecular bond is modelled as the crossing of a barrier within a free energy landscape (figure 3.5). To calculate the rupture forces that occur upon unbinding in dependence of a rate constant r , the fraction of bound states $Q(t)$, i.e. the probability of finding the system in the bound state, is employed. The fraction of bound states is related to the rate constant by the differential equation

$$dQ(t)/dt = -r(t)Q(t) \quad . \quad (3.30)$$

Typically, the fraction of bound states has a sigmoid shape over time (figure 3.12A). The time derivative of the fraction of bound states, $dQ(t)/dt$, denotes the flux over the barrier (figure 3.12B).

3.3 Non-equilibrium elastic behavior of macromolecules

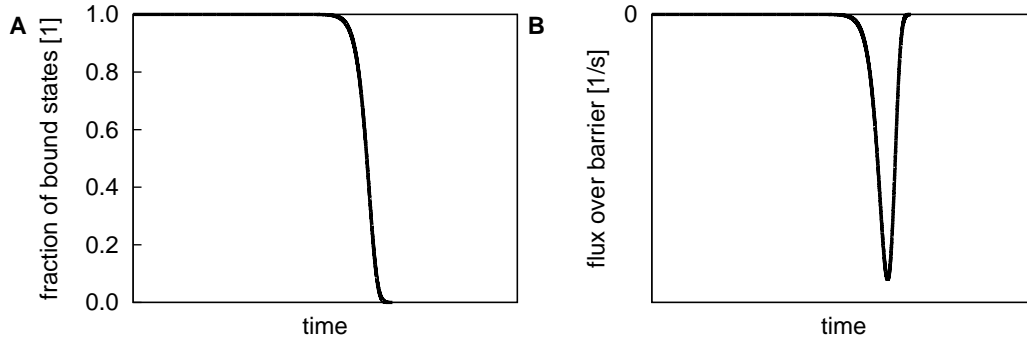


Figure 3.12.: A: Fraction of bound states $Q(t)$. B: Flux over barrier $dQ(t)/dt$.

A general formula for the rate constant for barrier crossing is given by [66]

$$r(t) = \frac{D}{\int_{\text{well}} e^{-\beta G} \int_{\text{barrier}} e^{\beta G}} \quad , \quad (3.31)$$

where D denotes a diffusion constant and G is an underlying free energy landscape. The rupture event is accompanied by a mean rupture force $\langle F_{\text{yield}} \rangle$ that is given by the integral of the product of force and the flux over a barrier over time,

$$\langle F_{\text{yield}} \rangle = - \int_0^\infty \langle F(t) \rangle \frac{dQ(t)}{dt} dt \quad . \quad (3.32)$$

However, the rate constant given by equation (3.31) can only be applied to systems that are near equilibrium and that are described by a one-dimensional free energy landscape. To calculate a rate constant for a system that is out of equilibrium and that is based on a multi-dimensional free energy landscape, such as the problem that is discussed here, this formalism has to be generalized.

As a first step, the equilibrium free energy $G(x, y, t)$ given by equation (3.22) is replaced by a non-equilibrium free energy $G_{NE}(x, y, t)$ that is defined as

$$G_{NE}(x, y, t) = -k_B T \ln(p(x, y, t)) \quad , \quad (3.33)$$

where $p(x, y, t)$ given by equation (3.26) is the probability distribution belonging to the enforced elongation of a macromolecule described by the free energy landscape $G(x, y, t)$.

To calculate the rate constant $r(t)$ for barrier crossing in a two-dimensional free energy landscape, it is not sufficient to account for one single path through the free energy landscape since multiple paths for barrier crossing are possible. Rather, this rate constant is the sum of all possible rate constants for single paths across the free energy landscape.

Therefore, two steps were taken to estimate the rate constant for barrier crossing in a two-dimensional free energy landscape:

1. A rate constant

$$r_x(y, t) = \frac{D}{\int_{\text{well}} e^{-\beta G_{NE}(x, y, t)} dx dy \int_{\text{barrier}} e^{\beta G_{NE}(x, y, t)} dx} \quad (3.34)$$

according to a single path along the reaction coordinate across the free energy landscape over the barrier is calculated as given by equation (3.31). This rate constant is evaluated for the path parallel to the reaction coordinate with the minimal barrier height (figure 3.13, red line), which is the most likely path in equilibrium.

2. This one-dimensional rate constant is integrated along the perpendicular coordinate to account for all possible paths through the two-dimensional free energy landscape (green lines), yielding

$$r(t) = \int_{-\infty}^{+\infty} r_x(y, t) dy \quad (3.35)$$

as the rate constant for barrier crossing in a two-dimensional free energy landscape.

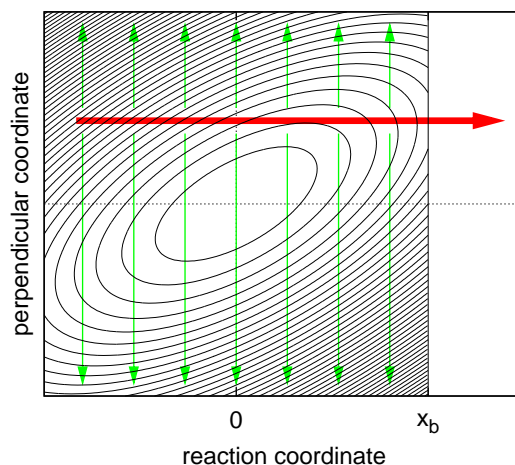


Figure 3.13.: Calculation of a rate constant for barrier crossing in a two-dimensional free energy landscape. First, a rate constant $r_x(y, t)$ accounting for one way along the reaction coordinate is calculated (red arrow). This rate constant is integrated over the perpendicular coordinate to account for all possible paths (green arrows).

As a first step, the rate constant $r_x(y, t)$ will be estimated. To evaluate the integrals in $r_x(y, t)$, a component $p_x(x, t)$ of $p(x, y, t)$ given by equation (3.26) that

3.3 Non-equilibrium elastic behavior of macromolecules

only depends on the reaction coordinate x must be determined. This is achieved by describing $p(x, y, t)$ as the product of two normalized probability distributions according to

$$p(x, y, t) = p_x(x, t) \cdot p_y(x, y, t) \quad . \quad (3.36)$$

These two components are given by

$$p_x(x, t) = \frac{1}{\sqrt{2\pi}\sigma_x(t)} \exp \left\{ -\frac{(x - \langle x(t) \rangle)^2}{2\sigma_x^2(t)} \right\} \quad (3.37)$$

and

$$p_y(x, y, t) = \frac{1}{\sqrt{2\pi}\sigma_y(t)} \exp \left\{ -\frac{(y - \langle y(t) \rangle + c \cdot (x - \langle x(t) \rangle))^2}{2\sigma_y^2(t)} \right\} \quad (3.38)$$

(figure 3.14). The additional parameters employed here are given by

$$\sigma_x(t) = \sqrt{(e_{1,x}\sigma_1(t))^2 + (e_{2,x}\sigma_2(t))^2} \quad , \quad (3.39)$$

$$\sigma_y(t) = \sqrt{1 / \left[\left(\frac{e_{1,y}}{\sigma_1(t)} \right)^2 + \left(\frac{e_{2,y}}{\sigma_2(t)} \right)^2 \right]} \quad , \quad (3.40)$$

and

$$c = \left[\frac{e_{1,x}e_{1,y}}{\sigma_1^2(t)} + \frac{e_{2,x}e_{2,y}}{\sigma_2^2(t)} \right] / \left[\left(\frac{e_{1,y}}{\sigma_1(t)} \right)^2 + \left(\frac{e_{2,y}}{\sigma_2(t)} \right)^2 \right] \quad . \quad (3.41)$$

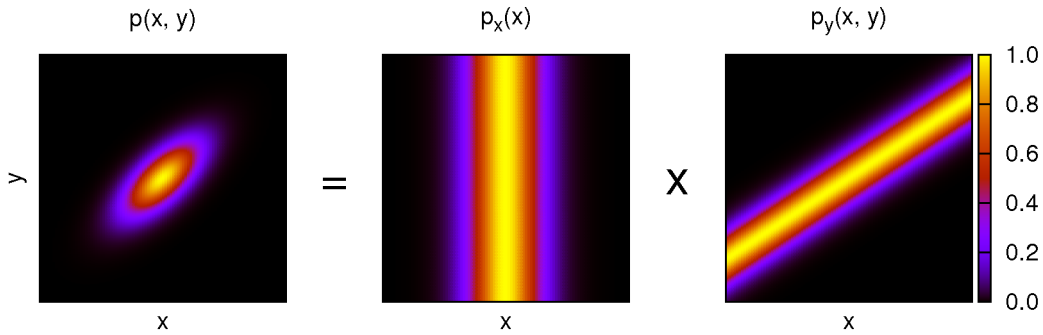


Figure 3.14.: A two-dimensional probability distributions $p(x, y)$ (left panel) is described as the product of two probability distributions $p_x(x)$ (center panel) and $p_y(x, y)$ (right panel).

With these solutions at hand, the rate constant $r(t)$ for barrier crossing in the two-dimensional free energy landscape given by the model for non-equilibrium elasticity is determined. As a first step, the rate constant for one path across the free energy

landscape $r_x(y, t)$ is evaluated. First, the well integral is evaluated as

$$\int_{\text{well}} e^{-\beta G_{NE}} = \int_{-\infty}^{x_b} p_x(x, t) dx \int_{-\infty}^{\infty} p_y(x, y, t) dy \quad (3.42)$$

$$= \frac{1}{2} \left\{ 1 + \operatorname{erf} \left[\frac{x_b - \langle x(t) \rangle}{\sqrt{2}\sigma_x(t)} \right] \right\} \quad (3.43)$$

For the barrier integral, the non-equilibrium free energy is replaced by a linear approximation to the barrier,

$$G_{NE}(x, y, t) \approx G_{NE}(x_b, y, t) + \Delta G \quad (3.44)$$

with

$$\Delta G = (x - x_b) \left. \frac{\partial}{\partial x} G(x, y, t) \right|_{x=x_b} \quad (3.45)$$

$$= (x - x_b)[(k_{\parallel} + k_p)x_b - k_p vt - \alpha k_{\perp} y] \quad (3.46)$$

This approximation is justified because the regions where this approximation becomes inaccurate only contribute little to the integral since the exponential of ΔG is integrated. Therefore, the error due to this approximation is small.

This expression is evaluated for $y = x_b/\alpha$, which is the minimum of $G(x_b, y, t)$ along y , to account for the path parallel to the reaction coordinate with the minimal barrier height, as required before. This results in

$$\Delta G = (x - x_b)[(k_{\parallel} - k_{\perp} + k_p)x_b - k_p vt] \quad (3.47)$$

Subsequently, this expression is further approximated by replacing ΔG with its mean over time for $0 \leq t \leq T$. Here, $T = (k_{\parallel} - k_{\perp} + k_p)x_b/(k_p v)$ is the time when the time-dependent minimum of $G(x, y, t)$ reaches $x = x_b$. Thus, a final expression for ΔG is given by

$$\Delta G = (x - x_b) \frac{1}{2} (k_{\parallel} - k_{\perp} + k_p)x_b \quad (3.48)$$

By employing that $e^{\beta G_{NE}(x_b, y, t)} = 1/p(x_b, y, t)$, the integral over the barrier is evaluated as

$$\int_{\text{barrier}} e^{\beta G_{NE}} = e^{\beta G_{NE}(x_b, y, t)} \cdot \int_{-\infty}^{x_b} e^{\Delta G} dx \quad (3.49)$$

$$= \frac{1}{\frac{1}{2}\beta(k_0 - k_{\perp} + k_p)x_b p(x_b, y, t)} \quad (3.50)$$

By combining equations (3.34), (3.43), and (3.50), the rate constant accounting for

3.3 Non-equilibrium elastic behavior of macromolecules

one path across the free energy landscape is given by

$$r_x(y, t) = \frac{\beta(k_{\parallel} - k_{\perp} + k_p)D}{1 + \operatorname{erf}\left[\frac{x_b - \langle x(t) \rangle}{\sqrt{2}\sigma_x(t)}\right]} p(x_b, y, t) \quad . \quad (3.51)$$

With this expression for $r_x(y, t)$, and by employing that $p(x_b, y, t) = p_x(x_b, t) \cdot p_y(x_b, y, t)$, the rate constant $r(t)$ for a non-equilibrium process in a two-dimensional free energy landscape given by equation (3.35) is evaluated as

$$r(t) = \frac{\beta(k_{\parallel} - k_{\perp} + k_p)D}{1 + \operatorname{erf}\left[\frac{x_b - \langle x(t) \rangle}{\sqrt{2}\sigma_x(t)}\right]} p_x(x_b, t) \int_{-\infty}^{\infty} p_y(x_b, y, t) dy \quad (3.52)$$

$$= \frac{\beta(k_{\parallel} - k_{\perp} + k_p)D}{1 + \operatorname{erf}\left[\frac{x_b - \langle x(t) \rangle}{\sqrt{2}\sigma_x(t)}\right]} p_x(x_b, t) \quad . \quad (3.53)$$

Along with mean forces $\langle F(t) \rangle$ given by equation (3.29) and the flux over the barrier $dQ(t)/dt$ defined by equation (3.30), mean rupture forces $\langle F_{\text{yield}} \rangle$ for barrier crossing within the free energy landscape from the model for non-equilibrium elasticity are calculated according to equation (3.32).

Application to importin- β

To test whether Bell's model is sufficient to describe the elastic response of importin- β at different probe velocities in FPMD simulations [2], mean forces $F(t)$ according to equation (3.18) were fitted to the respective stretching forces (figure 3.15A). Probe velocities ranged from 0.08 m/s to 20 m/s. Although these fits described the forces well, the spring constants determined from these fits were velocity-dependent (figure 3.15B). This result shows that Bell's model is not sufficient to describe the elastic behavior of importin- β present in these FPMD simulations.

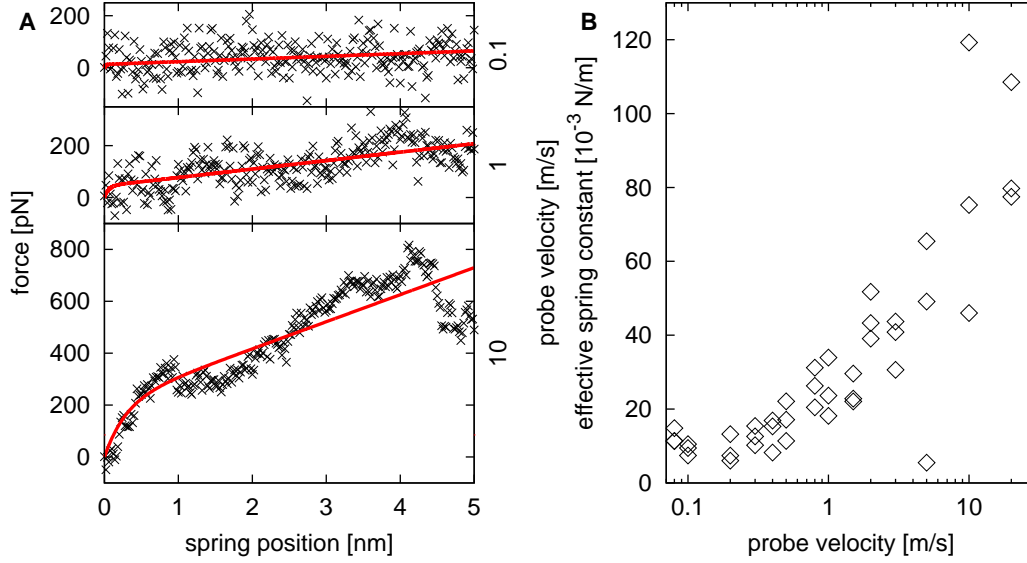


Figure 3.15.: A: Sample force curves obtained during stretching of importin- β at different stretching velocities (black symbols) with force fits under the assumption of a one-dimensional free energy landscape according to equation (3.18). B: Velocity dependence of spring constants determined by these fits.

However, the occurrence of velocity-dependent spring constants at fast probe velocities was predicted by the model for non-equilibrium elasticity. Hence, a force fit according to the model for non-equilibrium elasticity using equation (3.29) to a data set containing all force curves was carried out. This fit reproduced the force curves for all probe velocities with the same parameters. Only four parameters – $k_{\parallel} = 0.10$ N/m, $k_{\perp} = 0.09$ N/m, $\alpha = 0.51$, and $D = 1.64 \times 10^{-10}$ m²/s – describe all force profiles for probe velocities over two orders of magnitude (figure 3.16, green lines). The equilibrium spring constant $k_{eq} = k_{\parallel} - k_{\perp} = 11$ mN/m is in good agreement with the spring constant $k = (10 \pm 4)$ mN/m determined from equilibrium fluctuations and slow stretching simulation (figure 3 in reference [2]).

For comparison, additional individual fits to each single force curve were carried out (figure 3.16, blue lines). These individual fits reproduced the force curves very well and yielded similar parameters as the previous approach. These results show that the model for non-equilibrium elasticity reproduces the elastic behavior of importin- β very well.

3.3 Non-equilibrium elastic behavior of macromolecules

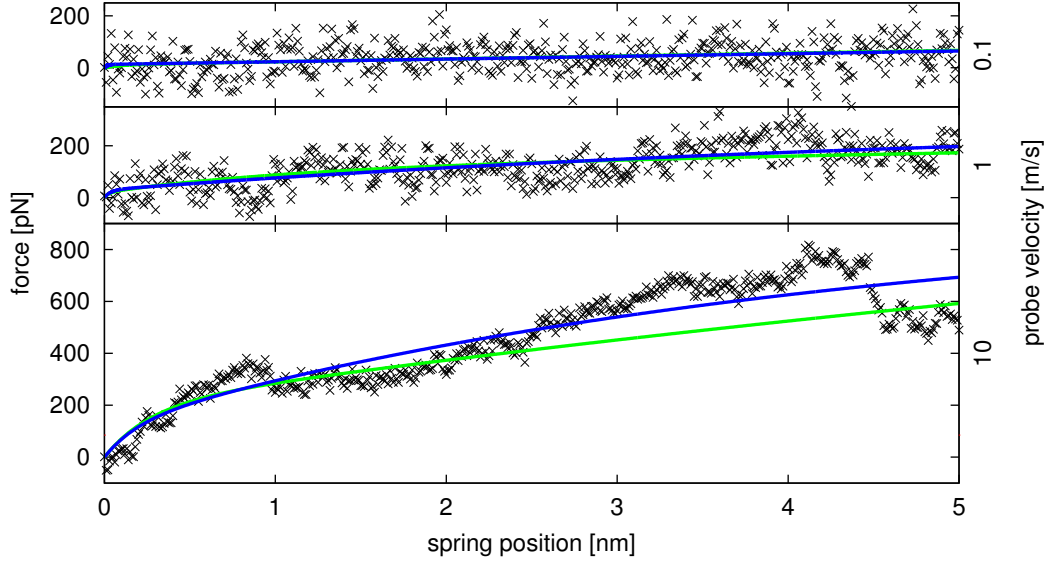


Figure 3.16.: Sample force curves obtained during the stretching of importin- β (black symbols) with fits under the assumption of a two-dimensional free energy landscape according to equation (3.29) (colored lines). Green and blue lines denote fits to all forces simultaneously and fits to individual forces, respectively.

Application to the southern bean mosaic virus

So far, the only test system for the model for non-equilibrium elasticity is the exceptionally flexible nuclear transport receptor importin- β . To test the universal applicability of this model, a mechanically different second test system was chosen. Hence, the elastic behavior of the very stiff outer shell of the southern bean mosaic virus, which was studied in recent FPMD simulations [29], was investigated.

