Internal Friction in Peptide Kinetics

am Fachbereich Physik der Freien Universität Berlin
eingereichte Dissertation

zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften (Dr. rer. nat.)

vorgelegt von

Julius C. F. Schulz

Berlin, April 2015

Disputation: 10. September 2015

1. Gutachter / Betreuer:

Prof. Dr. Roland R. Netz

Freie Universität Berlin
Fachbereich Physik
Arnimallee 14
D-14195 Berlin

2. Gutachter:

Prof. Dr. Frank Noé

Freie Universität Berlin
Fachbereich Mathematik
Arnimallee 6
D-14195 Berlin

Contents

Chapter 1

Introduction

Proteins belong to the most important substances throughout nature, since they account for most of the functionality in all living organisms. Known since the eighteenth century, they were first described by Gerardus Johannes Mulder in 1839 [1]. It took more than a hundred years from that until, for the first time, the chemical structure of a protein was derived by Kendrew et al. in their study of the myoglobin molecule. Surprisingly to the scientific world at this time, the structure lacked any signs of symmetry:

"Perhaps the most remarkable features of the molecule are its complexity and its lack of symmetry. The arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates, and it is more complicated than has been predicated by any theory of protein structure. Though the detailed principles of construction do not yet emerge, we may hope that they will do so at a later stage of the analysis." [2]

This statement by Kendrew already foreshadows the *problem of protein folding*, which entangles many scientists until today. As stated in a review on the fiftieth anniversary of the nobelprize awarded to Kendrew and Perutz [3], the *problem of protein folding* can be summarized by three main questions:

1. The physical folding code: How is the 3D native structure of a protein determined by the physicochemical properties that are encoded in its 1D amino-acid sequence?
2. The folding mechanism: A polypeptide chain has an almost unfathomable number of possible conformations. How can proteins fold so fast?
3. How to predict (with the aid of computers) protein structures?

Because proteins account by a large amount to the functionality of the human body, understanding their nature is crucial for the physiological sciences. Many genetic disorders are directly linked to misfolded proteins, such as Huntington's disease.

The physical folding code Since the experiments of Mulder, it was known that proteins consists of compounds, which are now known as aminoacids. He in fact believed proteins to be made up of a single compound species. It was the work of Hofmeister [4] and Fischer [5], that revealed their nature as polypeptides. During the 1930ies, William Astbury's crystallographic studies [6] of proteins raised the hypothesis of hydrogen bonds holding proteins together. But it was not until 1951, when Linus Pauling and coworkers [7] predicted the α and β structures, which together account for most secondary structure motifs in wild-type proteins [8, 9]. The findings of Kauzmann [10] provided another fundamental driving force in protein folding: the hydrophobic interaction, where aminoacids try to minimize the water exposed surface. Together with the fundamental interactions in molecular physics, the electrostatic and the van-der-Waals interaction, hydrogen bonding and the hydrophobic effect provide the major components that can explain the three dimensional structure of proteins. The folding mechanism is commonly described as a search for the lowest energy configuration in the high-dimensional phase space made up by the atoms constituting the protein.

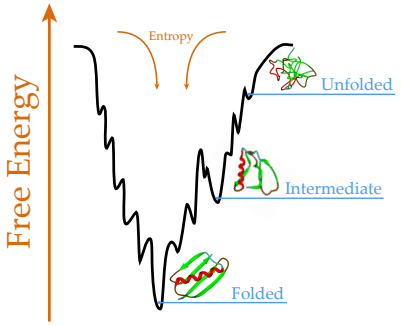

FIGURE 1.1: The protein folding process can be understood as a stochastical search in a high-dimensional landscape (here projected onto a single dimension) driven by entropic forces. The folded state(s) are corresponding to states of lowest free energy, so that the free energy landscape adopts a funnel-like shape.