To reduce noise from the force profiles, forces were smoothed by replacing them with one-dimensional linear functions $F = k_{\text{eff}}\Delta z$, where k_{eff} denotes an effective spring constant and Δz describes the indentation of the viral shell. Probe velocities ranged from 2 m/s to 50 m/s. As for importin- β , spring constants derived from this one-dimensional approach were found to be velocity-dependent (figure 3.17A, symbols).

Like in the case of importin- β , simultaneous fits of non-equilibrium forces using equation (3.29) to smoothed forces from all simulations provided fit results of $k_{\parallel} = 4.3 \text{ N/m}$, $k_{\perp} = 3.4 \text{ N/m}$, $\alpha = 0.058$, and $D = 4.2 \times 10^{-10} \text{ m}^2/\text{s}$. A conversion of these non-equilibrium forces into linear forces $k_{\text{eff}}\Delta z$ revealed that effective spring constants obtained by a one-dimensional model reach a limiting value at probe velocities below $\sim 0.1 \text{ m/s}$ (figure 3.17A, line).

To further validate the model for non-equilibrium elasticity, rupture forces occurring upon the rupture of the outer shell of the southern bean mosaic virus were predicted and compared to rupture forces obtained from the FPMD simulations. Therefore, mean velocity-dependent rupture forces $\langle F_{\text{yield}} \rangle$ according to equation (3.53) were calculated. This was done by employing mean forces given by equation (3.29) and rate constants $r(t)$ according to equation (3.32). The four parameters k_{\parallel} , k_{\perp} , α , and D obtained from the force fits were used. As an additional parameter, the barrier location x_b was needed to determine rate constants $r(t)$.

Least square fits of velocity-dependent mean rupture forces $\langle F_{\text{yield}} \rangle$ with the barrier location as additional fit parameter to results from the simulation (figure 3.17B, symbols) described the observed forces very well (line). In this way, a barrier location of $x_b = 1.56 \text{ nm}$ was predicted.

Also, a prediction for rupture forces at slow probe velocities $< 0.1 \text{ m/s}$ was made. The predicted rupture force of $\sim 1.4 \text{ nN}$ for slow probe velocities lies in the range of $\sim 0.6 - 2.5 \text{ nN}$ of experimentally determined rupture forces for viral shells [5, 67]

3.3 Non-equilibrium elastic behavior of macromolecules

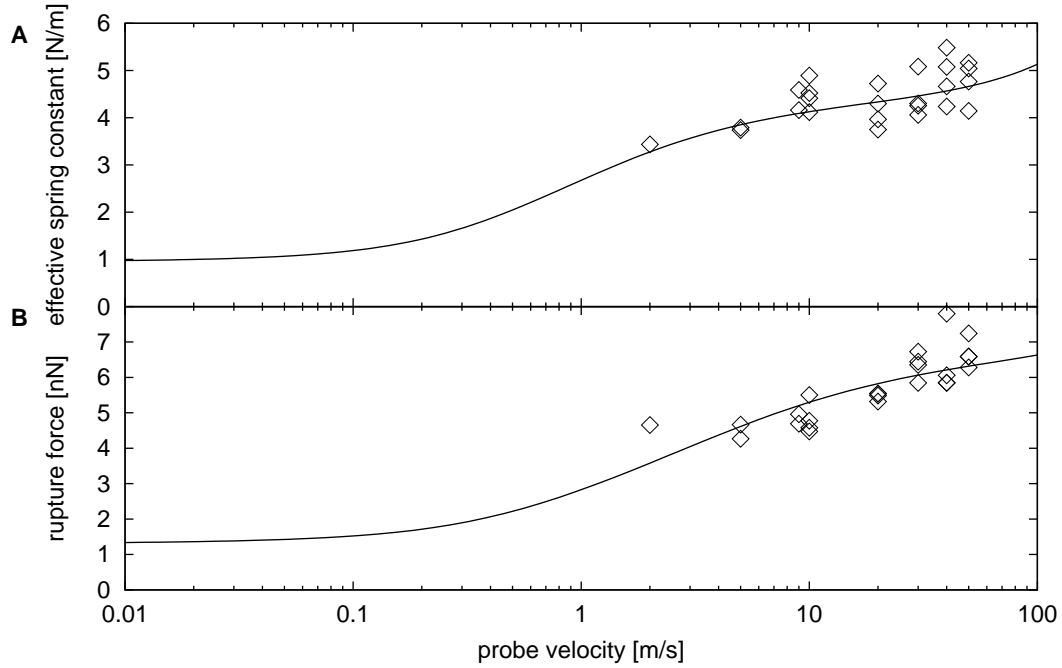


Figure 3.17.: Results from FPMD simulation of the outer shell of the southern bean mosaic virus and comparison to the model for non-equilibrium elasticity. A: Spring constants obtained from forces based on the assumption of a one-dimensional free energy landscape (symbols). The line shows the predicted velocity dependence of these spring constants under the assumption of a two-dimensional free energy landscape given by equation (3.19). B: Rupture forces obtained from the simulations (symbols). The line depicts predicted rupture forces based on the two-dimensional free energy landscape from the model for non-equilibrium elasticity.

4. Summary and Conclusion

4.1. General

In this PhD thesis, the behavior of several different proteins under force was studied. In the first part, the mechanical unfolding of the membrane protein bacteriorhodopsin was analyzed by FPMD simulations to quantify its structural stability. The results from the simulations were validated by comparison with results from AFM experiments with the same protein including forces and the locations of anchor point residues that resist mechanical unfolding. The analysis of atomic interactions and of structural rearrangements during unfolding revealed that anchor points are stabilized by intramolecular hydrogen bonds and hydrophobic contacts. This stabilization of anchor points, in turn, governs the mechanical stability of bacteriorhodopsin. Furthermore, these atomic interactions were found to form a dynamic interaction network.

In the second part, FPMD simulations were applied to study the flexibility of the nuclear transport receptor importin- β . In this way, the protein was characterized as a fully reversible molecular spring. This mechanical characterization allowed the determination of the amount of binding energy of importin- β -cargo complexes that is stored via mechanical deformation of importin- β . Investigation of the hydrophobic core of the protein and subsequent comparison to the structurally similar but mechanically different protein Rop revealed that a special highly mobile hydrophobic core is the origin of the flexibility of importin- β .

The last part of this thesis covers non-equilibrium elastic behavior of macromolecules. Such a behavior, which was observed in FPMD simulations of importin- β and of the outer shell of the southern bean mosaic virus, is manifested by velocity-dependent spring constants. To this end, a new model for macromolecular elasticity based on a multi-dimensional free energy landscape was introduced. This model was successfully applied to describe elastic forces present in both systems. The results gained from this model characterize internal protein motions that are otherwise not accessible.

4.2. Velocity-Dependent Mechanical Unfolding of Bacteriorhodopsin Is Governed by a Dynamic Interaction Network

The study “Velocity-Dependent Mechanical Unfolding of Bacteriorhodopsin Is Governed by a Dynamic Interaction Network” [1] addressed open questions concerning mechanical unfolding of bacteriorhodopsin. In this work, FPMD simulations were employed to extract single bacteriorhodopsin molecules from an atomistic purple membrane model (figure 1 in reference [1]). Different extraction velocities were employed to quantify a possible velocity dependence of unfolding pathways and associated observables. Furthermore, extraction simulations were carried out towards both sides of the membrane.

To facilitate complete extraction of a single protein within a reasonably sized simulation system, a new extraction protocol was developed and applied. In this protocol, extraction simulations were stopped at regular intervals and extracted and unfolded parts of the protein were removed. Voids generated by peptide removal were filled with water, the ion concentration was adjusted to keep the system neutral, and the extraction simulations were continued at the cleaved part of the protein with properly adjusted forces.

To validate the extraction simulations against the AFM experiments, forces from both methods were compared. The forces obtained during the extraction simulations showed strong similarities to forces obtained from AFM experiments. In both cases, four distinct peaks of decreasing height were observed (figure 3A and B in reference [1]). However, the peak forces from both methods did not follow a common logarithmic dependence (figure 3C in reference [1]). This results suggests that the unfolding pathways differ between both methods.

The question how the unfolding proceeds at an atomistic level was answered by investigating snapshots of bacteriorhodopsin taken from the extraction simulations (figure 4A-F in reference [1]). These snapshots revealed an unfolding pathway in which the single transmembrane helices unfolded turn-by-turn. Furthermore, the protein was unfolded along its amino acid sequence. The end of the protein that was attached to the probe potential unfolded first.

The question about the relation between the observed force profiles and the unfolding pathway was addressed by comparing the force development upon extraction to the different stages of unfolding (figure 4G in reference [1]). In this way, the extraction and unfolding of each of the first three helix pairs and the last helix alone along the unfolding sequence were revealed to induce a force peak.

This result is in contrast to the proposal from the AFM experiments that the concerted unfolding of helix pairs gives rise to a force peak. However, since the force profiles and the locations of the force peaks are similar between experiments and simulations, a sequential unfolding of single helices is also probable to occur in the

experiments.

To further validate the simulations against the experiments, locations of anchor points residues that stabilize bacteriorhodopsin were determined from the extraction trajectories and compared to anchor points obtained from the experiments. Instead of relying on worm-like chain polymer model fits to forces, as in the interpretation of the AFM experiments, atom coordinates were employed to determine anchor point locations. For this approach, that is only possible in FPMD simulations, a new method was developed and applied.

This method relies on the fact that anchor points interrupt the mechanical unfolding of bacteriorhodopsin. As a consequence, the transition point between the folded and the unfolded part of the protein remains stable for a longer period at anchor points than at residues that do not interrupt the mechanical unfolding. Accordingly, anchor points are the most frequent transition points. To that end, transition points were determined for all times. Transition points that were observed the most frequently were classified as anchor points (figure 2 in reference [1]).

By comparing anchor points from the simulations with anchor points from the experiments, a very good agreement between both methods was seen. This shows that the residues that govern stability are largely independent of the applied extraction velocity (figure 5A and B in reference [1]).

The question how anchor points are stabilized within the protein was addressed by looking at snapshots and interaction energies from extraction simulations before and after the detection of an anchor point. This way, hydrophobic contacts and hydrogen bonds within the protein were determined as the dominant causes for the stabilization of anchor points (figure 5C, D, and E in reference [1]).

More importantly, even on the short time scales of the simulations, a dynamic rearrangement of interactions was observed. This behavior raises the hypothesis of velocity-dependent unfolding pathways of bacteriorhodopsin. On the much longer time scales accessible by AFM, more such dynamic interaction rearrangements can be expected that alter the unfolding pathway of the protein. Furthermore, a change in unfolding pathways should be visible in FPMD simulations for sufficiently small extraction velocities.

To look for evidence of a change in unfolding pathways upon different probe velocities, bacteriorhodopsin unfolding at an extraction velocity of $v = 0$ m/s was simulated by stopping the unfolding process in special relaxation simulations (figure 6 in reference [1]). These simulations revealed two relaxation processes. First, a fast relaxation in the order of pico- to nanoseconds was seen. This relaxation is likely to be caused by the loss of friction build up during the fast non-equilibrium extraction simulations before. An estimation revealed that this relaxation process would vanish at extraction velocities below ~ 1 m/s, which is in good agreement with simulations performed at similar extraction velocities.

The second relaxation process takes place at much slower time scales of 5 to 50 ns. Here, structural rearrangements that are characteristic for the extraction of helix

4.3 Extreme Reversible Elasticity of Importin- β

pairs were observed. These rearrangements are due to a quasi-static pulling process. Furthermore, this behavior serves as an evidence for velocity-dependent unfolding pathways of bacteriorhodopsin.

The newly developed protocol for extraction simulations where unfolded parts of the protein are removed regularly proved to be useful in this work. With its easy implementation, this protocol can be applied to the mechanical unfolding of other proteins in FPMD simulations. This is of great use for protein that are too large to be unfolded otherwise.

It will be interesting to know, if the dynamic interaction network that stabilizes bacteriorhodopsin is unique for this protein or whether it is a common property of membrane proteins. Extraction simulations of other membrane proteins will serve to answer this question. Furthermore, it is not known whether soluble proteins also show this behavior. An according analysis of unfolding trajectories of soluble proteins will solve this question.

The successful determination of anchor points based on the frequency of transition points between folded and unfolded protein regions will be valuable for the analysis of future unfolding simulations. However, the current implementation of this algorithm based on line fits to residue deviations only works for proteins that unfold in sequential manner. For proteins that unfold non-sequentially, a different implementation for the determination of transitions between folded and unfolded protein regions is required.

A direct comparison of anchor points determined by transition points and by worm-like chain fits is still missing. To this end, FPMD simulations in which anchor points are determined by both methods are required. Unfortunately, this was not possible with the bacteriorhodopsin extraction simulations. This comparison must be carried out by future FPMD simulations on protein systems where both approaches can be applied. Furthermore, such a comparison would allow a validation of the accuracy of the indirect worm-like chain method by the direct atomistic anchor points approach.

4.3. An Unusual Hydrophobic Core Confers Extreme Flexibility to HEAT Repeat Proteins

The study “An Unusual Hydrophobic Core Confers Extreme Flexibility to HEAT Repeat Proteins” [2] focused on the flexibility of unbound importin- β and characterized its mechanical properties. Furthermore, the atomistic mechanism behind the flexibility of the protein was analyzed.

To answer the question about the extent of the flexibility of importin- β , FPMD simulations were carried out in which the protein was stretched. It turned out, that importin- β can be stretched from an equilibrium end-to-end distance of 13 nm up to 21 nm without structural break-ups (figure 1 in reference [2]). Also, these simulations answered the question whether changes between different conformations

of importin- β are continuous or discrete. The forces obtained upon stretching of the protein showed no distinct steps, thus suggesting that importin- β possesses a continuous flexibility (figure 3A in reference [2]).

To answer the questions whether the enforced structural changes of importin- β are reversible or irreversible, relaxation simulations were employed. To this end, snapshots of elongated conformations of importin- β from stretching simulations were used as starting point for free MD simulations, i.e. without the presence of a pulling potential. These simulations revealed that the protein relaxed back to its equilibrium conformation within only tens of nanoseconds (figure 2 in reference [2], colored lines). Furthermore, the RMSD of relaxed structures reached values of only 3 Å, which were also seen in equilibrium simulations of importin- β (figure 2 in reference [2], gray lines).

These results classify importin- β as a fully reversible molecular spring. The biological relevance of this behavior is the easy release of binding energy upon the release of binding partners, as proposed by Conti et al. [9].

With importin- β classified as a molecular spring, the question about the energy that is required to change the protein to a different conformation is easily answered. The deformation energy of importin- β is given by

$$E(d) = \frac{k}{2}(d - d_0)^2 \quad (4.1)$$

with k as the spring constant of the protein and d_0 as its equilibrium end-to-end distance. To this end, the spring constant of importin- β was determined from equilibrium fluctuations and from slow stretching simulations (figure 3 in reference [2]). The equilibrium end-to-end distance was determined from equilibrium simulations.

In this way, a spring constant of (10 ± 4) mN/m was determined. With this result, the amount of the binding energy between importin- β and different cargoes that is stored through deformation of importin- β was estimated (table 1 in reference [2]).

The large global changes governing the spring-like behavior of importin- β are the sum of small local changes. To identify these local changes, and thus to answer the question about the atomistic origin of the flexibility of importin- β , the hydrophobic core that connects adjacent HEAT repeats was investigated. To this end, it was compared with the hydrophobic core of the protein Rop. Similar to a succession of HEAT repeats in importin- β , Rop consists of two protein units with two antiparallel helices each. However, in contrast to importin- β , Rop is known to be very rigid.

To look for differences between the internal dynamics of the hydrophobic cores of both proteins, the mobility of hydrophobic amino acid side chains within each hydrophobic core was quantified. This was done by determining the torsional degrees of freedom of these side chains. A comparison of the side chain mobility of importin- β and Rop showed that the hydrophobic cores of both proteins behave very differently (figure 4 in reference [2]). The hydrophobic core of Rop showed only very little internal dynamics. In contrast, hydrophobic amino acid side chains within the

4.3 Extreme Reversible Elasticity of Importin- β

hydrophobic core of importin- β showed a high degree of flexibility.

This result suggests that successions of HEAT repeats allow importin- β to adapt to external changes while Rop likely does not have this ability. To test this hypothesis, a triplet of HEAT repeats was extracted from importin- β . Connecting loops were removed to allow better comparison to Rop. Upon external stress exerted in FPMD simulations, the HEAT repeats showed structural rearrangement at low forces, as quantified by their RMSD (figure 5 in reference [2]). In contrast, Rop remained stable up to high forces. These results confirmed the hypothesis that importin- β is easily able to adapt to external changes.

This raised a new question about how accumulated local changes of hydrophobic cores between adjacent HEAT repeats can cause the global spring-like energetic behavior of importin- β . To this end, Coulombic and Lennard-Jones energy contributions between and within HEAT repeats at different elongations were determined (figure 6 in reference [2]). The Lennard-Jones energy contributions between adjacent HEAT repeats, that characterize hydrophobic interactions, were found to increase with increasing end-to-end distance. This result showed that local changes within hydrophobic cores between HEAT repeats give rise to the global energetics of importin- β .

Furthermore, the sum of all energy contributions revealed a spring-like energy. However, the enthalpic spring constant derived in this way was still four times larger than the spring constant obtained from equilibrium and stretching simulations. This raised the question, whether an entropy compensation was present to decrease Gibbs free energy. To this end, the entropy of importin- β has to increase with increasing length d . This behavior was confirmed by entropy estimations of importin- β from simulations where the length of the protein was kept fixed (figure 7A in reference [2]). Moreover, enthalpic terms with entropy corrections covered the Gibbs free energy profile (figure 7B in reference [2]). In this way, local enthalpic interactions between adjacent HEAT repeats and an increasing entropy upon elongation determine the spring constant of importin- β .

This part of the thesis provided a thorough investigation of the flexibility of importin- β . It was revealed that this flexibility is governed by a special hydrophobic core between adjacent HEAT repeats of the protein. Moreover, its flexibility allows importin- β to bind to a variety of cargoes and thus designates its function as a universal nuclear transport receptor.

The flexibility of importin- β , that determines its function, points to a new view of proteins, where their function is governed by their structure as much as by their flexibility.

Besides importin- β , other nuclear transport receptors, such as transportin or exportin, are also built of HEAT repeats. Stretching simulations of these proteins will show whether flexibility is a common property of this class of proteins.

One restriction of this work is that only the flexibility of unbound importin- β was investigated. Thus, a new question is raised, concerning whether the elastic behavior

of importin- β within complexes with cargoes is different compared to its unbound state. The large entropy compensation revealed in this work suggests that the elastic properties of importin- β may change within complexes.

To address the question whether the type of hydrophobic core seen in importin- β is a general quantity of flexible proteins, MD simulations and corresponding analysis of hydrophobic cores of different flexible repeat proteins, such as ankyrin, will be necessary.

Although the special nature of the hydrophobic core of importin- β was revealed in this work, the mechanism how its amino acid composition affects flexibility is still unclear. To this end, future studies are necessary to link the origin of the flexibility of the protein to single amino acids within the hydrophobic core. Furthermore, based on this knowledge, mutations within the hydrophobic core of importin- β may lead to an alternation of its flexibility. If this approach would be successful, it may facilitate the design of flexible proteins.

4.4. Non-equilibrium elastic behavior of macromolecules

The study “Non-equilibrium elastic behavior of macromolecules” (chapter 3.3) investigated a previously unknown elastic behavior of macromolecules that is not explained by standard models. The starting point for this investigation was the observation of velocity-dependent spring constants in FPMD simulations of importin- β (figure 3.15B in chapter 3.3) and of the outer shell of the southern bean mosaic virus [29] (figure 3.17A in chapter 3.3).

To explain and quantify this behavior, a new model for the elasticity of macromolecules was introduced. In contrast to previous models, that are based on an underlying one-dimensional free energy landscape describing only the length of the macromolecule as a reaction coordinate, a two-dimensional free energy landscape was introduced. Using the additional dimension, perpendicular modes of molecular motion orthogonal to the reaction coordinate were described in addition to motions along the length of the macromolecule. Moreover, both modes are coupled.

The enforced elongation of a macromolecule through, e.g., an AFM cantilever was described by an additional moving probe potential that acts only along the reaction coordinate. In contrast, motions along the perpendicular coordinate were unrestricted. Analytical solutions to the Smoluchowski equation for the probability distribution of the macromolecule within the free energy landscape revealed velocity-dependent pathways. At slow probe velocities, a path along minimal energies is taken (figure 3.11A in chapter 3.3). At fast probe velocities, relaxation along the perpendicular coordinate does take place and, thus, a path along higher energies is taken. These different pathways lead to different energies and thus different elastic properties along the reaction coordinate (figure 3.11B in chapter 3.3). In turn, this causes

4.4 Non-equilibrium elastic behavior of macromolecules

an velocity dependence of spring constants. These results show that this new model explains qualitatively the non-equilibrium elastic behavior seen for importin- β and the southern bean mosaic virus.

This model was applied to quantitatively describe the elastic forces obtained from FPMD simulations of importin- β and of the southern bean mosaic virus. In both cases, all force curves obtained over a broad range of probe velocities could be described with a single set of only four parameters (figure 3.16 and figure 3.17A in chapter 3.3).

As a next test of the model, rupture forces occurring upon the rupture of the outer shell of the southern bean mosaic virus were calculated. To this end, a formalism for the calculation of rate constants for barrier crossing within the two-dimensional free energy landscape of the model for non-equilibrium elasticity was developed. With this formalism, rupture forces from the FPMD simulations were predicted very well with the barrier location as the only fit parameter (figure 3.17B in chapter 3.3).

Our newly introduced model for non-equilibrium elastic behavior of macromolecules correctly predicted elastic forces for two mechanically different systems: the soft nuclear transport receptor importin- β and the stiff outer shell of the southern bean mosaic virus. This shows that this model is universally applicable. The prediction of yielding forces further strengthened the model. Also, perpendicular molecular modes were quantified. This can only be achieved with this model.

The non-equilibrium elastic behavior describes in this study has been only observed in FPMD simulations that employ fast probe velocities above 0.1 m/s. Hence, the question whether this behavior can also be seen at slower probe velocities is still open.

So far, our model was only applied to two test systems. However, it is likely that other macromolecules exhibit the same behavior. To this end, force curves from existing experiments or simulations need to be analyzed with respect to the model for non-equilibrium elasticity. Also, experiments or simulations on proteins that are similar to both studied test systems and therefore are likely to show the same non-equilibrium elastic behavior may be carried out to look for a further validation of this model.

5. Declaration of own contribution

Hereby I declare that I wrote this dissertation on my own with only the cited references and the references within the publications as help.

The publication "Velocity-Dependent Mechanical Unfolding of Bacteriorhodopsin Is Governed by a Dynamic Interaction Network" presents an extension of my diploma thesis. Large parts of the work leading to this publication were done during the time of my PhD thesis, thus qualifying this work as a part of my dissertation. Specifically, the systematic comparison of anchor points between AFM and MD, the analysis of energetics stabilizing anchor points, the quantification of the relaxation processes, and the central thesis of velocity dependent unfolding pathways were done during my time as a PhD student.

The other two studies emerged solely during my PhD studies. Nicole Dölker and Ulrich Zachariae provided supervision for both studies. Mareike Zink was of great help by testing the physical model and the application to results from simulation performed by her.

6. Acknowledgements

First, I would like to thank my supervisor Prof. Dr. Helmut Grubmüller for his constant support and guidance during the time my thesis. Many thanks also go to Prof. Dr. Ralf Ficner and Prof. Dr. Müller who agreed to participate in my thesis committee and showed great interest in my work.

I would also like to thank my cooperation partners Nicole Dölker, Ulrich Zachariae, and Mareike Zink who helped me a lot with my work and who provided a wonderful cooperation within my department.

In this spirit, I also thank Prof. Dr. Hermann Gaub and Prof. Dr. Daniel Müller for their fruitful cooperation in Munich and Dresden, respectively. Big thanks go also to Achim Dickmanns and Thomas Monecke from the University of Göttingen for their cooperation on nuclear transport receptors.

I thank Ulrike Gerischer for the proof-reading of my manuscripts and of this current thesis.

Finally, I would like to thank the Department for Theoretical and Computational Biophysics. Special thanks go to Eveline Heinemann, who showed everyday the importance of having a good secretary. Special thanks go also to Ansgar Esztermann, David Köpfer, Jan Henning Peters, Jürgen Haas, Jürgen von Grünhagen, Magdalena Roos, Matthias Popp, and Rudolf Schemm.

Special thanks also go to my family, Wilfried, Gerti, and Wibke.

Bibliography

- [1] C. Kappel and H. Grubmüller. Velocity-dependent mechanical unfolding of bacteriorhodopsin is governed by a dynamic interaction network. *Biophysical Journal*, 100:1109–1119, 2011.
- [2] C. Kappel, U. Zachariae, N. Dölker, and H. Grubmüller. An unusual hydrophobic core confers extreme flexibility to HEAT repeat proteins. *Biophysical Journal*, 99:1596–1603, 2010.
- [3] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.
- [4] M. D. Shoulders and R. T. Raines. Collagen structure and stability. *Annual Review of Biochemistry*, 78:929–958, 2009.
- [5] W. H. Roos, I. L. Ivanovska, A. Evilevitch, and G. J. L. Wuite. Viral capsids: Mechanical characteristics, genome packaging and delivery mechanisms. *Cell. Mol. Life Sci.*, 64:1484–1497, 2007.
- [6] D. L. de Groot, S. Hayward, D. M. F. van Aalten, A. Amadei, and H. J. C. Berendsen. Domain motions in bacteriophage t4 lysozyme: A comparison between molecular dynamics and crystallographic data. *Proteins*, 31:116–127, 1998.
- [7] H. Nakamichi and T. Okada. Crystallographic analysis of primary visual photochemistry. *Angewandte Chemie*, 118:4376–4379, 2006.
- [8] D. P. Denning, S. S. Patel, V. Uversky, A. L. Fink, and M. Rexach. Disorder in the nuclear pore complex: The fg repeat regions of nucleoporins are natively unfolded. *PNAS*, 100:2450–2455, 2003.
- [9] E. Conti, C. W. Müller, and M. Stewart. Karyopherin flexibility in nucleocytoplasmic transport. *Curr. Opin. Struct. Biol.*, 16:237–244, 2006.
- [10] G. Binnig, C. F. Quate, and C. Gerber. Atomic force microscope. *Physical Review Letters*, 56:930–933, 1986.
- [11] A. Ashkin. Acceleration and trapping of particles by radiation pressure. *Physical Review Letters*, 24:156–159, 1970.
- [12] S. Chu, J. E. Bjorkholm, A. Ashkin, and A. Cable. Experimental observation of optically trapped atoms. *Physical Review Letters*, 57:314–317, 1986.
- [13] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276:1109–1112, 1997.

- [14] H. Dietz and M. Rief. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *PNAS*, 101:16192–16197, 2004.
- [15] Q. Peng and H. Li. Atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: Evidence for a kinetic partitioning mechanism. *PNAS*, 105:1885–1890, 2008.
- [16] 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.
- [17] C. Möller, D. Fotiadis, K. Suda, A. Engel, M. Kessler, and D. J. Müller. Determining molecular forces that stabilize human aquaporin-1. *Journal of Structural Biology*, 142:369–378, 2003.
- [18] K. T. Sapra, P. S.-H. Park, K. Palczewski, and D. J. Müller. Mechanical properties of bovine rhodopsin and bacteriorhodopsin: Possible roles in folding and function. *Langmuir*, 24:1330–1337, 2008.
- [19] M. R. Falvo, S. Washburn, R. Superfine, M. Finch, F. P. Brooks Jr., V. Chi, and R. M. Taylor II. Manipulation of individual viruses: Friction and mechanical properties. *Biophysical Journal*, 72:1396–1403, 1997.
- [20] Gwangrog Lee, Khadar Abdi, Yong Jiang, Peter Michaely, Vann Bennett, and Piotr E Marszalek. Nanospring behaviour of ankyrin repeats. *Nature*, 440:246–249, 2006.
- [21] E. L. Florin, V. T. Moy, and H. E. Gaub. Adhesion forces between individual ligand-receptor pairs. *Science*, 264:415–417, 1994.
- [22] H. Grubmüller, B. Heymann, and P. Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271:997–999, 1996.
- [23] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophysical Journal*, 72:1568–1581, 1997.
- [24] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophysical Journal*, 75:662–671, 1998.
- [25] F. Gräter, J. Shen, H. Jiang, M. Gautela, and H. Grubmüller. Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. *Biophysical Journal*, 88:790–804, 2005.
- [26] M. Sotomayor, D. P. Corey, and K. Schulten. In search of the hair-cell gating spring: Elastic properties of ankyrin and cadherin repeats. *Structure*, 13:669–682, 2005.
- [27] M. Kessler and H. Gaub. Unfolding barriers in bacteriorhodopsin probed from the cytoplasmic and the extracellular side by AFM. *Structure*, 14:521–527, 2006.

-
- [28] U. Zachariae and H. Grubmüller. Importin- β : Structural and dynamic determinants of a molecular spring. *Structure*, 16:905–916, 2008.
- [29] Mareike Zink and Helmut Grubmüller. Mechanical properties of the icosahedral shell of southern bean mosaic virus: a molecular dynamics study. *Biophysical Journal*, 96:1350–1363, 2009.
- [30] G. I. Bell. Models for the specific adhesion of cells to cells. *Science*, 200:618–627, 1978.
- [31] O. Dudko, G. Hummeln, and A. Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Physical Review Letters*, 96:108101–1–4, 2006.
- [32] B. Heymann and H. Grubmüller. Dynamic force spectroscopy of molecular adhesion bonds. *Physical Review Letters*, 84:6126–6129, 2000.
- [33] E. Evans and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophysical Journal*, 72:1541–1555, 1997.
- [34] M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, 289:457–484, 1927.
- [35] 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.
- [36] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.
- [37] W. F. van Gunsteren and H. J. C. Berendsen. *Gromos-87 manual*. Biomos BV Nijenborgh 4, 9747 AG Groningen, 1987.
- [38] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4:187–217, 1983.
- [39] R. W. Hockney and S. P. Goel. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics*, 14:148–158, 1972.
- [40] 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.
- [41] J.-L. Popot and D. M. Engelman. Membrane protein folding and oligomerization: The two-stage model. *Biochemistry*, 29:4031–4037, 1990.
- [42] W. Stoeckenius and R. A. Bogomolni. Bacteriorhodopsin and related pigments of halobacteria. *Ann. Rev. Biochem.*, 52:587–616, 1982.

- [43] U. Haupts, J. Tittor, and D. Oesterhelt. Closing in on bacteriorhodopsin: Progress in understanding the molecule. *Annual Review of Biophysics and Biomolecular Structure*, 28:367–399, 1999.
- [44] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of λ -phage DNA. *Science*, 265:1599–1600, 1994.
- [45] H. Janovjak, J. Struckmeier, M. Hubain, A. Kedrov, M. Kessler, and D. J. Müller. Probing the energy landscape of the membrane protein bacteriorhodopsin. *Structure*, 12:871–879, 2004.
- [46] M. P. Rout and G. Blobel. Isolation of the yeast nuclear pore complex. *The Journal of Cell Biology*, 123:771–783, 1993.
- [47] A. Radu, M. S. Moore, and G. Blobel. The peptide repeat domain of nucleoporin nup98 functions as a docking site in transport across the nuclear pore complex. *Cell*, 81:215–222, 1995.
- [48] D. Görlich and U. Kutay. Transport between the cell nucleus and the cell cytoplasm. *Annu. Rev. Cell Dev. Biol.*, 15:607–660, 1999.
- [49] S. Frey and D. Görlich. A saturated fg-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *Cell*, 130:512–523, 2007.
- [50] D. Görlich, S. Prehn, R. A. Laskey, and E. Hartmann. Isolation of a protein that is essential for the first step of nuclear protein import. *Cell*, 79:767–778, 1994.
- [51] K. Stade, C. S. Ford, C. Guthrie, and K. Weis. Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell*, 90:1041–1050, 1997.
- [52] M. Fornerod, M. Ohno, M. Yoshida, and I. W. Mattaj. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell*, 90:1051–1060, 1997.
- [53] M. A. Andrade and P. Bork. HEAT repeats in the huntington’s disease protein. *Nature Genet.*, 11:115–116, 1995.
- [54] Y. M. Chook and G. Blobel. Karyopherins and nuclear import. *Current Opinion in Structural Biology*, 11:703–715, 2001.
- [55] K. Ribbeck, G. Lipowsky, H. M. Kent, M. Stewart, and D. Görlich. NTF2 mediates nuclear import of ran. *The EMBO Journal*, 17:6587–6598, 1998.
- [56] A. Smith, A. Brownawell, and I. G. Macara. Nuclear import of ran is mediated by the transport factor NTF2. *Current Biology*, 8:1403–1406, 1998.
- [57] G. Cingolani, C. Petosa, K. Weis, and C. W. Müller. Structure of importin- β bound to the IBB domain of importin- α . *Nature*, 399:221–229, 1999.
- [58] G. Mitrousis, A. S. Olia, N. Walter-Kopp, and G. Cingolani. Molecular basis for the recognition of snurportin 1 by importin- β . *J. Biol. Chem*, 283:7877–7884, 2008.

-
- [59] S. J. Lee, T. Sekimoto, E. Yamashita, E. Nagoshi, A. Nakagawa, N. Imamoto, M. Yoshimura, H. Sakai, K. T. Chong, T. Tsukihara, and Y. Yoneda. Crystal structure of importin- β and SREBP-2 complex. *Science*, 302:1513–1514, 2003.
 - [60] S. J. Lee, Y. Matsuura, S. M. Liu, and M. Stewart. Structural basis for nuclear import complex dissociation by RanGTP. *Nature*, 435:693–696, 2005.
 - [61] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–304, 1940.
 - [62] P. Hänggi, P. Talkner, and M. Borkovec. Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys.*, 62:251–341, 1990.
 - [63] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Review of Modern Physics*, 15:1–89, 1943.
 - [64] M. V. Smoluchowski. Über Brownische Molekularbewegung unter Einwirken äusserer Kräfte und deren Zusammenhang mit der verallgemeinerten Diffusionsgleichung. *Annalen der Physik*, 353:1103–1112, 1915.
 - [65] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Physical Review*, 36:823–841, 1930.
 - [66] G. Hummer and A. Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophysical Journal*, 85:5–15, 2003.
 - [67] W. H. Roos, R. Bruinsma, and G. J. L. Wuite. Physical virology. *Nature Physics*, 6:733–743, 2010.

Bibliography

**A. Velocity-Dependent Mechanical
Unfolding of Bacteriorhodopsin Is
Governed by a Dynamic
Interaction Network**

Velocity-Dependent Mechanical Unfolding of Bacteriorhodopsin Is Governed by a Dynamic Interaction Network

Christian Kappel and Helmut Grubmüller*

Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

ABSTRACT Bacteriorhodopsin is a model system for membrane proteins. This seven transmembrane helical protein is embedded within a membrane structure called purple membrane. Its structural stability against mechanical stress was recently investigated by atomic force microscopy experiments, in which single proteins were extracted from the purple membrane. Here, we study this process by all-atom molecular dynamics simulations, in which single bacteriorhodopsin molecules were extracted and unfolded from an atomistic purple membrane model. In our simulations, key features from the experiments like force profiles and location of key residues that resist mechanical unfolding were reproduced. These key residues were seen to be stabilized by a dynamic network of intramolecular interactions. Further, the unfolding pathway was found to be velocity-dependent. Simulations in which the mechanical stress was released during unfolding revealed relaxation motions that allowed characterization of the nonequilibrium processes during fast extraction.

INTRODUCTION

Membrane proteins mediate a broad range of biochemical functions. Although ~20–30% of all proteins are membrane proteins (1), only little is known about their energetics and their stability against mechanical stress. One important class of membrane proteins are G-coupled protein receptors (GPCRs) (2), which consist of seven transmembrane helices and mediate signals from the extracellular to the cytoplasmic side of a membrane, e.g., light, hormones, or neurotransmitters, by conformational changes. Therefore, knowledge of the mechanical stability and energetics of these proteins is crucial.

The light-driven proton-pump bacteriorhodopsin (BR) is a model system for seven-transmembrane helical proteins and shares many similarities to GPCRs (3,4). Photoinduced isomerization of the chromophore retinal drives a reaction cascade in which the protein undergoes several conformational changes. In vivo, BR trimers and lipid molecules arrange into a remarkably stable two-dimensional hexagonal structure called “purple membrane”.

One versatile tool to study single proteins and their structural stability is atomic force microscopy (AFM) (5), which served, e.g., to study the mechanical properties of biological macromolecules such as polysaccharides (6), DNA (7), and heparin (8). Furthermore, AFM experiments provided insight into the strength of ligand-receptor bonds (9) and the stability of soluble proteins such as titin (10) and fibronectin (11). Also, membrane bound proteins such as sodium-proton antiporters (12) and aquaporin (13) were studied.

The underlying intramolecular forces that govern macromolecular mechanical stability and unfolding have been

studied by force probe molecular dynamics (FPMD) simulations (14,15), e.g., for the unfolding of titin domains (16), folding intermediates (17), the function of titin kinase (18), and the elastic properties of ankyrin (19). Also, the unfolding of membrane proteins was studied using this method (20–22).

In recent AFM experiments (23–32), single BR molecules were mechanically extracted and unfolded from the purple membrane. The forces obtained during extraction revealed a characteristic sawtooth pattern. Using the wormlike chain polymer model, these force peaks were linked to the unfolding of single helices. Contrary to former AFM experiments on soluble multidomain proteins, the order of unfolding was not from the weakest to the strongest link, but along the amino-acid sequence. Also from wormlike chain fits to the force profiles, anchor residues that exhibit considerable resistance against mechanical unfolding were located. These relatively few residues are obviously important for stabilizing the protein.

Although the forces that counteract mechanical unfolding have been precisely measured by AFM, the atomistic origin of the quite diverse behavior of the individual residues is unclear. In this work, the extraction and unfolding of single BR molecules is therefore simulated at atomic level using FPMD simulations. To validate the results from the simulations with the experiments, peak forces and anchor residues were compared. Subsequently, the simulations enabled us to characterize the interactions that provide the anchor residues with such remarkable stability.

To assess a possible influence of the loading rate, which is much larger in the simulations than in the experiments, the extraction simulations were carried out at various loading rates. Simulations in which the extraction process was stopped provided further insight into extraction at very slow velocities.

Submitted November 19, 2010, and accepted for publication January 5, 2011.

*Correspondence: hgrubmu@gwdg.de

Editor: Scott Feller.

© 2011 by the Biophysical Society
0006-3495/11/02/1109/11 \$2.00

doi: 10.1016/j.bpj.2011.01.004

METHODS

Simulation system and details

As starting structure, the BR x-ray structure 1QHJ (33) from the Protein DataBank was chosen. Residue side-chain atoms not resolved in the x-ray structure (Met¹⁶³, Arg²²⁷, and Glu²³²) were modeled using WHAT IF (34). Terminal residues not resolved in the x-ray structure (1–4 and 233–248) were omitted from the model. BR trimers were created by applying the appropriate symmetry operations to the protein.