The folding mechanism The second question in above's list is the classic Levinthal's paradox [11]: Given the extremely high dimensionality of the phase space, how is it possible to find proteins that fold on timescales as short as microseconds? He then postulated the concept of the *folding pathway*, along which the protein is able to carry out that search quickly. Given that most protein have only a single *native state*, this led to the picture of a funnel-like free energy landscape along that pathway, where the lowest energy states are identified with the folded state. This is depicted in Figure 1.1. Here we show the free energy of the peptide as a function of some arbitrary projection of the phase-phase, also known as *reaction coordinate*. Ideally this projection is chosen such that the folding pathway is clearly visible and the folded state can be distinguished well from its intermediate and unfolded conformations. This depiction, the free energy as a function of some projection, is called a free energy landscape. The choice of the reaction coordinate depends strongly on the specific peptide and is crucial for a good description of the folding dynamics. An example for a reaction coordinate that can be

obtained both experimentally and from simulations is the end-to-end distance between the first and last residue of the peptide or protein.

This simple model however tells little about the dynamics of protein folding, since the free energy landscape is only providing static information on the equilibrium population. The search for the folded state is typically assumed as a diffusional process driven by Brownian motion. In the most simple form, the folding process can be described by a two state process: two metastable states, *folded* and *unfolded*, form two minima in a one-dimensional free energy landscape, which are seperated by an energy barrier of height ΔE (cf. Fig.1.2). In the limit of strong frictional damping, the barrier-crossing time τ (i.e. the folding time) is given according to Kramers theory [12] as:

$$\tau^{-1} = \frac{\omega_a \, \omega_b \, m}{2 \, \pi \, \gamma} \mathrm{e}^{-\beta \, \Delta E}, \tag{1.1}$$

where ω_a and ω_b are the curvatures of the free energy as depicted in Figure 1.2, m is the mass and γ is the friction constant in the underlying Langevin equation. According to Stokes' Law, the frictional drag γ on a small particle in a fluid is proportional to the viscosity η of the fluid. We can subsequently rewrite Eq. 1.1 as:

$$\tau^{-1} \sim \frac{1}{\eta} \mathrm{e}^{-\beta \, \Delta E}, \tag{1.2}$$

Thus, for a simple two-state folder the folding time should scale with the viscosity of the solvent: $\tau \sim \eta$.

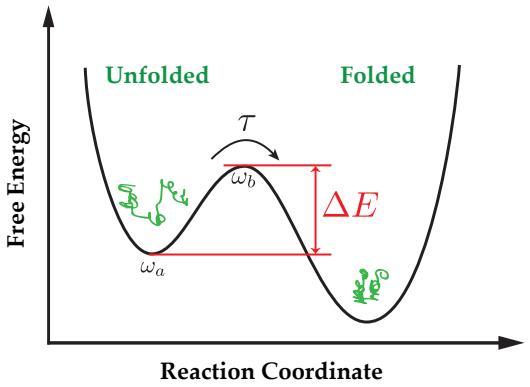

FIGURE 1.2: The dynamics of a protein that has two distinct states, folded and unfolded, is projected onto a suitable one-dimensional reaction coordinate. In this picture the two states correspond to energy minima seperated by an energy barrier of height ΔE. Given the curvatures ω_a and ω_b of the free energy function, the barrier-crossing time τ according to Kramers theory is given by Eq 1.1.

This result has two important implications. First it sets a lower limit to protein folding, i.e. the protein cannot fold faster than it can diffuse in the free energy landscape. Secondly, this is a prediction that can in principle be easily tested by experiments. The challenge is to find peptides whose folding speed is close or even below this limit and to find experimental methods to also monitor this process.

Fortunately, in the last twenty years, several such studies have been carried out and a number of fast folding proteins have been found. Probably the first to do so were Eaton and coworkers in 1992 [13]. Using nanosecond lasers they measure the relaxation of myglobin after ligand dissociation. By adding glycerol as a viscogen, they are able to measure the conformational relaxation as a function of the viscosity. While they confirm the linear relationship of Eq. 1.2 in priciple, they make an important discovery: For low viscosities, the rate constant does not further increase with lowering the viscosity but levels off and thus they propose a new relationship between folding time and solvent viscosity: $\tau \sim \eta + \sigma$, where they attribute the contribution σ to what they call internal friction, i.e. friction effects not originating from the solvent but from the internal barriers that slow down the motion of the atoms. This finding is also supported by studies of Kiefhaber and coworkers [14], who study the loop-folding kinetics of small model peptides using FRET experiments.