To obtain a full membrane system, the proteins were embedded within a hydrated POPC lipid bilayer (provided by Peter Tieleman (35)) using the method described in Faraldo-Gómez et al. (36) and employing the program MSMS (37). Nine lipid molecules per monomer (four at the cytoplasmic and five at the extracellular side) were placed in the simulation box (33). We consider POPC a reasonable mimic for the purple membrane, because BR is known to fold into its native functional state within a POPC bilayer (38). Details are provided in the [Supporting Material](#).

A water layer of 5 nm with sodium and chloride ion concentration of 300 mM, similar to the experiments (23), was added to the hydrated system at the cytoplasmic side of the purple membrane model.

Next, energy minimization of 150 steps was performed using steepest descent. Three equilibration runs with increasingly relaxed harmonic position restraints were performed. First, restraints with a force constant of $k = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ were applied to all nonhydrogen atoms of protein and lipid molecules for 200 ps. Second, only the protein restraints were kept for 1 ns. Finally, a 3.5-ns equilibration run was performed without any position restraints.

To provide sufficient space for extracted polypeptides, one more 5-nm water layer was appended at the cytoplasmic side of the system, yielding a water layer of 10 nm in total (Fig. 1 A). This enlarged system was equilibrated for 500 ps.

To avoid periodic boundary artifacts during the extraction of one monomer, the system was replicated to construct a system of four trimers according to the crystal symmetry of a hexagonal lattice. The full simulation box thus contained 12 BR monomers (Fig. 1 B). Periodic boundary conditions were used for all simulations.

The system was equilibrated for 500 ps and used as starting structure for all subsequent cytoplasmic extraction simulations described below. For extraction simulations toward the extracellular side, the 10-nm water layer was placed at the extracellular side of the purple membrane model, and the N-terminal was subjected to a force acting downwards in Fig. 1 A. Both simulation systems comprised 236,124 atoms, with a box size of $12.16 \times 12.16 \times 15.32 \text{ nm}^3$.

All simulations were carried out using the software package GROMACS 3.3 (39). Proteins and ions were described by the OPLS all-atom force field (40). The TIP4P water model (41) was employed. Lipid molecules were described with a unified atom model (42). Partial charges for the retinal were taken from Kandt et al. (43). All other force-field parameters were converted into the OPLS force field as described.

All simulations were run in the NPT ensemble. The system was coupled to a constant temperature of 300 K using a Berendsen thermostat (44) with a relaxation time constant of $\tau = 0.1 \text{ ps}$. A Berendsen barostat with a relaxation time constant of $\tau = 1.0 \text{ ps}$ and a compressibility of $4.5 \cdot 10^{-5} \text{ bar}^{-1}$ was used to keep the pressure at 1 bar. Semiisotropic pressure coupling was used; no coupling was applied in x and y direction, whereas in z direction the box was free to adapt to pressure changes.

Long-range electrostatic interactions beyond 1.0 nm were calculated using particle mesh Ewald summation (45,46). A grid dimension of 0.12 nm and fourth-order b-spline interpolation was used. Lennard-Jones interactions were truncated at 1.0 nm. The length of bonds involving hydrogen atoms were constrained using LINCS (47). An integration time step of 2 fs was used.

FPM simulations with various extraction velocities were carried out for extraction toward both the cytoplasmic and the extracellular side. In each simulation, the C_α atom of the C-terminus (N-terminus, respectively)

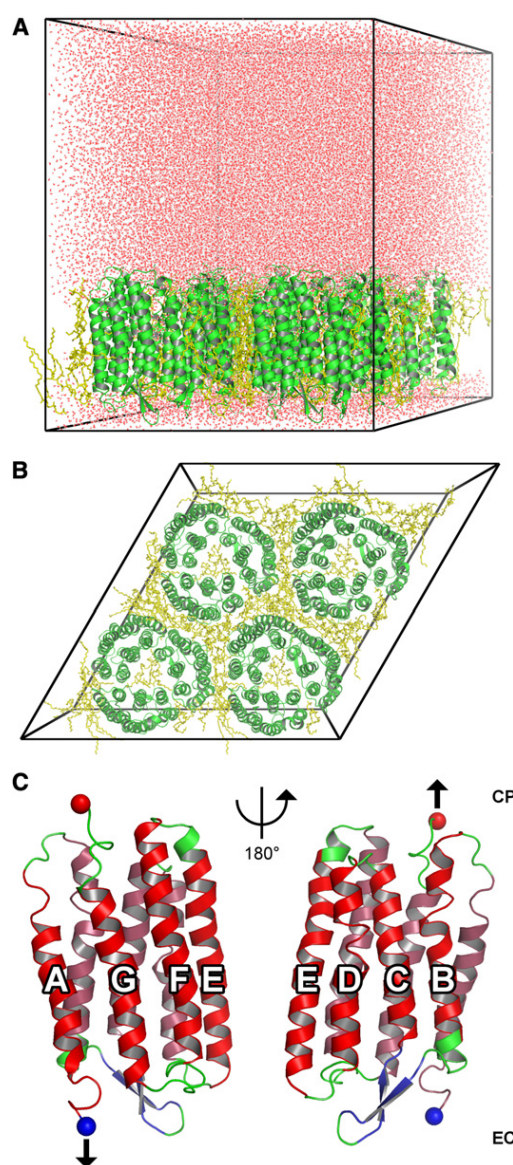


FIGURE 1 Simulation system setup. (A, *Side view*) The expanded volume at the cytoplasmic side provides sufficient space for extracted polypeptides. BR molecules are shown in green, lipid molecules in yellow, and water molecules in red. (Black lines) Simulation box. (B, *Top view*) Four unit cells, each containing a BR trimer, form a hexagonal two-dimensional crystal. Colors as before. (C) Mechanical unfolding setup. Either the C- or the N-termini (red and blue sphere, respectively) were subjected to a pulling potential as described in [Methods](#) (arrow), enforcing extraction toward the cytoplasmic and extracellular side, respectively. Colors represent the secondary structure of the crystal structure (33). Helices A–G are indicated.

(Fig. 1 C) was subjected to a harmonic pulling potential V_{pull} , which was moved with constant velocity in positive (negative, respectively) z direction away from the membrane, parallel to the membrane normal,

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

where $k = 500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ is the spring constant, $z_{C_\alpha}(t)$ the z position of the respective C_α atom, and v the extraction velocity. An offset z_{offset} was introduced to allow for nonzero initial forces, as will be described below.

The center of mass of the nonextracted proteins was kept fixed. No position restraints were applied.

Aiming at extracting and unfolding a complete BR monomer, the fully extended polypeptide chain would be too long to fit into the chosen simulation box. To circumvent this problem and to keep the system computationally tractable, we repeatedly cut off unfolded parts of the protein that had moved sufficiently far away from the membrane to render interactions, in particular electrostatic ones, with the membrane negligible. Accordingly, whenever the pulled C_α atom approached the simulation box boundary to below 1 nm, the unfolding simulation was stopped, extracted residues that reside >1 nm above the membrane in z direction were removed, and the voids caused by peptide removal were filled with water molecules. Ions were added or removed to keep the system neutral. New charged termini were built and the system was energy-minimized and equilibrated at 300 K for 20 ps with position restraints on heavy atoms of proteins and lipid molecules.

The extraction simulation was then resumed, with the new terminal C_α atom subjected to the new V_{pull} and appropriately chosen z_{offset} to match the initial force of the new simulation to the final force of the previous one. Specifically, the new force

$$F_{new}(t) = k(z_{C_\alpha, new}(t) - z_{offset, new} - vt)$$

was required to equal the force

$$F_{old}(t) = k(z_{C_\alpha, old}(t) - z_{offset, old} - vt)$$

applied before removal of the residues. Accordingly, and using Eq. 1 and with $F = -\nabla V$, the new offset position $z_{offset, new}$ was chosen as

$$z_{offset, new} = z_{offset, old} + z_{C_\alpha, new} - z_{C_\alpha, old}. \quad (2)$$

The spring constant was kept unchanged during the simulations. This procedure was iterated until complete protein extraction.

Anchor points

As observed in the AFM experiments (23), the unfolding of BR proceeds stepwise. Apparently, some residues referred to as anchor points are able to withstand a much higher force against mechanical unfolding than others. In the experiments, anchor points have been located from peaks in the force curves by wormlike chain polymer model fits (48). To facilitate direct comparison to AFM results, anchor points were also determined from our simulations as described below. In contrast to the experiments, detailed structural information is here available, which therefore has been used to determine anchor points without referring to the wormlike chain model.

We proceeded in two steps (Fig. 2). First, for each frame of an extraction simulation a transition point (Fig. 2 A, arrows) between the unfolded (red) and folded (gray) part of the protein was determined. To this end, unfolded parts of the protein were identified via their markedly increased deviation Δz of C_α positions from the x-ray structure in z direction (Fig. 2 B).

Transition points were defined in an automated manner by using the fact that Δz values of the unfolded part of the protein increase nearly linearly with residue number, in contrast to the folded part. Accordingly, transition points were determined as the start of this linear range. To that end, the quality of fits to varying ranges (Fig. 2 C, different colors) with varying start residues N_0 but the same end residue was used. (For mechanical extraction toward the cytoplasmic side, conversely, ranges with the same start residues but different end residues were used.)

Accordingly, a linear function with slope m and intercept b was fitted to the Δz_i values for all residues i in each of the considered ranges with length n (Fig. 2 C) using

$$m = \frac{n(\sum i \Delta z_i) - (\sum i)(\sum \Delta z_i)}{n(\sum i^2) - (\sum i)^2} \quad (3)$$

and

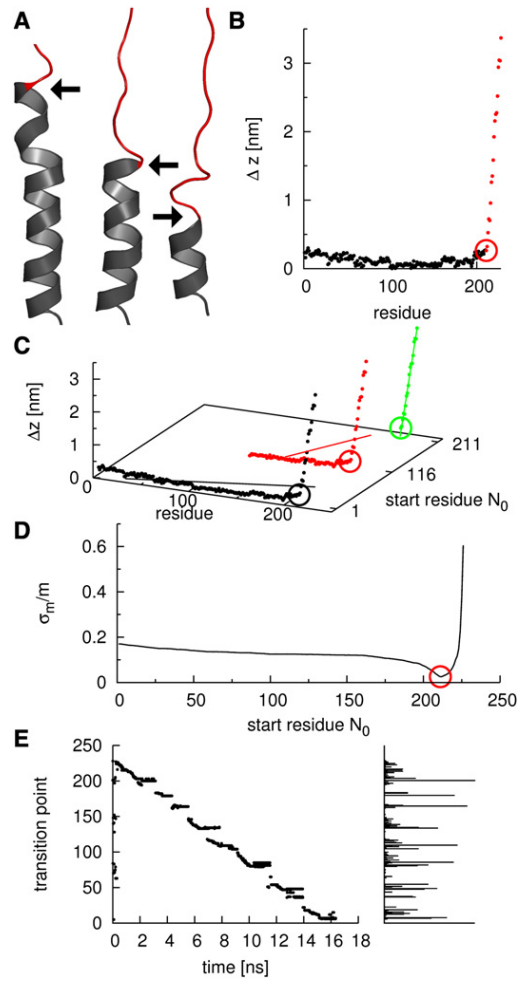


FIGURE 2 Determination of anchor points. (A) Snapshots of an unfolding helix. Black and red regions indicate folded and unfolded parts, respectively, with arrows depicting transition points. (B) Displacement Δz of residues from the crystal structure in z direction; colors as in panel A. (Circle) Transition point. (C) Transition points are determined via a line fit. Example data sets with line fits are shown. (Black) Full data set. (Red) Data set without the first 115 residues. (Green) Data set with only unfolded residues. (D) Relative error σ_m/m of the slope m of the line fits as a function of the first residue in the data sets. The minimal relative error was used to define the transition point (red circle). (E) Time development (left) and frequency (right) of transition points determined from an extraction simulation.

$$b = \frac{(\sum i^2)(\sum \Delta z_i) - (\sum i)(\sum i \Delta z_i)}{n(\sum i^2) - (\sum i)^2}. \quad (4)$$

The fit quality was quantified by the relative error σ_m/m of the slope, with

$$\sigma_m = \sqrt{\frac{n \sum (\Delta z_i - b - mi)^2}{(n-2)(n \sum i^2 - (\sum i)^2)}}. \quad (5)$$

Fig. 2 D shows the fit quality as a function of start residue number N_0 for the chosen example. Transition points were then defined as the minimum of σ_m/m (red circle).

The left part of Fig. 2 E shows transition points determined from an extraction simulation as described above. As can be seen—and as quantified by the histogram (right side)—some transition points occur markedly more often than others. This histogram served to determine the location of the anchor points for comparison with the AFM data.

Due to the stochastic nature of the unfolding process, histograms obtained from separate trajectories will differ from each other. To improve statistics, all histograms for a given extraction velocity were merged and filtered using a Gaussian filter with a one-residue half width.

RESULTS AND DISCUSSION

After equilibration, 27 extraction simulations were carried out, both toward the cytoplasmic and the extracellular side. To study to which extent the unfolding path and anchor points are velocity-dependent, unfolding simulations were carried out at five different extraction velocities ranging from 50 m/s to 1 m/s. Each simulation was terminated after complete extraction of the respective BR monomer. As also observed in the AFM experiments (23), adjacent proteins were essentially unperturbed, leaving a stable water-filled hole at the former position of the extracted protein. Except otherwise noted, we will focus at those observables that were similar for all simulations and velocities.

Force profiles

Fig. 3, A and B, shows typical force profiles obtained from simulations at different extraction velocities for extraction toward the cytoplasmic (A) and the extracellular side (B). All force profiles show four main force peaks labeled 1, 2, 3, and 4 at similar spring positions. Their overall shape and position is similar to the force peaks obtained by AFM.

Some of the main force peaks split up into subpeaks, e.g., the first and third peak of the cytoplasmic trajectories. Overall, the force pattern does not depend on the extraction velocity.

For high extraction velocities (20 m/s and 50 m/s), abrupt force jumps are occasionally observed. These drops occur when peptides are removed from the system as described in Methods and are due to transiently reduced frictional drag of the unfolded peptide after cutting.

As in the experiments, the force peaks observed in the simulations decreased from peak 1 to peak 4. We attribute this effect to the decreasing number and strength of interactions within the nonextracted protein due to the extraction of unfolded parts during the later parts of the simulations. Force minima reach 0 pN for extraction velocities below 5 m/s. This indicates that at these velocities, as also reported in Gräter et al. (18), frictional forces become small with respect to unfolding forces.

Comparison to AFM forces

Fig. 3 C compares the heights of the force peaks between simulation (triangles) and experiment (26) (circles) for different extraction velocities. In accordance to the AFM force peaks, the MD force peaks increase with faster extraction velocities. To compare the unfolding forces observed in the simulations with those measured by AFM at much slower extraction velocities, a simple model for the logarithmic velocity dependence of activated barrier crossing is used (18,49),

$$F(v) = \gamma v + a \ln(v/v_0). \quad (6)$$

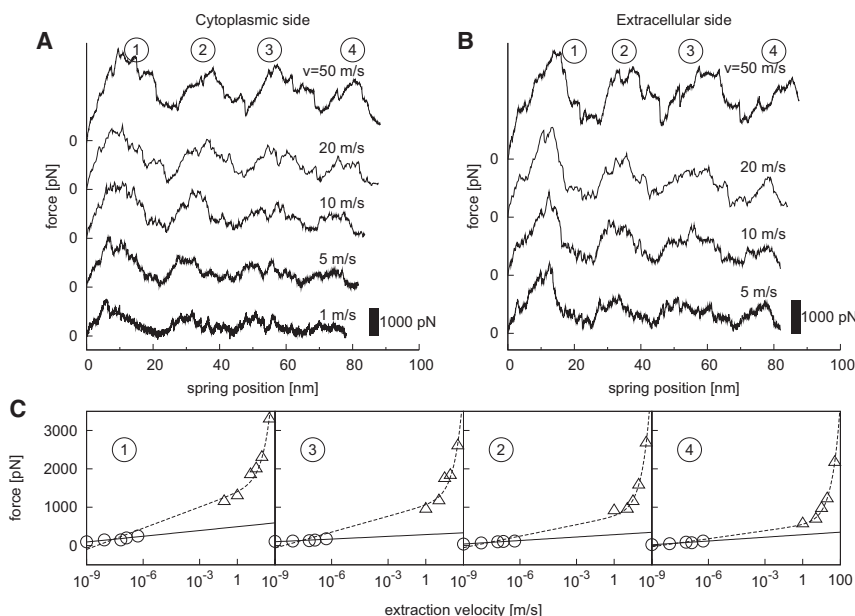


FIGURE 3 (A and B) Sample force profiles from extraction simulations for different extraction velocities toward the cytoplasmic (A) and the extracellular side (B). All force profiles start at 0 pN. (C) Comparison between peak forces from AFM experiments (circles) and extraction simulations (triangles) for extraction toward the cytoplasmic side. Two fits are shown for each peak; a logarithmic fit to only the AFM data (solid lines) and a fit to both AFM and MD data using Eq. 6 (dashed lines).

Here, γ denotes a friction coefficient and a and v_0 are fitting parameters.

Fits to both AFM and MD values (Fig. 3 C, dashed lines) yield a good description for all MD force peaks and for AFM force peak 4. For AFM force peaks 1, 2, and 3, the logarithmic slope of the AFM force peaks (solid lines) is smaller than that of the fit to both AFM and MD, indicating that for these cases the unfolding pathway between both methods may differ.

Remarkably, this discrepancy decreases, and eventually vanishes, for the later stages of the unfolding process, where fewer sterical hindrances obstruct further unfolding. This indicates that, besides the intrinsic unfolding forces of the helices, interactions to the remaining protein contribute to the overall forces. The additional velocity dependence can be explained by structural rearrangements that may occur during AFM extraction but that are too slow for MD timescales.

To examine the causes of the force peaks during extraction, the unfolding pathway of the protein during the extraction simulations will be analyzed in more detail and

compared to the unfolding pathway proposed by the AFM experiments.

Unfolding pathway

Fig. 4 shows unfolding snapshots of a BR monomer. As example, the unfolding and extraction of helices G and F toward the cytoplasmic side at an extraction velocity of 5 m/s is shown. The snapshots show a continuous and sequential unfolding and extraction of the two helices. As in the experiments, the order of the extraction on the helix scale is given by the extraction direction: For extraction toward the cytoplasmic side, helix G unfolds first and helix A last. This order is reversed for extraction toward the extracellular side.