Another particularly revealing experiment in this direction was executed by Jas, Eaton and Hofrichter [15]. They use temperature jump experiments to measure the relaxation time of a small α-helix and a β-hairpin. While they confirm Kramers law for the β-hairpin, they find a completely different relationship for the α-helix, where the folding rate τ^{-1} follows a power-law: $\tau \sim \eta^{\alpha}$. In Chapter 3 of this thesis we will investigate this in more detail and give possible explanations for the origin of such different results.

When talking about internal friction, this of course raises the question, what actually creates internal friction. Manke and Williams [16] start from the perspective of an ideal flexible polymer and identify deviations to be responsible for the internal friction. Many such deviations can be envisioned for peptides: hydrogen bonding, backbone reorientation, hydrophobic interactions etc. In a broader sense we can look at internal friction as the rugeddness of the free energy landscape. This perception nicely links the microscopic friction of a peptide to the well known macroscopic friction, that is caused by pulling a large object over a microscopically rugged surface. The role of friction is very much the same in both cases. If the object of interest is dragged along a certain path, energy is dissipated.

It is important to note that the simple model of a two-state folder given above is not valid in general. Not only are folding processes more complex involving intermediate states, the concept of a single native folded structure would be misleading. Instead an entire ensemble of conformations should be considered, whose population is not only depending on outer conditions but is dynamic and not neccessarily representing the equilibrium population in this free energy landscape. Not only are energetics competing but also timescales. This is of significant importance when the stability of different secondary structures are compared. In Chapter 4 we will discuss how internal friction differs between different secondary structure motifs, namely α-helical and β-hairping structures, and how that influences the kinetic properties of the peptide.

Chapter 2

Computational Methods

Studying objects on a molecular and submolecular level requires a careful choice of methods. Since peptides, as studied in this work, live on a nanometer scale, one has to take into account the peculiarities of inter- and intramolecular interactions. The modelling on this length-scale is particularly challenging, since long-range interactions that bridge into mesoscopic length scales and short-ranged interactions originating from quantum effects are highly correlated. This correlation usually hinders the treatment of long-ranged interactions in a continuum model, thus requiring the inclusion of a relatively large environment around the actual object of interest.

Once the modelling of the intra- and intermolecular interactions is accomplished, it is, on a formal level, straightforward to obtain the quantities of interest by applying the methods of statistical mechanics. The high degree of correlation in the system and the large dimension of the phase space due to the size of the system however makes it more or less impossible to evaluate the resulting Hamiltonian analytically or with conventional numerical methods. When the correlations in the system are small, mean-field theories can provide suitable solutions, however they typically fail on objects like peptides.

One therefore has to switch to more approximate ways of evaluating the system's Hamiltonian. If these methods attempt to be a computational representation of a real or model system rather than some mathematical equation, they are usually called computer simulations. Two main classes may be distinguished: static evaluations of the phase space using e.g. Monte Carlo simulations, and kinetically oriented methods such as Molecular Dynamics (MD) or Brownian Dynamics (BD) simulations. While the former are well suited for problems where only static properties are of interest and they are typically computationally less expensive, the latter give insight into dynamic properties as well. However the overcoming of high energy barriers can be challenging.

When simulating any system, one has to choose the level of abstraction. There is a wide range of methods between fully solving the quantum mechanics and coarse graining up the level of entire molecules being the building blocks. The influence of the water on peptide dynamics is significant, producing hydrogen bonds and the hydrophobic effect. As these effects reach down to the interaction between single atoms, they are

hard to describe by continuous, implicit water models. Therefore, we mostly study all-atom systems, meaning that atoms are the smallest object in the system and all-atoms refers to also taking solvent atoms into account, namely water.

All of these methods provide an exact solution of the model to a certain statistical uncertainty and, in the limit of enough computational effort, they evaluate the relevant phase space and yield an exact solution of the underlying Hamiltonian. Therefore, we are able to judge the quality of the modelling of the system's interactions. Or we can, if sufficiently convinced of the quality of the molecular modelling, make predictions on certain properties of interest. We thus have a powerful tool at hand to study nanometer sized objects and gain an understanding at the submolecular level.