On the level of the individual helices, a sequential unfolding of individual helix turns is seen (Fig. 4, A–F). During unfolding of part of each helix, only small structural changes are seen in its remaining folded part and the remaining helices. One exception is helix E, part of which

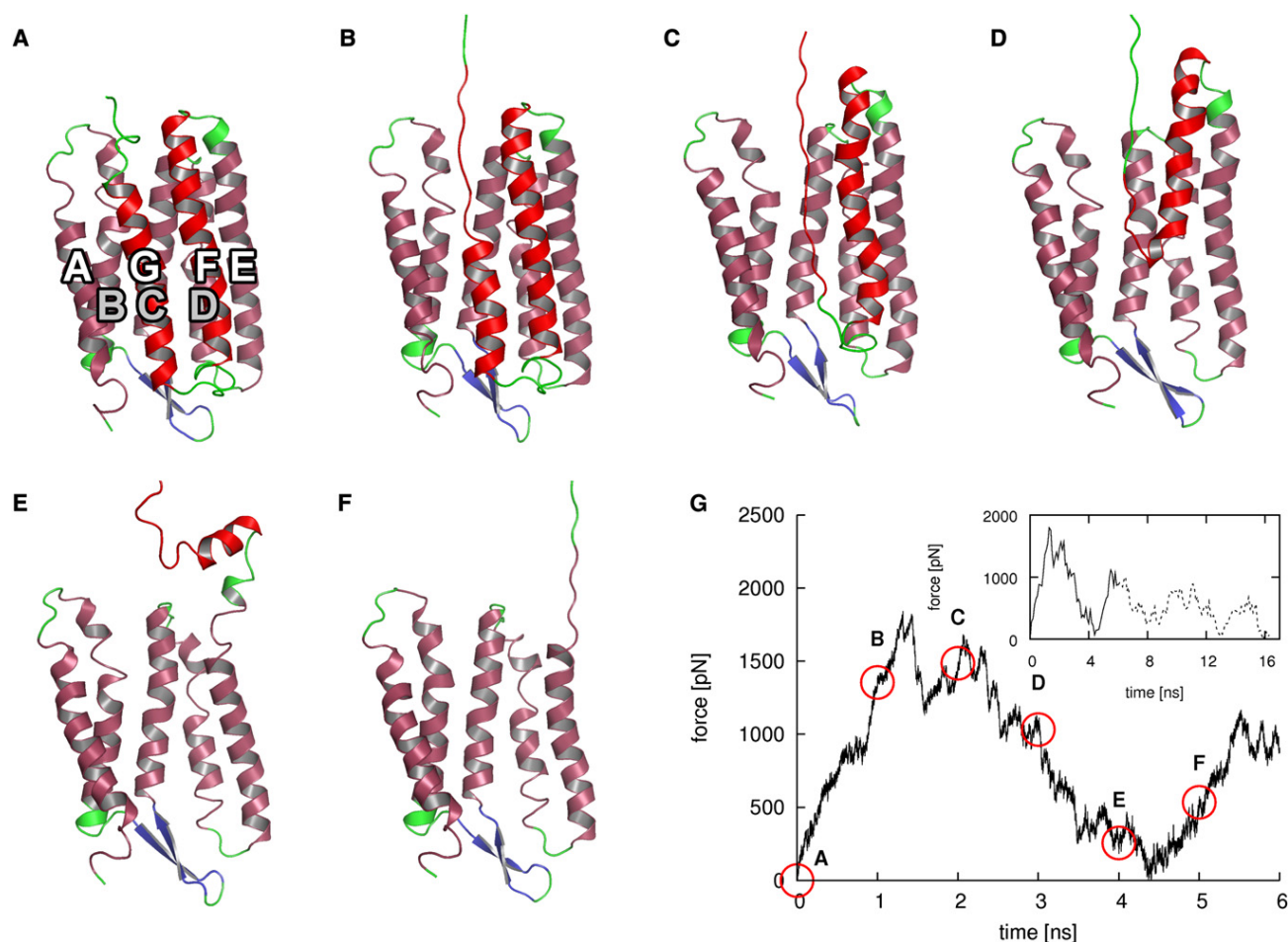


FIGURE 4 Unfolding pathway of helices G and F of BR (A–F) and corresponding forces (G). Snapshots were taken at 1 ns intervals. Colors as in Fig. 1 C. Bold letters denote transmembrane helices. Gray letters belong to helices that are only partly visible. Red circles in the force profile denote snapshot times. (Inset) Complete force profile. The solid part corresponds to the main plot.

unfolds together with helix F (Fig. 4 E). Most unfolding events occur within the membrane, with the unfolding helix nearly in place and oriented parallel to the remaining helices (Fig. 4, A–D). An exception are the last parts of helices F, D, and B, which are dragged out of the membrane partly intact, then kink with respect to their original alignment, and unfold in bulk water parallel to the membrane surface.

The rise and fall of the force during extraction correlates with the unfolding of single helices (Fig. 4 G). During the unfolding of helix G and F the force increases and decreases, respectively. Upon complete unfolding of helix G and F, the force reaches its maximum and minimum, respectively. The extraction and unfolding of helix pairs E and D as well as C and B creates a force peak in a similar manner. The last peak in the force spectrum is caused by the unfolding and extraction of helix A. Extraction toward the extracellular side proceeds in a similar manner.

Anchor points

One main result of the experiments was that at certain residues, referred to as anchor points, larger forces were required for unfolding to proceed. Apparently, these residues resist mechanical unfolding particularly strongly, which indicates strong interactions relevant for the general stability of the protein. In the experiments, anchor point locations were determined via wormlike chain fits to the force curves (48). To locate anchor points in the extraction simulations, all trajectories were subjected a systematic search protocol.

To that aim, transition points between folded and unfolded parts of the protein were determined during the course of the simulations. Because anchor points are expected to delay unfolding, transition points should occur more frequently at their locations, and should therefore show up as peaks in transition point histograms.

Indeed, transition point histograms reveal a noncontinuous unfolding in the simulations (Fig. 5, A and B, *light gray areas*). Maxima from this distribution can be clearly identified and define anchor points. Anchor points from the extraction simulations show no velocity-dependent locations and nearly all are seen for all extraction velocities at each respective extraction direction.

To better compare anchor points from AFM and MD, we divided the latter into three groups: The first group consists of MD anchor points that cover AFM anchor points within their respective uncertainties. A second group is defined as MD anchor points occurring close to AFM anchor points. The first two groups cover ~40% of all anchor points for both extraction directions, showing a good agreement between both methods. For the last group, no match between AFM and MD data is seen.

In the first group, almost all AFM anchor points are also seen in the simulations. In the case of extraction toward the cytoplasmic side, AFM anchor points Gly⁶, Arg⁸², Val¹⁰¹, Lys¹²⁹, and Phe¹³⁵ are well matched by MD anchor points. For extrac-

tion toward the extracellular side, Glu⁷⁴, Leu¹⁰⁰, and Asp¹⁰² show a good agreement. Given their occurrence in both AFM and MD, these anchor points seem to be independent of the extraction velocity, and the interactions revealed by the simulations are likely to also dominate the AFM forces.

For the second group, a noticeable deviation of up to three residues between anchor points from MD and from AFM is seen. Notable examples are AFM anchor points Ile¹⁹⁸ and Val¹⁸⁸ for extraction toward the cytoplasmic side and Arg¹⁶⁴ and Leu²²³ for extraction toward the extracellular side. These deviations may arise either from uncertainties in the experimental determination of the anchor points via the wormlike chain model (28), or, alternatively from a possible velocity dependence of anchor point locations. Like the first group, this group agrees remarkably well with the AFM experiments.

The third group, where no match between AFM and MD is observed, falls into two categories. The first, less frequent category contains anchor points from AFM that are not observed by MD. Only three anchor points fall into this category, Tyr⁸³, Pro⁹¹, and Leu¹¹¹ from the extraction toward the extracellular side. The fact that these anchor points are only seen by AFM indicates a velocity dependence of the underlying reaction paths. The second, more frequent category includes all MD anchor points that are not seen by AFM. Most notably, MD anchor points are observed in helices B, D, F, and G during the extraction toward the cytoplasmic side and in helices A, B, F, and G and the loop region connecting helices D and E during the extraction toward the extracellular side.

We assume that these anchor points are not seen by AFM because they occur during force drops, whereas rising forces are needed to detect anchor points in the analysis of the experiments (48). Because the transition point method used for the simulations does not involve forces, anchor points can also be detected during decreasing forces. This idea is also confirmed by the strongly decreasing forces that accompany the unfolding of helices B, D, and F in both extraction directions (see Fig. 3) and explains almost all MD anchor points not seen by AFM.

Remarkably, anchor points occur more frequently in the extraction simulations toward the cytoplasmic side than in the opposite direction. We attribute this finding to interactions between the retinal and the helices in contact with it that stabilize the protein core. As the retinal is covalently bound to Lys²¹⁶ located in helix G, it is removed upon extraction of helix G. Because in the extraction simulations toward the cytoplasmic side, helix G is removed from the protein core already very early in each simulation, the stabilization is lost, and a larger number of weaker interactions dominate. In contrast, for extraction toward the extracellular side, the retinal remains within the protein for nearly the complete process, and, due to the interactions with the retinal, more pronounced but possibly fewer anchor points are expected. This is seen in Fig. 5, A and B. In this scenario, one would also expect more details and anchor points in the

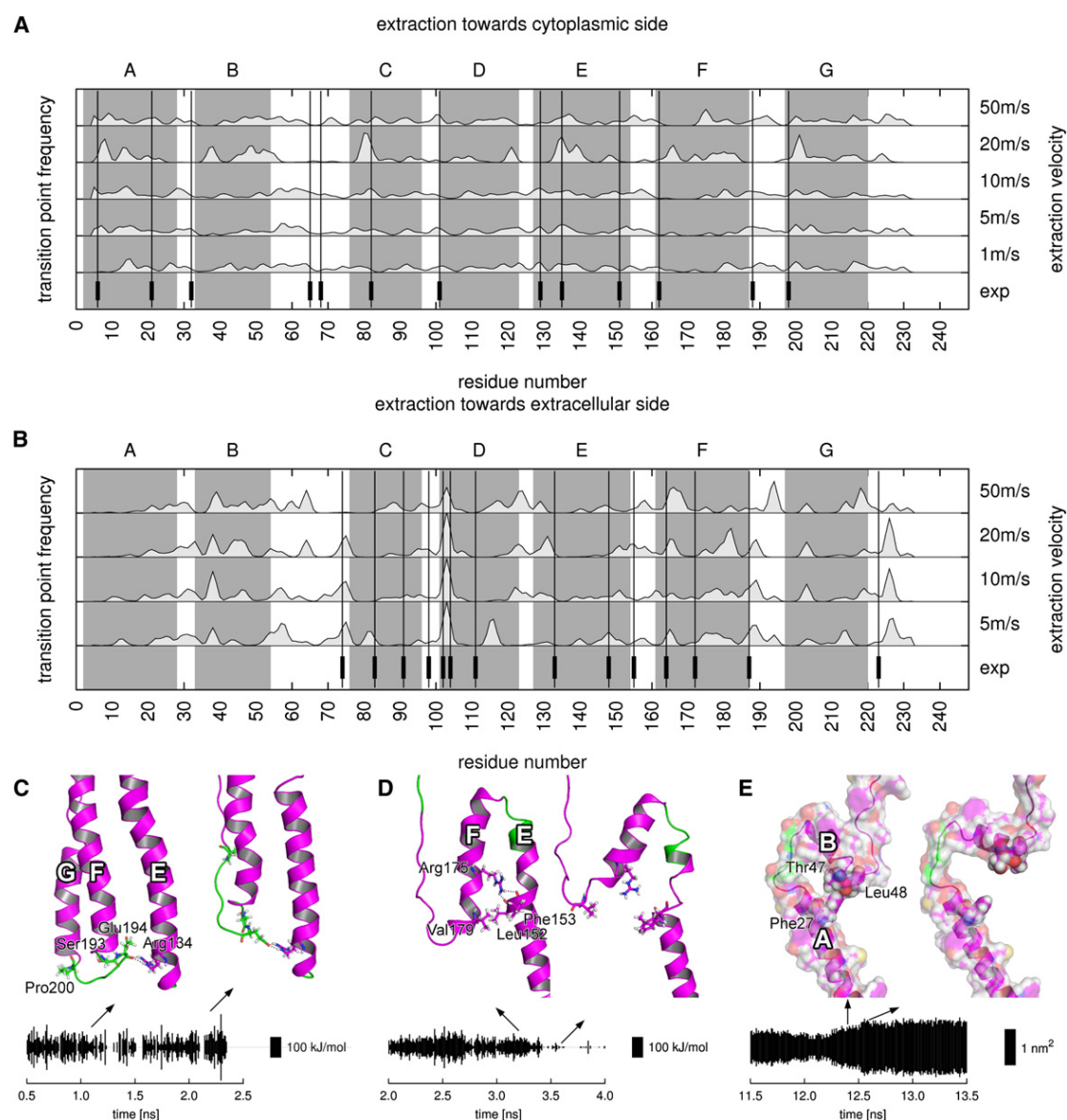


FIGURE 5 (A and B) Comparison between AFM anchor points (*vertical bars and lines*) (28) and frequency of transition points determined by MD (*light gray areas*). Transition point histograms are filtered with a Gaussian filter with a one residue half width. (*Dark gray areas*) Helices. (C–E) Structural and energetic determinants of anchor points. Each panel shows snapshots before and after rupture of an anchor point. The plots show hydrogen bond energies (C and D) or hydrophobic solvent-accessible surface areas (E) of selected residues. Bold letters denote helices. Key residues are shown as sticks (C and D) or spheres (E). (C) Anchor point Pro²⁰⁰ in a stable conformation and shortly before rupture. The plot shows the summed energy of hydrogen bonds between Arg¹³⁴ and Ser¹⁹³ and Glu¹⁹⁴. (D) Anchor point Val¹⁷⁹ before and after rupture. The plot shows the summed energy of hydrogen bonds belonging to Arg¹⁷⁵. (E) Anchor point Leu⁴⁸ before and after rupture. The plot shows the hydrophobic solvent accessible surface area of Phe²⁷.

force profile for bacterioopsin extraction than seen for BR, which indeed has been observed (24).

Energetics causing anchor points

Which are the interactions that stabilize intermediate structures and thus cause the observed anchor points? To address this question, we analyzed the sequence of structural changes and the rupture of intramolecular interactions leading to the occurrence of these points in more detail.

Analysis of interactions between each residue and the remaining protein revealed that the residues lose their interactions in sequential order (see Fig. S1 in the Supporting Material). This behavior even applies to hot-spot residues with higher interaction energies. Fluctuations of residuewise energies suggest a rearrangement of interactions within the protein.

Based on their location within the helices and on their stabilizing energetics, we divided anchor points into three groups. The first group consists of anchor points that are

located at the ends of the transmembrane helices and are stabilized by interhelical hydrogen bonds. Anchor points in the second group are also stabilized by interhelical hydrogen bonds but reside in the center of the helices. The third group consists of anchor points in the center of helices that are stabilized by hydrophobic contacts.

Fig. 5 C depicts Pro²⁰⁰ as a typical example of anchor points from the first group. Upon the unfolding of helix G, the extracellular end of the adjacent helix F is held back by hydrogen bonds between backbone atoms from Ser¹⁹³ and Glu¹⁹⁴ from helix F to side-chain atoms of Arg¹³⁴ from helix E (*first snapshot*). Before complete unfolding of helix G, a rearrangement of interactions is seen: New hydrogen bonds between side-chain atoms from Arg¹³⁴ to side-chain atoms from Ser¹⁹³ and Glu¹⁹⁴ are formed (*second snapshot*). The nearly constant energy of hydrogen bonds belonging to Arg¹³⁴ (*lower plot*) shows that this rearrangement preserves the strength of interactions that hold back helix F. To this end, a dynamic network of hydrogen bonds connects terminal regions of adjacent transmembrane helices.

Anchor points from the second group are also stabilized by hydrogen bonds but are located in the center of the helices, like Val¹⁷⁹ located in helix F (Fig. 5 D). Upon extraction, hydrogen bonds between side-chain atoms of the nearby Arg¹⁷⁵ to backbone atoms from residues from helix E maintain the orientation of helix F (*first snapshot*). Once these interactions break up, the remaining part of helix F kinks with respect its former orientation (*second snapshot*).

In contrast, anchor points belonging to the third group, like Leu⁴⁸ located in helix B (Fig. 5 E), are stabilized by nonpolar interactions rather than hydrogen bonds. Phe²⁷, Thr⁴⁷, and Leu⁴⁸ form a hydrophobic core between helices A and B that disintegrates during unfolding of helix B, indicated by the increasing hydrophobic surface area of Phe²⁷ (*lower graph*). Once the hydrophobic core is completely disrupted (maximal hydrophobic surface), the remaining part of helix B tilts away from its former alignment (*second snapshot*).

A table of anchor points observed in extraction simulations toward the cytoplasmic side is available in Table S1.

Our analysis of mechanically relevant interactions in BR thus revealed as the main determinants for the observed anchor points a network of hydrogen bonds and hydrophobic contacts. Further, a highly dynamic rearrangement of these interactions during extraction is seen. The competition of quite different kinetics suggests that the observed unfolding pathway depends on the timescale set by the extraction velocity. Thus, the question arises, how slower extraction velocities and correspondingly longer timescales of extraction would influence the unfolding process.

Relaxation processes

Indeed, the different kinetics of interactions described in the previous section suggests a velocity-dependent unfolding

pathway. This was also proposed by fits to the peak forces obtained by AFM and MD (see Comparison to AFM Forces). Thus, for sufficiently small extraction velocities, we would expect to observe also in the simulations a simultaneous unfolding of helix pairs rather than the sequential unfolding of single helices seen so far.

To test this hypothesis, we determined how the unfolding pathway of partially unfolded helices proceeds once the extraction velocity is changed from fast MD values (~ 10 m/s) to slow AFM values ($\sim 10^{-9}$ m/s). Because the difference in velocities is very large, extraction at AFM velocities is quasistatic within MD timescales. To this end, we took snapshots from an extraction simulation with $v = 20$ m/s as start for simulations with a resting spring, $v = 0$ m/s. To study an extended unfolding pathway, we chose snapshots where helices G or F were partly unfolded.

In both simulations, forces began to decrease instantaneously (Fig. 6 A). The force development can be described by a biexponential decay

$$F(t) = F_{\text{final}} + F_1 \cdot e^{-t/\tau_{\text{fast}}} + F_2 \cdot e^{-t/\tau_{\text{slow}}} \quad (7)$$

with slow and fast relaxation times τ_{slow} and τ_{fast} , respectively. F_{final} , F_1 , and F_2 are fit constants (values are given in Table S2). The fast relaxation times ((730 ± 5) ps and (95 ± 2) ps for the simulation with a partially unfolded helix G and F, respectively) can be attributed to the loss of friction built up during the preceding extraction simulation. By assuming that the attachment point for the harmonic potential moves ~ 1 nm during the fast relaxation, a minimal velocity of vanishing relaxation can be determined using $v < 1 \text{ nm}/\tau_{\text{fast}}$. For a time constant of $\tau_{\text{fast}} \sim 1$ ns, one obtains $v < 1$ m/s. This is in agreement with the previous observation that frictional forces become negligible below extraction velocities of 5 m/s (see Force Profiles).

The slow relaxation seen in the later part of the simulations revealed much longer relaxation times ($\tau_{\text{slow}} = (55,000 \pm 6000)$ ps and $\tau_{\text{slow}} = (4810 \pm 60)$ ps for the first and second simulation, respectively). Because in the slowest extraction simulation ($v < 1$ m/s) the unfolding of the protein takes ~ 80 ns, relaxation times of ~ 50 ns can only be covered by much slower extraction velocities not accessible to our simulations. However, it should be possible to observe structural relaxation in the 30-ns relaxation simulations.

In the first relaxation simulation, a shift of both helices G and F in extraction direction was observed (Fig. 6 B). Helix G did not further unfold but remained parallel to helix F during further extraction, which was not observed in the extraction simulations. This clearly indicates mutual extraction and unfolding of helices G and F at long timescales. We conclude that for very small extraction velocities, the unfolding pattern observed by AFM already shows up in the simulations.