2.1 Molecular Dynamics Simulations

The major focus of this thesis lies on friction effects, which are mainly a dynamic property and thus molecular dynamics simulations are a natural choice as motivated above. Molecular dynamics simulations are a semiclassical approximation to the actual dynamics. Semiclassical in a sense that the simulations follow Newtonian mechanics, while the interactions forces are an empirical approximation to the underlying quantum description. Thus a molecular dynamics simulation is in principal simply numerically integrating Newtons equations of motion.

In practice however, the efficient implementation of such an integrator is a major challenge. In this work an integration timestep of 2 fs is used, resulting in several hundred million integration steps for typical timescales in the order of several microseconds. Therefore, molecular dynamics simulations are usually run using ready-to-use packages, such as Gromacs [17], which is used here.

As will be discussed in detail further below, proper care has to be taken when setting up the interactions of an MD simulation. Typically, in a molecular dynamics simulation only three types of interactions are present:

- Bonded interactions: usually rigid or harmonic potentials that keep distances and angles between atoms in a desired range
- Lennard-Jones (LJ) interaction: The LJ interaction represents the non-electrostatic part of the interactions between nonbonded particles.

$$V_{LJ} = 4\,\epsilon \left(\frac{\sigma^{12}}{r^{12}} - \frac{\sigma^{6}}{r^{6}} \right) \tag{2.1}$$

  It consistst of two parts, an attractive part $\sim r^{-6}$ due to van-der-Waals forces and a repulsive part in order to obey the Pauli principle. The parameters ϵ and σ are empirically chosen to approximate the complex many-body quantum mechanical interactions in the condensed state.

- Electrostatic interaction: every atom is assigned a certain charge, which results in a Coulomb interaction

While angular potentials are typically implemented as harmonic, this can be unfavorable for bonds. The stiff potentials that would reproduce the desired bonded properties lead to vibrations, especially for the light hydrogen atoms. To properly calculate these without corrupting the integration, small time steps are required, which in return leads to a higher computational effort. Since these vibrations are lying in a quantum dominated regime, they are anyway not properly calculated and as such unphysical, thus removing these is not wrong either. The methods to keep atom bonds stiff without a harmonic potential are called constraint algorithms. For 3-site molecules such as water an analytical solution named SETTLE [18] exists, for more complicated molecules we use the LINCS [19] algorithm.

Lennard-Jones interactions can be efficiently treated using cut-offs, since they are short ranged on a atomar length scale due to their r^{-6} decay. This is however not true for the electrostatic potential, which decays as r^{-1}. For typical box sizes of a few nanometers, the electrostatic interaction is still of non-negligible order. Therefore, assumptions have to be made of how the system should look beyond the box size. The typical assumption is that the system is repeated infinitely in each cartesian direction (for square or rectangular boxes). Rather than summing up the interactions of the repetitions of the boxes, it is more convenient to use a Fourier transformed expression of the Coulomb interaction. With a proper resummation, called Ewald summation [20], the series can be truncated and thus be computed efficiently. More elaborate versions are parallelisable and are known as Particle Mesh Ewald (PME) methods [21] and are used in this work. Also in this method, the short ranged part of the Coulomb interaction is calculated separately in real space and cut off just as the LJ-potential. We use a cutoff of $r_c = 0.9\,\mathrm{nm}$.

The standard formulation of a Hamiltonian in a periodic box would lead to what's known as an NVE ensemble, i.e. fixed particle number, volume and energy, where the energy is given by the particle positions (potential energy) and initial speed (kinetic energy). When comparing to thermodynamic systems that ought to model real systems, this is usually not the correct ensemble. The most common ensemble is probably the isobaric-isothermic or NPT ensemble, where measures are taken to keep the temperature and pressure at a certain value. For the temperature the v-rescale thermostat [22] is used and as a barostat we use the Parinello-Rahman [23] algorithm. Due to computational reasons, we also use the isothermal (NVT) ensemble, which does not use a barostat but keeps the volume fixed. Since in the thermodynamic limit all ensembles should give the same results, one can reason that given proper equilibration, the NVT ensemble can give the proper dynamics when pressure fluctuations are not too large.

2.1.1 Mass Scaling of the Solvent

One property we are especially interested in in this work is the solvent viscosity. For studying a protein at a wide range of solvent viscosities, we use a modification of the

standard MD protocol, where the solvent's (i.e. water) mass is scaled by a certain factor c. Under mass scaling, Newton's equations of motion stay invariant

$$cm\frac{d^2\mathbf{r}}{d(\sqrt{c}t)^2} = \mathbf{F}, \tag{2.2}$$

and thus the equilibrium properties will not be change. However, all dynamics properties will scale als $\sqrt{c}$. In Fig. 2.1 one can see that the solvent viscosity indeed scales as $\eta/\eta_0 \sim \sqrt{m/m_0}$. While this method has been previously applied to speed up equilibration, we use it in a novel way to study internal friction effects. It is particularly suited, since the free energy landscape of the protein is by definition not altered, whereas in experiments the effects of added viscogens (e.g. polyethylenglycol) might be present. With this method we can study a wide range of solvent viscosities and most importantly also viscosities far lower than water viscosity, which is experimentally unreachable, since typical viscogens only increase the viscosity of water.

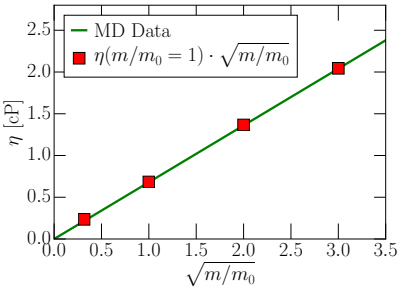

FIGURE 2.1: Viscosity η of SPC/E water for different water masses measured using a Green-Kubo formalism [24]: $\eta = \lim_{\tau\to\infty} \frac{V}{k_B T} \int_0^\tau \left\langle \sigma_{xz}(t_0)\, \sigma_{xz}(t_0 + t) \right\rangle_{t_0} \mathrm{d}t$, where σ_{xz} is one off-diagonal element of the stress-tensor.

2.1.2 Steered Molecular Dynamics

The basic concepts of molecular dynamics simulations laid out above often assume systems to be in equilibrium. This is however no necessity when using the MD method. One particular example of non-equilibrium molecular dynamics simulations is called *Steered Molecular Dynamics Simulations*, where an external force is applied to a group of atoms in the system. For example, this method can be used to obtain and thereby study processes that would otherwise be out of reach on the typical timescales of molecular dynamics simulations, such as protein unfolding or ligand-receptor bindings. Experimentally there exist conceptually very similar techniques, such as Atomic Force Microscopy (AFM) [25] und optical tweezers [26] and this can provide a link between theory and experiment [27].

In principal, there are two types of pulling scenarios: constant force and constant velocity. In this work, we investigate the dissipative effects during unfolding, thus constant velocity pulling is more appropriate. This is achieved by putting the atom group in

a moving harmonic potential (described in detail in chapter 4). Proper care has to be taken of the periodic boundaries.

2.2 Force Fields

As laid out above, the modelling of some molecular system relies on the proper choice of interactions and in the case of standard molecular dynamics simulations this includes bond parameters, LJ-parameters and partial charges. A combined set of such parameters for some class of molecules is called a force field. Since it is virtually impossible at this time to fulfill all experimental observations using a single forcefield, a subset of parameters of interest out of all available observables is chosen to benchmark the forcefield against. For simple molecules such as water this can be properties such as density, viscosity or dielectric permittivity. For complex molecules like peptides one usually looks at folding properties, such as helix propensities.

2.2.1 Water

Although being a very small molecule with only three atoms, water is notoriously complex in its behaviour and has been a major challenge to simulate up to today. Due to its complex interaction with itself and the peptide of interest, it is hard to replace water by a coarse grained model, although attempts exist and do work under certain circumstances. But once an atomistic description of water is necessary, a large number is needed since no simple continuum transition exists. This makes all-atom simulation of peptides and other large systems computationally expensive.