The partially unfolded helix F (Fig. 6 C) showed further unfolding during the slow relaxation phase at small forces

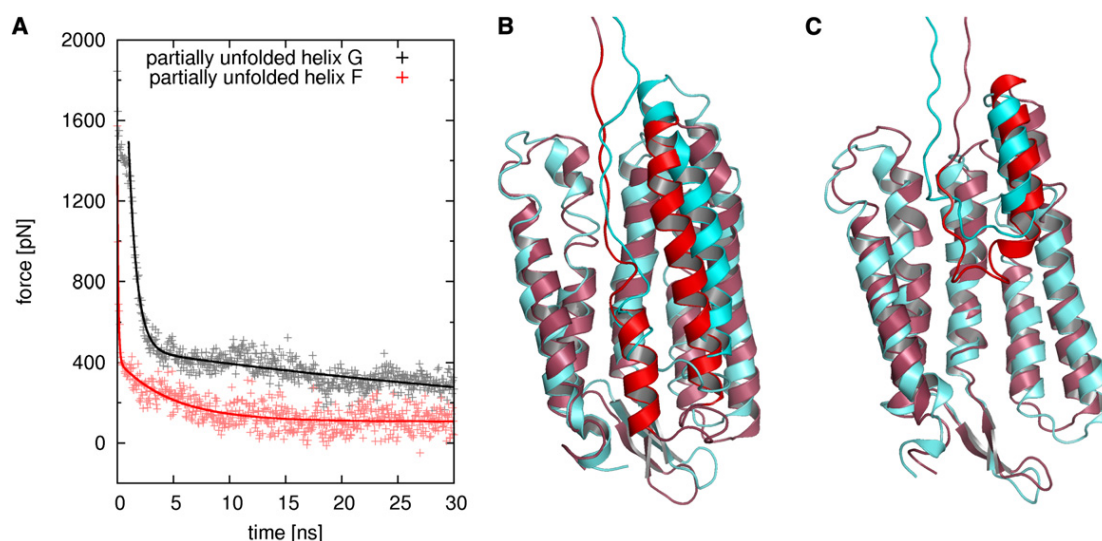


FIGURE 6 Relaxation simulations with partially unfolded helices G (gray) and F (red), respectively. (A) Forces (symbols) and biexponential fits (lines). The force fit for the first simulation starts after the retinal was extracted from the protein core. (B) Snapshots before (red) and after (cyan) the relaxation simulation with a partially unfolded helix G. (C) Similar, but for a partially unfolded helix F (same colors).

(~100 pN) (Fig. 6 C). Because peak forces in the experiments are in a similar range and helix F unfolds after a peak, a spontaneous unfolding of helix F in the experiments can be assumed. Fast unfolding is further corroborated by the short relaxation time of ~5 ns.

The results from both relaxation simulations show that the unfolding pathway of BR under external force changes with extraction velocity and is therefore timescale-dependent. Based on these observations, the discrepancies in the proposed unfolding pathways can be explained. At slow extraction velocities in AFM experiments ($<10^{-7}$ m/s), the extraction and unfolding of pairs of helices is visible, whereas at high extraction velocities in MD simulation (>1 m/s), single helix turns unfold. This is also in line with the observation of the unfolding of single helices at higher AFM extraction velocities ($>10^{-7}$ m/s) (26).

SUMMARY AND CONCLUSION

We presented all-atom FPMD simulations in which we extracted and unfolded single BR monomers from an atomistic purple membrane model. The extraction was carried out at various velocities both toward the cytoplasmic and the extracellular side of the membrane. To maintain a computationally tractable box size, a new protocol was introduced and successfully tested, in which the extracted and unfolded part of the protein was removed regularly and the extraction continued at the cleaved part of the protein, with the respective forces adjusted appropriately.

Force profiles obtained from the simulations were similar to those obtained by AFM in many respects. In both methods, four distinctive peaks with decreasing values were observed, pointing to a sequential unfolding of the protein. For extrac-

tion toward the cytoplasmic side, extraction and unfolding of each helix pair GF, DE, and BC, as well as of helix A, caused a force peak. In contrast to the experiments, which suggested a concerted unfolding of helix pairs, in the simulations we observed a sequential unfolding of individual helix turns, although—as in the experiments—only every second helix actually gave rise to a pronounced force peak. This observation suggests a second possible interpretation of the AFM force profiles, according to which the experiments would also be compatible with sequential unfolding of individual helices. In any case, details of the sequence of events during unfolding may depend on the loading rate.

Despite these possible differences, similar anchor points were seen by MD, suggesting that the underlying molecular interactions are largely independent of extraction velocity. MD anchor points fall into three groups. The first group comprises anchor points that fully agree with those seen by AFM, the second group consists of anchor points close to AFM anchor points, and the third group includes anchor points not seen by AFM. Most anchor points of the last group occur after the main force peaks, during force drops, where they are unlikely to be seen in the experiments, which can explain this discrepancy. Overall, remarkably good agreement is seen, with the position of most anchor points being largely unaffected by the quite different timescales of AFM and MD.

As the dominant molecular interactions that stabilize the anchor points and, hence, govern the mechanical stability of BR, we identified hydrogen bonds and hydrophobic contacts. During extraction, these interactions form a highly dynamic network, where transient interactions are established subsequent to the rupture of old ones. Thus, the unfolding of BR does not follow a predefined path; rather, the competition between the kinetics of the

interaction network and the unfolding timescale can give rise to different velocity dependent unfolding pathways.

To explore this possibility further, relaxation simulations were carried out to detect slow relaxation motions that are not visible in the fast extraction simulations. To this end, extraction simulations with a partially unfolded protein were taken and continued without further extraction. Indeed, an onset of spontaneous collective unfolding of both helices G and F was seen, which deviated from the fast unfolding pathway, toward the one suggested by AFM carried out at much longer timescales.

The good agreement between the results from the experiments and the simulations suggest that our simulations capture the essential characteristics of BR extraction. Because the simulations were performed with a POPC membrane model for the purple membrane, one might expect slightly changed interactions between the protein and the bilayer. However, as our model fully includes the hexagonal symmetry of the purple membrane and therefore also all intermolecular contacts, our results support the role of this specific arrangement of BR within the bilayer.

It will be interesting to see whether dynamic interaction networks of the type observed in this work also govern the timescale-dependent mechanical properties of other GPCRs or even members of other membrane protein families. Further, this approach may also reveal atomistic details of the origin of recently observed mechanical fingerprints of ligand binding (50,51).

SUPPORTING MATERIAL

One figure and two tables are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(11\)00046-4](http://www.biophysj.org/biophysj/supplemental/S0006-3495(11)00046-4).

The authors thank Daniel Müller, Hermann Gaub, and Max Kessler for helpful discussions, and Ulrike Gerischer and Ulrich Zachariae for carefully reading the manuscript.

REFERENCES

- Krogh, A., B. Larsson, ..., E. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305:567–580.
- Pierce, K. L., R. T. Premont, and R. J. Lefkowitz. 2002. Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.* 3:639–650.
- Stoeckenius, W., and R. A. Bogomolni. 1982. Bacteriorhodopsin and related pigments of halobacteria. *Annu. Rev. Biochem.* 51:587–616.
- Haupts, U., J. Tittor, and D. Oesterhelt. 1999. Closing in on bacteriorhodopsin: progress in understanding the molecule. *Annu. Rev. Biophys. Biomol. Struct.* 28:367–399.
- Binnig, G., C. F. Quate, and C. Gerber. 1986. Atomic force microscope. *Phys. Rev. Lett.* 56:930–933.
- Rief, M., F. Oesterhelt, ..., H. E. Gaub. 1997. Single molecule force spectroscopy on polysaccharides by atomic force microscopy. *Science*. 275:1295–1297.
- Lee, G. U., L. A. Chrisey, and R. J. Colton. 1994. Direct measurement of the forces between complementary strands of DNA. *Science*. 266:771–773.
- Marszalek, P. E., A. F. Oberhauser, ..., J. M. Fernandez. 2003. The force-driven conformations of heparin studied with single molecule force microscopy. *Biophys. J.* 85:2696–2704.
- Florin, E. L., V. T. Moy, and H. E. Gaub. 1994. Adhesion forces between individual ligand-receptor pairs. *Science*. 264:415–417.
- Rief, M., M. Gautel, ..., H. E. Gaub. 1997. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*. 276:1109–1112.
- Li, L., H. H.-L. Huang, ..., J. M. Fernandez. 2005. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J. Mol. Biol.* 345:817–826.
- Kedrov, A., C. Ziegler, ..., D. J. Müller. 2004. Controlled unfolding and refolding of a single sodium-proton antiporter using atomic force microscopy. *J. Mol. Biol.* 340:1143–1152.
- Möller, C., D. Fotiadis, ..., D. J. Müller. 2003. Determining molecular forces that stabilize human aquaporin-1. *J. Struct. Biol.* 142:369–378.
- Grubmüller, H., B. Heymann, and P. Tavan. 1996. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*. 271:997–999.
- Izrailev, S., S. Stepaniants, ..., K. Schulten. 1997. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* 72:1568–1581.
- Lu, H., B. Isralewitz, ..., K. Schulten. 1998. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* 75:662–671.
- Gräter, F., and H. Grubmüller. 2007. Fluctuations of primary ubiquitin folding intermediates in a force clamp. *J. Struct. Biol.* 157:557–569.
- Gräter, F., J. Shen, ..., H. Grubmüller. 2005. Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. *Biophys. J.* 88:790–804.
- Sotomayor, M., D. P. Corey, and K. Schulten. 2005. In search of the hair-cell gating spring elastic properties of ankyrin and cadherin repeats. *Structure*. 13:669–682.
- Cieplak, M., S. Filipek, ..., K. A. Krzyżko. 2006. Pulling single bacteriorhodopsin out of a membrane: comparison of simulation and experiment. *Biochim. Biophys. Acta*. 1758:537–544.
- Seeber, M., F. Fanelli, ..., A. Caffisch. 2006. Sequential unfolding of individual helices of bacteriorhodopsin observed in molecular dynamics simulations of extraction from the purple membrane. *Biophys. J.* 91:3276–3284.
- Fanelli, F., and M. Seeber. 2010. Structural insights into retinitis pigmentosa from unfolding simulations of rhodopsin mutants. *FASEB J.* 24:3196–3209.
- Oesterhelt, F., D. Oesterhelt, ..., D. J. Müller. 2000. Unfolding pathways of individual bacteriorhodopsins. *Science*. 288:143–146.
- Müller, D. J., M. Kessler, ..., H. Gaub. 2002. Stability of bacteriorhodopsin α -helices and loops analyzed by single-molecule force spectroscopy. *Biophys. J.* 83:3578–3588.
- Janovjak, H., M. Kessler, ..., D. J. Müller. 2003. Unfolding pathway of native bacteriorhodopsin depend on temperature. *EMBO J.* 22:5220–5229.
- Janovjak, H., J. Struckmeier, ..., D. J. Müller. 2004. Probing the energy landscape of the membrane protein bacteriorhodopsin. *Structure*. 12:871–879.
- Cisneros, D. A., D. Oesterhelt, and D. J. Müller. 2005. Probing origins of molecular interactions stabilizing the membrane proteins halorhodopsin and bacteriorhodopsin. *Structure*. 13:235–242.
- Kessler, M., and H. E. Gaub. 2006. Unfolding barriers in bacteriorhodopsin probed from the cytoplasmic and the extracellular side by AFM. *Structure*. 14:521–527.
- Kessler, M., K. E. Gottschalk, ..., H. E. Gaub. 2006. Bacteriorhodopsin folds into the membrane against an external force. *J. Mol. Biol.* 357:644–654.

30. Sapra, K. T., J. Doehner, ..., D. J. Muller. 2008. Role of extracellular glutamic acids in the stability and energy landscape of bacteriorhodopsin. *Biophys. J.* 95:3407–3418.
31. Sapra, K. T., G. P. Balasubramanian, ..., D. J. Muller. 2008. Point mutations in membrane proteins reshape energy landscape and populate different unfolding pathways. *J. Mol. Biol.* 376:1076–1090.
32. Sapra, K. T., P. S.-H. Park, ..., D. J. Muller. 2008. Mechanical properties of bovine rhodopsin and bacteriorhodopsin: possible roles in folding and function. *Langmuir* 24:1330–1337.
33. Belrhali, H., P. Nollert, ..., E. Pebay-Peyroula. 1999. Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution. *Structure* 7:909–917.
34. Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* 8:52–56, 29.
35. Tieleman, D. P., and H. J. C. Berendsen. 1998. A molecular dynamics study of the pores formed by *E. coli* OmpF porin in a fully hydrated palmitoyl-oleoylphosphatidylethanolamine bilayer. *Biophys. J.* 74:2786–2801.
36. Faraldo-Gómez, J. D., G. R. Smith, and M. S. P. Sansom. 2002. Setting up and optimization of membrane protein simulations. *Eur. Biophys. J.* 31:217–227.
37. Sanner, M. F., A. J. Olson, and J.-C. Spehner. 1996. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 38:305–320.
38. Curran, A. R., R. H. Templer, and P. J. Booth. 1999. Modulation of folding and assembly of the membrane protein bacteriorhodopsin by intermolecular forces within the lipid bilayer. *Biochemistry* 38:9328–9336.
39. van der Spoel, D., E. Lindahl, ..., H. J. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.
40. Jorgensen, W. L., and J. Tirado-Rives. 1988. The OPLS potential functions for proteins—energy minimizations for crystals of cyclic-peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.
41. Jorgensen, W. L., J. Chandrasekhar, ..., M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Comput. Phys.* 79:926–935.
42. Berger, O., O. Edholm, and F. Jähnig. 1997. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* 72:2002–2013.
43. Kandt, C., J. Schlitter, and K. Gerwert. 2004. Dynamics of water molecules in the bacteriorhodopsin trimer in explicit lipid/water environment. *Biophys. J.* 86:705–717.
44. Berendsen, H. J. C., J. P. M. Postma, ..., J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
45. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
46. Essmann, U., L. Perera, ..., L. G. Pedersen. 1995. A smooth particle mesh Ewald potential. *J. Chem. Phys.* 103:8577–8592.
47. Hess, B., H. Bekker, ..., J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.
48. Bustamante, C., J. F. Marko, ..., S. Smith. 1994. Entropic elasticity of λ -phage DNA. *Science* 265:1599–1600.
49. Heymann, B., and H. Grubmüller. 1999. AN02/DNP-hapten unbinding forces studied by molecular dynamics atomic force microscopy simulations. *Chem. Phys. Lett.* 303:1–9.
50. Kedrov, A., C. Ziegler, and D. J. Müller. 2006. Differentiating ligand and inhibitor interactions of a single antiporter. *J. Mol. Biol.* 362:925–932.
51. Cisneros, D. A., L. Oberbarnscheidt, ..., D. J. Muller. 2008. Transducer binding establishes localized interactions to tune sensory rhodopsin II. *Structure* 16:1206–1213.

B. An Unusual Hydrophobic Core Confers Extreme Flexibility to HEAT Repeat Proteins

An Unusual Hydrophobic Core Confers Extreme Flexibility to HEAT Repeat Proteins

Christian Kappel, Ulrich Zachariae, Nicole Dölker, and Helmut Grubmüller*

Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

ABSTRACT Alpha-solenoid proteins are suggested to constitute highly flexible macromolecules, whose structural variability and large surface area is instrumental in many important protein-protein binding processes. By equilibrium and nonequilibrium molecular dynamics simulations, we show that importin- β , an archetypical α -solenoid, displays unprecedentedly large and fully reversible elasticity. Our stretching molecular dynamics simulations reveal full elasticity over up to twofold end-to-end extensions compared to its bound state. Despite the absence of any long-range intramolecular contacts, the protein can return to its equilibrium structure to within 3 Å backbone RMSD after the release of mechanical stress. We find that this extreme degree of flexibility is based on an unusually flexible hydrophobic core that differs substantially from that of structurally similar but more rigid globular proteins. In that respect, the core of importin- β resembles molten globules. The elastic behavior is dominated by nonpolar interactions between HEAT repeats, combined with conformational entropic effects. Our results suggest that α -solenoid structures such as importin- β may bridge the molecular gap between completely structured and intrinsically disordered proteins.

INTRODUCTION

Solenoid proteins, consisting of repeating arrays of simple basic structural motifs, account for >5% of the genome of multicellular organisms (1). Their fold often forms extended superhelical structures, which expose large surface areas to solvent. Because of this feature, solenoid proteins play central roles in the shaping of cellular scaffolds and many other cellular processes that require multiple protein-protein interactions. Examples include pathways of endocytosis, receptor-mediated nucleocytoplasmic transport, protein sorting, and the scaffold structure of the nuclear pore complex (2–5).

In structural studies, it was suggested that α -solenoids form especially flexible structures, which may fall between typical globular structured and intrinsically disordered proteins (6,7). Their tertiary structure usually displays high flexibility, whereas their secondary structure elements are thought to remain stable and well defined (6–8). This feature actually also bears close resemblance to the so-called molten globule state of proteins—a third protein structural state lying between the fully folded and completely unfolded states, in which stable secondary structure elements are believed to be arranged into a flexible, loosely packed tertiary fold (9–12). Despite their enormous biological importance, the principles governing flexibility and structural integrity of solenoid proteins have received limited attention so far.

Recently, single molecule spectroscopy methods, such as atomic force microscopy and optical or magnetic tweezers, have become useful tools to study the response of macromolecules to external forces (13). Whereas most studies

on repeat proteins focus on the folding and unfolding mechanism (14–16), an atomic force microscopy study on different ankyrin repeat constructs showed that ankyrin repeats behave as Hookean springs with spring constants of $1.5\text{--}23 \times 10^{-3}$ N/m (17). This finding suggests that such springlike features may be of general importance for the function of α -helical repeat proteins.

Importin- β is an archetypical α -solenoid and one of the best-studied repeat proteins. It is the main player in nucleocytoplasmic transport, mediating the principal pathway of protein import into the cell nucleus (3). Its extended superhelical structure, built from 19 so-called HEAT repeats, a structural motif composed of two α -helices linked by a short loop (18), is instrumental in binding and wrapping around cargo proteins destined to the nucleus. Simultaneously, importin- β interacts with nuclear pore complexes (19). To perform its function, importin- β must be able to structurally adapt to cargo proteins of various size and shape as well as to its effector protein RanGTP (20–22). Intriguingly, it has been suggested that importin- β complexes are capable of storing energy through a huge springlike deformation of this superhelix that is then released, upon dissociation, to drive disassembly and support high transport rates (20,22).

To test this hypothesis, we have performed extended equilibrium and nonequilibrium molecular dynamics (MD) simulations of the yeast importin- β solenoid (Kap95p, yImp β). Our stretching simulations show that yImp β exhibits an extraordinary degree of elasticity. Most unexpectedly, extensions by nearly a factor of two are found to be fully reversible on very short timescales. Our studies enabled us to determine the molecular spring constant of yImp β and thus to calculate the amount of energy stored in molecular complexes of yImp β . They also reveal the

Submitted May 10, 2010, and accepted for publication June 7, 2010.

*Correspondence: hgrubmu@gwdg.de

Editor: Axel T. Brunger.

© 2010 by the Biophysical Society
0006-3495/10/09/1596/8 \$2.00

doi: 10.1016/j.bpj.2010.06.032

structural determinants underlying the unique level of elasticity in yImp β , and highlight a marked difference to the usual construction principle of rigid proteins.

METHODS

All simulations were based on an open unbound conformation of the importin- β homolog Kap95p (yImp β). The relaxed structure with an elongation of ~ 13.5 nm was simulated from PDB file 2BKU (22). The structure was obtained by free MD simulations of yImp β without RanGTP as described in Zachariae and Grubmüller (8).

All subsequent MD simulations were performed using GROMACS 3.3 and 4.0 (23,24). The OPLS all-atom force field (25) and the TIP4P water model (26) were used. In all simulations, sodium and chloride ions corresponding to a salt concentration of 150 mM were added. All bond lengths were restrained using the LINCS algorithm (27). The temperature was kept constant at 300 K using the Berendsen thermostat (28) with a coupling time of 0.1 ps. The Berendsen barostat (28) was used with a coupling time of 1 ps and an isotropic compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$ to maintain a constant pressure of 1 bar. Lennard-Jones interactions were cut off beyond 1 nm. Electrostatic interactions were calculated explicitly below 1 nm and via particle-mesh Ewald summation (29,30) beyond that distance. An integration time step of 2 fs was used. All simulations performed are summarized in Table S1 in the Supporting Material.