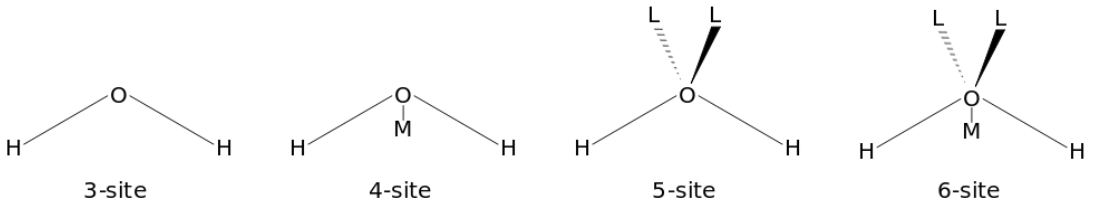

FIGURE 2.2: Different classes of water models based on their number of interaction sites. Sites denoted *M* and *L* are called *dummy atoms* and they don't carry a mass but typically carry the negative charge so that the true charge distribution of the water molecule is better represented. Thus they usually give better results than the 3-site water model, albeit with a higher computational and modelling effort. [28]

Over the time several different models have been proposed, but they can loosely be classified using the number of interaction sites assumed. The most simple models are 3-site models, one mass-carrying site for each atom of the water molecule. Models with more interaction sites have several mass-less particles, that can however still be carrying charges. Throughout this work we use the SPC/E water model [29], a 3-site water model, which, is widely used in the field of protein simulations and several protein forcefields are benchmarked using this water model. This water model is well known

for reproducing the dielectric properties of water well, a property that makes it well suited for peptide simulations.

2.2.2 Peptides

While water is a very simple molecule of only (in this case) three interaction sites with a rigid conformation, peptides and proteins are far more complex. Since peptides are flexible, one also has to find proper parameters for angle potentials. For a chain of four atoms, cis and trans configurations are usually favorable over all other possible alignments. As this is a four-site interaction, angle potentials are usually not able to reproduce this correctly. Thus another type of bonded interactions is introduced, that has to be modelled properly, so called torsional potentials.

In the field of protein simulations, a large number of forcefields exist, each with even more variations attached to cope with special cases and minor improvements. A proper choice is therefore hard to make and due to restrictions on computational power available, proper testing of several different forcefields might be difficult to accomplish. In this work we choose the Amber03 forcefield, which is widely used for studying peptides and proteins. However one has to remember the weaknesses of the choice made. In the case of the Amber03 [30] forcefield, it is known that helices are overstabilised [31] and a comparison to experimental setups might be difficult. But this can also be of use, since small homopolymeric helices are not stable in an actual experiment. If the governing principles would be the same however, this way makes it much easier to study peptides that can be easily grasped from the perspective of a theoretician.

Chapter 3

Peptide Chain Dynamics in Light and Heavy Water: Zooming in on Internal Friction

Frictional effects due to the chain itself, rather than the solvent, may have a significant effect on protein dynamics. Experimentally, such "internal friction" has been investigated by studying folding or binding kinetics at varying solvent viscosity; however the molecular origin of these effects is hard to pinpoint. We consider the kinetics of disordered glycine-serine and α-helix forming alanine peptides, and a coarse-grained protein folding model in explicit-solvent molecular dynamics simulations. By varying the solvent mass over more than two orders of magnitude, we alter only the solvent viscosity and not the folding free energy. Folding dynamics at the near-vanishing solvent viscosities accessible by this approach suggest that solvent and internal friction effects are intrinsically entangled. This finding is rationalized by calculation of the polymer end-to-end distance dynamics from a Rouse model that includes internal friction. An analysis of the friction profile along different reaction coordinates suggests a connection between friction and the formation of hydrogen bonds upon folding.

3.1 Introduction

It seems intuitively clear that the kinetics of protein folding and conformational transitions is not dictated by water viscosity alone, but also by internal dissipation processes. In essence, if one could experimentally lower the viscosity of water without changing the folding free energy landscape, one would – for large enough proteins and in the hypothetical limit of vanishing solvent viscosity – still expect folding to be diffusive, but with a diffusivity entirely determined by the internal friction of the protein. Apart from early stopped-flow studies on ribonuclease A [32], the majority of experiments have demonstrated kinetic slowing down with increasing solvent viscosity: In a spectroscopic study on the folding of the α-subunit of tryptophan synthase, the folding rate