For unbiased simulations of yImp β , the protein was placed in a dodecahedral box with box vectors of ~ 18 -nm length. Together with added water molecules and ions, these systems consisted of $\sim 550,000$ particles. Preceding the production runs, 1000 steps of steepest-descent energy minimization, with a subsequent 1-ns simulation, and position restraints of 1.66 N/m on heavy protein atoms, were performed.

For stretching simulations, the protein was oriented along the box vectors, so that the main dimension of the protein pointed in z direction. The length of the box vectors was chosen to ensure that the minimal distance from protein atoms to the box boundaries does not fall below 1 nm. Subsequently, the box was enlarged by 15 nm along z , resulting in a box of $10 \times 10 \times 30$ nm. Water and ions were added, resulting in $\sim 390,000$ particles. One-thousand steps of steepest-descent energy minimization with a subsequent 1-ns simulation and position restraints of 1.66 N/m on heavy protein atoms were performed. All stretching simulations are based on the coordinates of this system. To stretch the protein, the C_α -atom of the N-terminus was fixed by position restraints with a force constant of 1.66 N/m, while a pulling potential with a force constant of 0.83 N/m was moved with constant velocity in z direction (see Fig. 1) (31). The pulling potential was acting exclusively in z direction. The simulations were stopped after 15 nm of spring movement. To determine the spring constant, a second set of simulations was performed with the setup as described above, but which were stopped after 5 nm of spring movement.

Relaxation simulations formed a third group of simulations. Here, structures from stretching simulations were taken and simulated freely, i.e., without the stretching potential. Two approaches were used to generate the simulation systems. In the first case (elongations of 15.3, 17.1, and 19.1 nm), a snapshot of only the protein was used. Subsequently, a new box with a minimal distance to the protein of 1 nm was built and filled with water and ions. Before production runs, energy minimization and equilibration were performed as described before. In the second case (elongation of 21.0 nm), a snapshot of the whole simulation system including solvent atoms was taken and used as input for further unbiased simulations. In this instance, further energy minimizations or equilibration simulations were not needed.

To estimate entropy changes upon extension, in a fourth group of simulations, snapshots from elongated conformations of yImp β at 13.4, 14.6, 15.3, 16.3, 17.1, 18.1, and 19.1 nm were taken and the C_α atoms of the termini were subjected to position restraints of 1.66 N/m. In this way, extended trajectories of the protein at these fixed elongations were obtained.

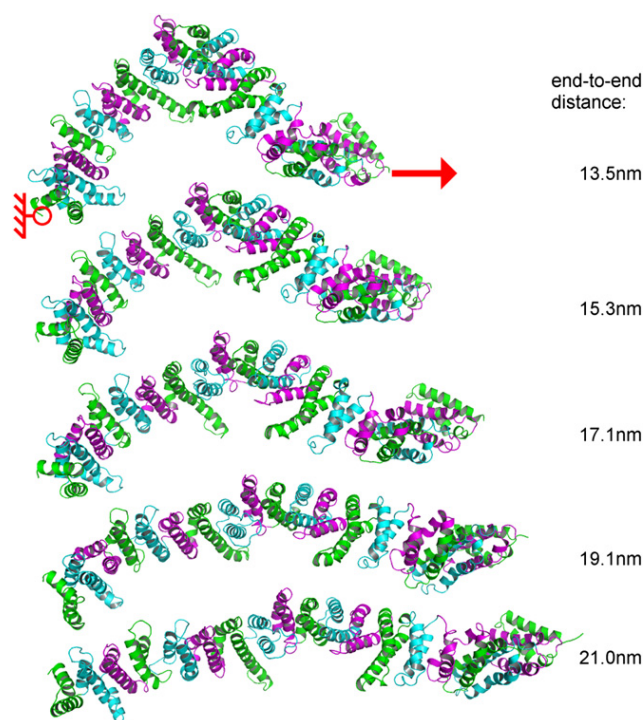


FIGURE 1 Snapshots during stretching of yImp β . Different colors represent different HEAT repeats. The C_α atom of the N-terminus was kept fixed (red circle) while a moving harmonic potential was applied to the C_α atom of the C terminus (red arrow). (Right) Numbers denote the end-to-end distance of the protein.

To quantify the equilibrium fluctuations of yImp β , a sum of n weighted Gaussian functions with center μ_i , width σ_i , and weighting factor w_i according to

$$f(d) = \sum_{i=1}^n w_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(d - \mu_i)^2}{2\sigma_i^2}\right) \quad (1)$$

was used to describe the histogram of elongations d of the protein. From the width of the Gaussians, force constants k_i were derived via

$$\sigma_i = \sqrt{k_B T / k_i},$$

where k_B is the Boltzmann constant and T the temperature used in the simulations. The statistical error was estimated as follows: For all five subsets of four out of five trajectories, spring constants and center positions of the Gaussians were determined as above. The variance of the obtained values was used as an estimate of the error.

To quantify the flexibility of amino-acid side chains, we used dihedral order parameters for torsion angles from side chains (χ -angles), as described in van der Spoel and Berendsen (32). Briefly, from using the probability distribution $p(\theta)$ of a given dihedral angle θ , the autocorrelation function $C(\infty)$ of this dihedral angle at infinite times is calculated,

$$C(\infty) = \left(\int_0^{2\pi} \cos(\theta) p(\theta) d\theta \right)^2 + \left(\int_0^{2\pi} \sin(\theta) p(\theta) d\theta \right)^2, \quad (2)$$

yielding an order parameter $S_D^2 = C(\infty)$ (33) between zero (full flexibility) and one (no flexibility). For each residue, the lowest calculated order

parameter was used. Calculations were performed using the GROMACS tool *g_chi* (32).

Entropies were calculated according to Schlitter (34), based on a principal component analysis of the backbone motions of yImp β . To obtain a sufficiently equilibrated ensemble, the first 5 ns of simulated time were omitted from the entropy calculations. Entropies were calculated from sections of the trajectory of increasing length starting at a length of $t_{\text{window}} = 20$ ns. Each subsequent trajectory part t_{window} was chosen to be 5 ns longer than the previous one. A linear fit of the calculated entropies as a function of $1/t_{\text{window}}$ was used to extrapolate entropy estimates for infinite trajectory length. As an estimate for the accuracy of the determined entropies, the statistical uncertainty of the intersection of the line fit with the y axis was used.

Simulations of the four-helix bundle protein Rop were based on the PDB structure 1ROP (35). The protein was placed in a dodecahedral box with box vectors of length 6.8 nm. Water molecules and sodium and chloride ions according to a salt concentration of 150 mM were added. Energy minimization and equilibration were performed as described above. Subsequently, a simulation of 300-ns length was carried out. All other parameters were identical to the simulations described above.

RESULTS AND DISCUSSION

Fully reversible elastic stretching

To characterize the flexibility of yImp β , we carried out 42 independent force probe simulations. In all simulations, the protein was stretched by applying a moving harmonic potential to the C-terminus, while the position of the N-terminus was kept fixed. The elongation of yImp β was measured as the distance between the C $_{\alpha}$ atoms of the N- and C-termini.

Fig. 1 shows snapshots from a representative slow ($v = 1$ m/s) stretching simulation. It is evident that, while yImp β adopted a strikingly more elongated tertiary structure under mechanical stress, its helical secondary structure remained intact. No unfolding events were observed for elongations below ~ 22 nm. Beyond that critical point, the structure showed local intermediate unfolding and separated into two segments of HEAT repeats, most frequently between HEAT repeats 4/5 and 14/15. These locations had also been suggested previously to form dynamic hot spots (8). Here, we focus at the elastic regime before unfolding.

To examine whether or not stretching of yImp β is reversible, six simulations were carried out in which the stretching potential was turned off at different elongations of the protein. Four of these simulations were started at elongations of 15.3, 17.1, 19.1, and 21.0 nm from snapshots after stretching, carried out at 1 m/s (Fig. 2 A). To investigate a possible influence of stretching velocity, two more relaxation simulations were started from a stretching simulation with a 10-fold lower stretching velocity (0.1 m/s) at elongations of 15.1 and 17.3 nm (Fig. 2 B).

In each case, yImp β was found to relax back to its original elongation. Strikingly, the initial structure was fully recovered, as testified by the low backbone root-mean-square deviation (RMSD) values with respect to the initial state (magenta, blue, and orange lines), which reached the equi-

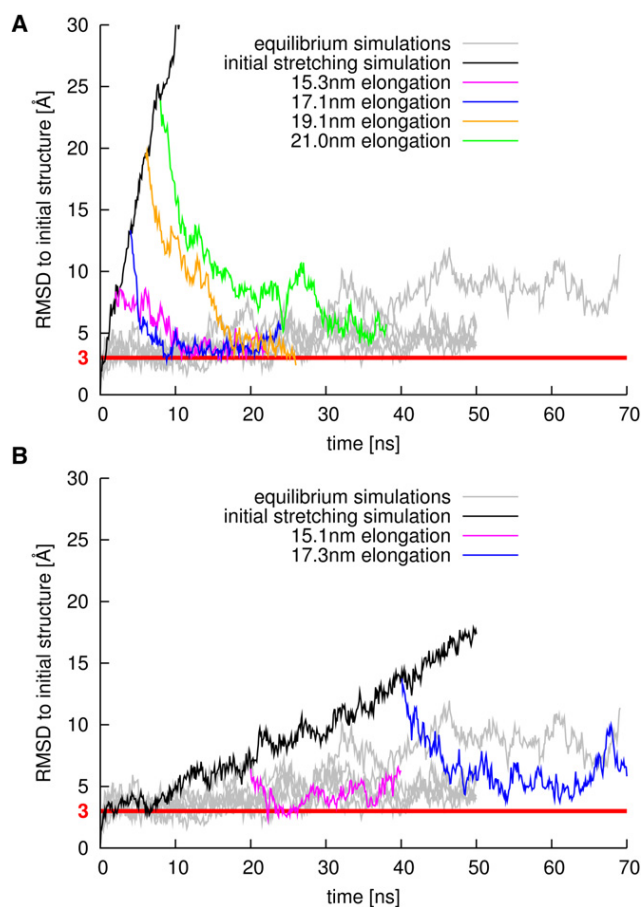


FIGURE 2 Backbone RMSD, with respect to the initial structure of yImp β during stretching (black lines) and subsequent release (magenta, blue, orange, and green lines). For comparison, gray lines display data from equilibrium simulations. (Red horizontal line) RMSD value of 3 Å. (A) Stretching at 1 m/s and subsequent release at different elongations. (B) Stretching at 0.1 m/s and subsequent release at different elongations.

librium RMSD value of simulations of free yImp β (gray lines). Relaxation simulations starting from initial RMSD values up to 20 Å returned to values at ~ 3 Å. Even when stretched close to the limit of elastic elongation, the RMSD still dropped from 25 Å to 5 Å (green line).

These results demonstrate that yImp β tolerates extreme extensions of more than twice the end-to-end distance of its RanGTP-bound structure (~ 8.9 nm) (22), showing fully reversible recovery. Moreover, all relaxation processes occurred on timescales faster than ~ 60 ns (Fig. S1 in the Supporting Material). This timescale is similar to that observed for the complete conformational transition between the nuclear and cytosolic states of the related nuclear transport receptor Cse1p (36), underscoring the functional relevance of these fast timescales for karyopherins. Furthermore, it can be seen from Fig. 2 B that fast relaxation occurs independent of the velocity of the stretching simulation carried out before. Such fully reversible elasticity (Fig. 2, A and B) is, to our knowledge, the largest seen for proteins to date,

markedly exceeding that of ankyrin repeats (37) and that observed in the recent 2-ns simulations on the protein PR65, which also consists of HEAT repeats (38).

Spring constant of yImp β

Stretching and subsequent fully reversible relaxation of yImp β suggests that the protein indeed can act as a molecular spring. We have determined its molecular spring constant via two complementary and independent approaches.

First, yImp β was subjected to weak mechanical stress, and its spring constant was obtained from the force-extension curve. The moving spring potential was chosen to advance slowly ($v = 0.1$ m/s) to minimize friction. The mean yImp β spring constant k_{str} , determined from the slope of the force profiles of three independent stretching simulations (Fig. 3 A), was found to be $(8 \pm 1) \times 10^{-3}$ N/m. The size of the observed force fluctuations is that expected for the probe spring constant of 0.83 N/m used in the stretching simulations (Fig. S2).

Secondly, extended equilibrium simulations were performed, from which the spring constant was obtained from the end-to-end distance distributions. Shaded bars in Fig. 3 B display equilibrium fluctuations of yImp β elongations obtained from a 250-ns equilibrium simulation. As can be seen from the histogram, nearly all fluctuations of yImp β can be described by a Gaussian function centered at 13 nm. The additional minor peaks at elongations of 11 nm and below belong to transient excursions to local energy minima (Fig. S1 and Fig. S3). The spring constant as well as its uncertainty was estimated from the width of the main peak by a statistical analysis as described in Methods. A spring constant of $k_{\text{eq}} = (10 \pm 4) \times 10^{-3}$ N/m was found at an equilibrium elongation of (13.0 ± 0.2) nm (dashed line in Fig. 3 B). For comparison, the solid line in Fig. 3 B shows a Boltzmann distribution for a harmonic potential with spring constant k_{str} .

The values for the spring constant of yImp β agree very well, characterizing yImp β as a very soft spring, allowing large end-to-end equilibrium fluctuations of up to 2 nm. With this stiffness, yImp β is slightly softer than the ankyrin repeats studied before ($k_{\text{eq}}(\text{experimental}) = 1.5\text{--}23 \times 10^{-3}$ N/m (17), $k_{\text{eq}}(\text{calculated}) = 16.4 \times 10^{-3}$ N/m (37)). It is far more rigid than DNA, which has a spring constant of $0.2\text{--}4.7 \times 10^{-6}$ N/m upon elastic stretching (39). Viral shells, on the other hand, need to resist high pressures of up to 60 atm, and exhibit force constants ranging from ~ 0.1 N/m to up to ~ 4 N/m upon indentation (40,41).

As calculated from its spring constant, yImp β can store up to 40 kJ/mol in its compact RanGTP-bound state. This result suggests that this mechanical strain can indeed drastically reduce the binding free energy of the tight complex and thus enable disassembly by GTP hydrolysis which ultimately drives nucleocytoplasmic transport. Table 1

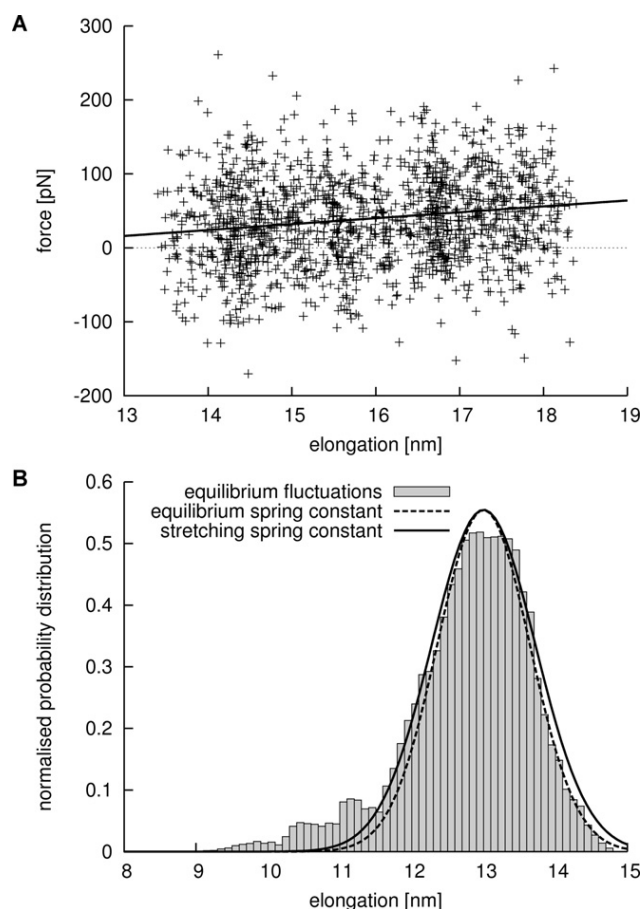


FIGURE 3 Determining the spring constant of yImp β . (A) Typical force curve from a slow stretching simulation (crosses) and linear fit to the data (solid line). (B) Equilibrium fluctuation of elongations of yImp β (shaded bars) and derivation of the spring constant. (Broken line) Boltzmann distribution of fluctuations according to the spring constant determined from equilibrium simulations (see text). For comparison, a Boltzmann distribution according to the spring constant determined in the stretching simulations is shown (solid line).

summarizes an extrapolation of these findings to complexes of human and mouse importin- β .

Structural basis for the reversible elasticity of yImp β

HEAT repeat proteins contain a well-described and extensive hydrophobic core, already detectable on a sequence level (18,21,42,43). Despite this pronounced hydrophobic core, α -solenoids appear to differ strongly from other α -helical proteins by their exceptionally high flexibility, characterized above. As a possible explanation, we hypothesized that the hydrophobic core connecting helices, despite its apparent regularity, shows differences compared to typical soluble proteins.

To test this idea, we compared successions of HEAT repeats (Fig. 4, B and D) with the repressor of primer (Rop) protein from *Escherichia coli*, a soluble protein that

TABLE 1 Estimation of stored energy in bound importin- β complexes according to an equilibrium elongation of 13 nm and a spring constant of 8×10^{-3} N/m, as derived from equilibrium and stretching simulations of yImp β

PDB id	Organism	Cargo	Elongation	Stored energy	Reference
1QGK	Human	Importin- α	8.9 nm	~41 kJ/mol	(21)
2P8Q	Human	Snurportin1	8.5 nm	~48 kJ/mol	(48)
2BKU	Yeast	RanGTP	8.9 nm	~40 kJ/mol	(22)
1UKL	Mouse	SREBP-2	10.5 nm	~15 kJ/mol	(49)

has a paradigmatic, well-studied regular hydrophobic core (Fig. 4, A and C). Indeed, Rop forms a dimer of two α -helices each, connected by a short loop, bearing close resemblance to a pair of HEAT repeats. The length of the α -helices is ~45 amino acids in HEAT repeats and ~60 in Rop, and the interhelical angle is ~20° in both cases, demonstrating close structural similarity (35,43–45). Also, no marked differences in the packing efficiency of the hydrophobic cores are observed (Fig. 4, A and B). Therefore,

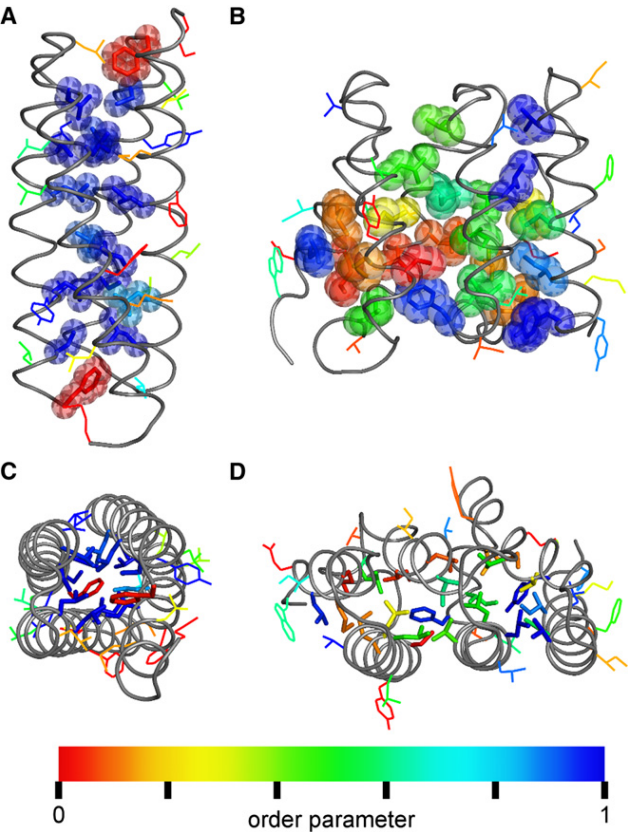


FIGURE 4 Comparison of the flexibility of hydrophobic side chains of Rop and yImp β . (A and C) Side and top view of Rop. (B and D) Side and top view of HEAT repeats 4–6 from yImp β . (Gray tubes) The protein backbone. Hydrophobic residues belonging to the respective hydrophobic core are shown as van der Waals spheres (A and B) or sticks (C and D), and as lines otherwise. The coloring reflects dihedral order parameters S^2_D (32). Values near 1 indicate rigid side chains (blue); low values reveal increased rotameric flexibility (red).

a simple structural explanation is not sufficient to explain the enhanced flexibility of yImp β . Moreover, during stretching, the contact pattern between the residues in the hydrophobic core of yImp β remains unchanged (not shown).

We therefore focused on the internal dynamics of the hydrophobic core as a second possible determinant for global yImp β flexibility. The dihedral order parameter S^2_D (32) was used as a measure for the flexibility of the side chains (color-coded in Fig. 4). As can be seen, Rop core side chains are very rigid, i.e., they are confined to single conformers, whereas most yImp β core side chains are rotationally nearly as flexible as surface-exposed side chains, despite their buried position.

This observation suggests that, by virtue of this molten globulelike conformational flexibility, the hydrophobic core of yImp β can adapt to external forces, exerted by mechanical stress or binding interactions with other proteins. The mechanical properties of such an adaptable hydrophobic core are clearly seen in force-probe simulations, by comparing Rop and a succession of HEAT repeats from yImp β . The Rop dimer did not display significant RMSD changes up to mechanical forces of 1200 pN, but then ruptured abruptly at a high force of 2900 pN. In contrast, yImp β responded to mechanical stress gradually and in a controlled manner by small changes in the structure between HEAT repeats, as evidenced by stepwise small increases in backbone RMSD (Fig. 5).

This behavior is also reflected in the energetics of the system. According to the modular composition of yImp β , we first focused on energy contributions within and between adjacent HEAT repeats. Sequence-distal interactions were found not to contribute significantly to the overall energetics of the protein and were therefore not considered here.

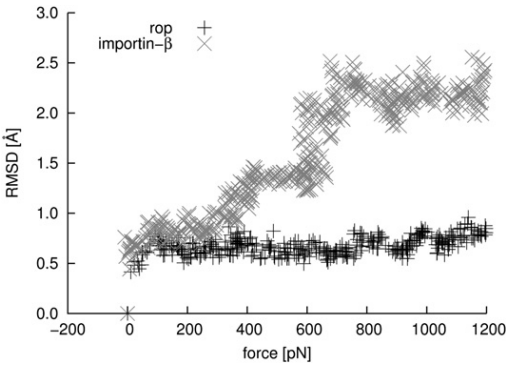


FIGURE 5 RMSD of Rop (shaded) and HEAT repeats 11 and 12 from yImp β (solid) during stretching in force-probe MD simulations. For Rop, a moving harmonic potential was applied to the center-of-mass of both chains. For yImp β , HEAT repeats 11–13 were simulated without connecting loops between HEAT repeats. A moving harmonic potential was applied to the centers-of-mass of HEAT repeats 11 and 13. In both cases, the harmonic potentials were moved perpendicular to the main longitudinal axis. The x axis shows the sum of the applied forces on both chains and HEAT repeats, respectively.

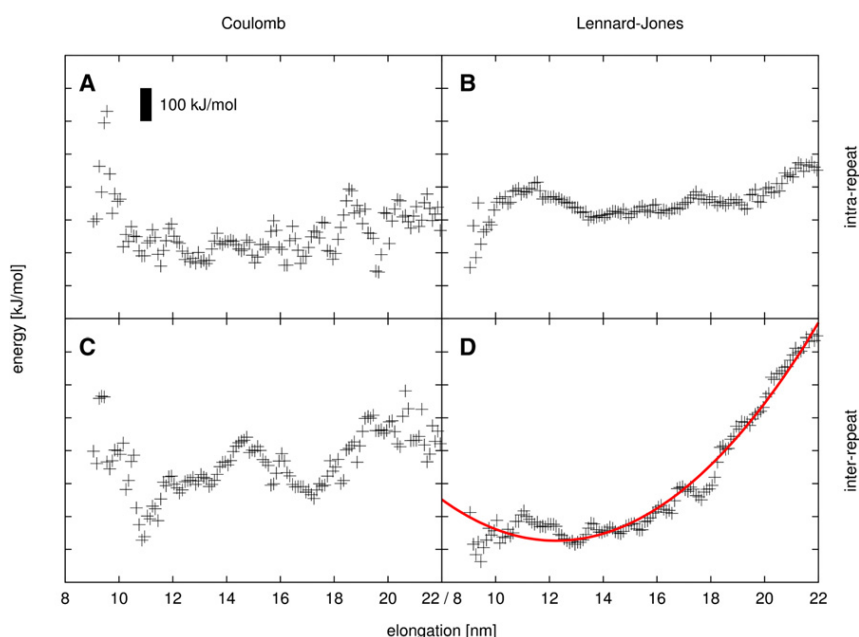


FIGURE 6 Dependence of enthalpic interactions in yImp β on molecular elongation. (A) Intra-HEAT-repeat Coulombic interactions. (B) Intra-HEAT-repeat Lennard-Jones interactions. (C) Inter-HEAT-repeat Coulombic interactions. (D) Inter-HEAT-repeat Lennard-Jones interactions and harmonic fit (red line). (Black bar) Energy interval of 100 kJ/mol.

Fig. 6 displays intrarepeat and interrepeat Coulombic and Lennard-Jones energies, respectively, as a function of the elongation of the protein, averaged over 0.1-nm intervals. For averaging, 250-ns equilibrium and ~740-ns stretching simulations were merged, altogether totaling ~990 ns of simulation. It is evident from Fig. 6 D that the largest contribution to the total increase in enthalpy upon elongation comes from Lennard-Jones interactions between neighboring HEAT repeats. By comparison, intrarepeat Lennard-Jones interactions (Fig. 6 B) as well as intra- and intermolecular Coulomb interactions (Fig. 6, A and C) only show relatively low increases with elongation.

As the quadratic fit shows, the dominant interrepeat Lennard-Jones energy is well described by a harmonic potential with a spring constant of 23×10^{-3} N/m. When combined, all four energy contributions yield a spring constant of 47×10^{-3} N/m and an equilibrium elongation of 13.2 nm (blue line in Fig. 7 B, see below).

This equilibrium elongation agrees very well with the equilibrium end-to-end distance obtained from our free and stretching simulations. In contrast, the obtained spring constant is ~4 times larger, which suggests that additional, compensating contributions must be relevant.

To identify these contributions, the change of entropy with elongation was estimated (Fig. 7 A), using the quasi-harmonic approximation formulated by Schlitter (34). To this end, we carried out extended simulations, in which the end-to-end-distance of yImp β was kept fixed at different values, whereas motions on all other degrees of freedom of the protein were left unrestricted.

Fig. 7 A shows that, indeed, the entropy of yImp β rises with increasing elongation of the protein. This finding is in contrast to the simple worm-like chain model of polymer chains, which would predict an entropy decrease. We attri-

bute this rise in entropy to the unusual flexibility of the hydrophobic core discussed above, which, upon stretching, facilitates relaxation and leads to an enhanced flexibility of the entire protein.

Fig. 7 B shows that this rise in entropy indeed accounts for the discrepancy between the spring constant derived from purely enthalpic terms (blue line) and the overall spring constant observed for yImp β in equilibrium and stretching simulations (red line). When the entropic contribution to the free energy (Fig. 7 A) is subtracted from the enthalpic part (Fig. 7 B, blue line), agreement within the error bars with the overall spring constant of yImp β (Fig. 7 B, red line) is obtained (Fig. 7 B, black bars). In summary, the reversibly elastic, springlike behavior of yImp β can be largely ascribed to contributions from Lennard-Jones interactions of the hydrophobic core between HEAT repeats and to substantial entropy changes associated with different degrees of elongation.

CONCLUSIONS

Our results characterize the α -solenoid yImp β as a soft nanospring that tolerates up to twofold extensions without rupture of its hydrophobic core and is capable of fully reverting back to its original structure after stress release. Upon stretching, the hydrophobic core between HEAT repeats is able to adapt to the external changes by an internal rearrangement of hydrophobic residues, while the contact pattern between hydrophobic core residues is fully preserved. Our analysis of side-chain flexibilities suggests that yImp β and, by similarity, also most other α -solenoid structures, exhibit a molten globulelike hydrophobic core and that this core structure underlies the flexibility of these proteins. Molten globule states are defined as possessing

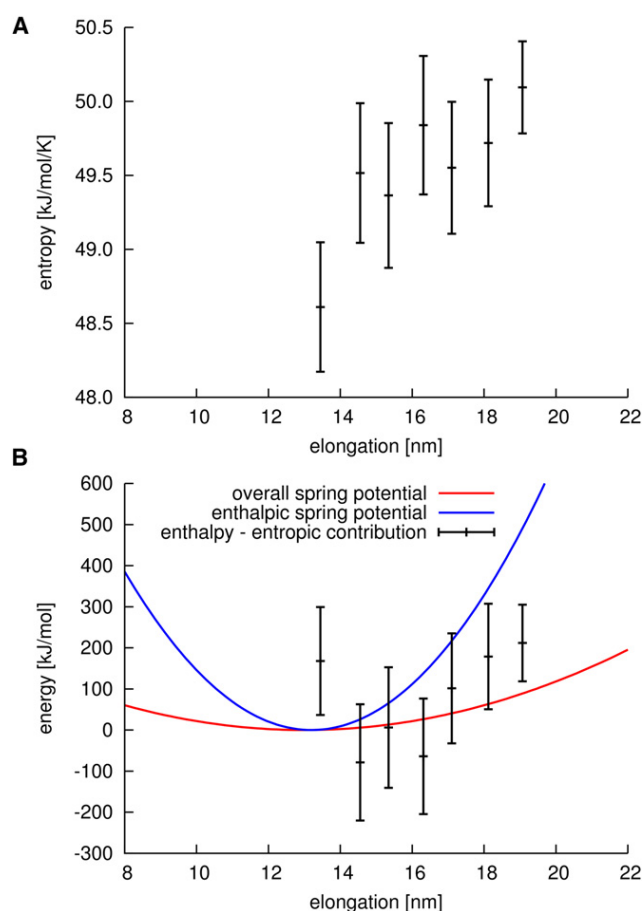


FIGURE 7 (A) Entropy estimation for different elongations of yImp β . (B) Spring potential derived from purely enthalpic terms (blue line), and spring potential observed in equilibrium and stretching simulations (red line). The discrepancy between the two potentials is resolved when the entropic contribution to the free energy is subtracted from the purely enthalpic potential (black bars).

stable secondary structure elements and variable tertiary structure, as seen here, and they are assumed to differ from fully ordered states by relatively loose packing of the hydrophobic core, allowing side-chain rearrangements (11). Our simulations demonstrate that large conformational changes of the global structure of α -solenoids can be effected by the accumulation of many small changes between repeating secondary structure units. Alpha-solenoids thus bridge the structural gap between fully folded and intrinsically unfolded proteins both in terms of global structure and the mobility of core side chains (6–8,46). These unexpectedly highly flexible proteins also will deserve further studies addressing the question of whether conformational selection or induced fit is the predominant mechanism in their protein-protein binding pathways (47).

SUPPORTING MATERIAL

One table and four figures are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(10\)00775-7](http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)00775-7).

We thank Bostjan Kobe, Peter Hinterdorfer, and Ziv Reich for helpful discussions.

This work was partially funded by grant No. RGP53/2004 from the Human Frontier Science Program.

REFERENCES

- Kajander, T., A. L. Cortajarena, ..., L. Regan. 2005. A new folding paradigm for repeat proteins. *J. Am. Chem. Soc.* 127:10188–10190.
- Hierro, A., A. L. Rojas, ..., J. H. Hurley. 2007. Functional architecture of the retromer cargo-recognition complex. *Nature*. 449:1063–1067.
- Stewart, M. 2007. Molecular mechanism of the nuclear protein import cycle. *Nat. Rev. Mol. Cell Biol.* 8:195–208.
- Fotin, A., Y. Cheng, ..., T. Walz. 2004. Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature*. 432:573–579.
- Brohawn, S. G., J. R. Partridge, ..., T. U. Schwartz. 2009. The nuclear pore complex has entered the atomic age. *Structure*. 17:1156–1168.
- Kobe, B., and A. V. Kajava. 2000. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* 25:509–515.
- Wright, P. E., and H. J. Dyson. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293:321–331.
- Zachariae, U., and H. Grubmüller. 2008. Importin- β : structural and dynamic determinants of a molecular spring. *Structure*. 16:906–915.
- Ptitsyn, O. B., R. H. Pain, ..., O. I. Razgulyaev. 1990. Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett.* 262:20–24.
- Tsou, C. L. 1995. Inactivation precedes overall molecular conformation changes during enzyme denaturation. *Biochim. Biophys. Acta*. 1253:151–162.
- Ptitsyn, O. B. 1995. How the molten globule became. *Trends Biochem. Sci.* 20:376–379.
- Vamvaca, K., B. Vögeli, ..., D. Hilvert. 2004. An enzymatic molten globule: efficient coupling of folding and catalysis. *Proc. Natl. Acad. Sci. USA*. 101:12860–12864.
- Neuman, K. C., and A. Nagy. 2008. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods*. 5:491–505.
- Serquera, D., W. Lee, ..., L. S. Itzhaki. 2010. Mechanical unfolding of an ankyrin repeat protein. *Biophys. J.* 98:1294–1301.
- Werbeck, N. D., P. J. E. Rowling, ..., L. S. Itzhaki. 2008. Shifting transition states in the unfolding of a large ankyrin repeat protein. *Proc. Natl. Acad. Sci. USA*. 105:9982–9987.
- Rief, M., J. Pascual, ..., H. E. Gaub. 1999. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J. Mol. Biol.* 286:553–561.
- Lee, G., K. Abdi, ..., P. E. Marszalek. 2006. Nanospring behavior of ankyrin repeats. *Nature*. 440:246–249.
- Andrade, M. A., and P. Bork. 1995. HEAT repeats in the Huntington's disease protein. *Nat. Genet.* 11:115–116.
- Liu, S. M., and M. Stewart. 2005. Structural basis for the high-affinity binding of nucleoporin Nup1p to the *Saccharomyces cerevisiae* importin- β homologue, Kap95p. *J. Mol. Biol.* 349:515–525.
- Conti, E., C. W. Müller, and M. Stewart. 2006. Karyopherin flexibility in nucleocytoplasmic transport. *Curr. Opin. Struct. Biol.* 16:237–244.
- Cingolani, G., C. Petosa, ..., C. W. Müller. 1999. Structure of importin- β bound to the IBB domain of importin- α . *Nature*. 399:221–229.
- Lee, S. J., Y. Matsuura, ..., M. Stewart. 2005. Structural basis for nuclear import complex dissociation by RanGTP. *Nature*. 435:693–696.
- van der Spoel, D., E. Lindahl, ..., H. J. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.

24. Hess, B., C. Kutzner, ..., E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.
25. Jorgensen, W. L., and J. Tirado-Rives. 1988. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic-peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.
26. Jorgensen, W. L., J. Chandrasekhar, ..., M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Comput. Phys.* 79:926–935.
27. Hess, B., H. Bekker, ..., J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.
28. Berendsen, H. J. C., J. P. M. Postma, ..., J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
29. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an $n\log(n)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
30. Essmann, U., L. Perera, ..., L. G. Pedersen. 1995. A smooth particle mesh Ewald potential. *J. Chem. Phys.* 103:8577–8592.
31. Grubmüller, H., B. Heymann, and P. Tavan. 1996. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*. 271:997–999.
32. van der Spoel, D., and H. J. C. Berendsen. 1997. Molecular dynamics simulations of Leu-enkephalin in water and DMSO. *Biophys. J.* 72:2032–2041.
33. Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. I. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546–4559.
34. Schlitter, J. 1993. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* 6:617–621.
35. Banner, D. W., M. Kokkinidis, and D. Tsernoglou. 1987. Structure of the ColE1 Rop protein at 1.7 Å resolution. *J. Mol. Biol.* 196:657–675.
36. Zachariae, U., and H. Grubmüller. 2006. A highly strained nuclear conformation of the exportin Cse1p revealed by molecular dynamics simulations. *Structure*. 14:1469–1478.
37. Sotomayor, M., D. P. Corey, and K. Schulten. 2005. In search of the hair-cell gating spring elastic properties of ankyrin and cadherin repeats. *Structure*. 13:669–682.
38. Grinthal, A., I. Adamovic, ..., N. Kleckner. 2010. PR65, the HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis. *Proc. Natl. Acad. Sci. USA*. 107:2467–2472.
39. Bustamante, C., S. B. Smith, ..., D. Smith. 2000. Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* 10:279–285.
40. Ivanovska, I. L., P. J. de Pablo, ..., G. J. Wuite. 2004. Bacteriophage capsids: tough nanoshells with complex elastic properties. *Proc. Natl. Acad. Sci. USA*. 101:7600–7605.
41. Zink, M., and H. Grubmüller. 2009. Mechanical properties of the icosahedral shell of southern bean mosaic virus: a molecular dynamics study. *Biophys. J.* 96:1350–1363.
42. Andrade, M. A., C. Petosa, ..., P. Bork. 2001. Comparison of ARM and HEAT protein repeats. *J. Mol. Biol.* 309:1–18.
43. Kresse, H. P., M. Czubayko, ..., H. Bloecker. 2001. Four-helix bundle topology re-engineered: monomeric Rop protein variants with different loop arrangements. *Protein Eng.* 14:897–901.
44. Munson, M., S. Balasubramanian, ..., L. Regan. 1996. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci.* 5:1584–1593.
45. Willis, M. A., B. Bishop, ..., A. T. Brunger. 2000. Dramatic structural and thermodynamic consequences of repacking a protein's hydrophobic core. *Structure*. 8:1319–1328.
46. Gsponer, J., and M. M. Babu. 2009. The rules of disorder or why disorder rules. *Prog. Biophys. Mol. Biol.* 99:94–103.
47. Lange, O. F., N. A. Lakomek, ..., B. L. de Groot. 2008. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*. 320:1471–1475.
48. Mitrousis, G., A. S. Olia, ..., G. Cingolani. 2008. Molecular basis for the recognition of snurportin 1 by importin β . *J. Biol. Chem.* 283:7877–7884.
49. Lee, S. J., T. Sekimoto, ..., Y. Yoneda. 2003. Crystal structure of importin- β and SREBP-2 complex. *Science*. 302:1513–1514.

Lebenslauf

Persönliche Daten

Name	Christian Kappel
Adresse	Otto-Laufer-Straße 4b, 37077 Göttingen
Geburtsdatum/-ort	1. November 1980, Kassel
Nationalität	Deutsch

Ausbildung

06/2000	Abitur an der Jacob-Grimm-Schule, Kassel
07/2000 - 05/2001	Zivildienst in der Seniorenwohnanlage Habichtswald, Kassel
10/2001 - 03/2007	Studium der Physik an der Georg-August-Universität Göttingen
04/2005 - 10/2006	Diplomarbeit am Max-Planck-Institut für biophysikalische Chemie bei Prof. Dr. Helmut Grubmüller
03/2007 - 02/2011	Promotion am Max-Planck-Institut für biophysikalische Chemie bei Prof. Dr. Helmut Grubmüller