was found to scale inversely with increased solvent viscosity, confirming that the rate-limiting step involves a solvent-dominated diffusional process, but the extrapolation to vanishing solvent viscosity resulted in a negligible folding time, therefore no indication of internal friction [33], in agreement with more recent studies on protein L [34]. A study on the conformational relaxation following CO photodissociation in myoglobin on the other hand yielded a substantial viscosity-independent component [13]. For various proteins that essentially fold down-hill in the μs range close to the "speed-limit", the existence of a non-zero internal friction was confirmed [35–37], showing that folding over low barriers is particularly susceptible to internal friction effects [38]; internal friction has only been observed in one case for a larger protein [39]. Likewise, the loop formation dynamics of intrinsically disordered short peptides was found to be dramatically slowed down with added viscosifier [14, 40]. In a particularly revealing study, the kinetics of α-helix and β-hairpin formation in two short model peptides was studied by laser temperature jumps and fluorescence detection [15]: The data were compared with two different laws for the folding time as a function of solvent viscosity, a pure power-law predicting infinitely fast folding at vanishing solvent viscosity, and a linear law with a finite limiting folding time. Both forms essentially described the data equally well, partly reflecting the restricted available data range since the aqueous solvent viscosity cannot be lowered, but only increased by the addition of viscogenic cosolutes. With the exception of very few protein studies [37], the unwanted effects of viscogens on equilibrium properties cannot be excluded, which is particularly disturbing since denaturants (that are commonly added to counteract the stabilization due to viscogens) have also been shown to influence the peptide kinetics [41].

In an effort to assist and interpret these experimental efforts, Langevin simulations of protein folding in implicit solvent were performed [42–44]. As solvent friction is reduced, folding times decrease linearly with viscosity, as expected. But for viscosities orders of magnitude less than water, the functional dependence changes [43] and eventually a turnover is observed, leading to a counterintuitive decrease of folding rate with decreasing viscosity [42, 44]. This turnover is related to suppressed momentum dissipation into the Langevin heat bath at low friction [12] but presumably has no experimental relevance. The main drawback of standard implicit-solvent simulations is that viscous damping, which mimics the solvent, acts in the same way on all residues, regardless of whether they face the solvent or are in the protein interior [45]. The increase of the importance of internal friction as one goes from disordered, solvent-rich conformations to compact ones is therefore not fully accounted for in such models. In explicit-solvent simulations, on the other hand, mass-rescaling provides a simple method for modifying the water viscosity: changing the water mass by a scaling factor c, the inertial force $\mathbf{F}$ following from Newton's equation of motion is invariant when simultaneously rescaling time by a factor $\sqrt{c}$,

$$cm\frac{d^2\mathbf{r}}{d(\sqrt{c}t)^2} = \mathbf{F}, \tag{3.1}$$

meaning that at reduced water mass all transport properties are accelerated and in particular viscosity is reduced. By construction, equilibrium properties and therefore free energies are not influenced: consequently, this trick has previously been used to accelerate the equilibration of protein simulations in explicit solvent [46, 47]. In the present

paper we employ this procedure to study the viscosity-dependent kinetics of short peptides and a two-state protein, changing the solvent mass by more than two orders magnitude and in particular reaching the relevant regime of reduced solvent viscosities. Since all time scales are shifted, small solvent friction is accompanied by high heat conductivity, and therefore Kramer's turnover is not expected, suggesting that our simulation approach more closely captures the essence of experiments at modified solvent viscosity. Our simulations show clearly that internal friction effects exist. The separation into internal and solvent friction is nevertheless not straightforward, even when we locally resolve the friction profile along different reaction coordinates [48, 49] and although in simulations we can substantially reduce the solvent viscosity.

Historically, the concept of internal friction arose for polymers in the context of dissipation due to the thermally activated crossing of dihedral barriers [50]. Shortly after, an alternative mechanism based on contacts between monomers that are not necessarily close neighbors along the backbone was suggested [51]. These contacts can be attractive (hydrogen bonds (HBs) are obvious candidates for this type) or steric and thus purely repulsive, giving rise to considerable kinetic slowing down in polymeric globules [52]. While local friction, e.g., due to dihedral barriers constitutes only a small correction to the long-time chain relaxation, it can be dominant at short times or for strongly stretched biopolymers [53, 54]. Using a simple Rouse model that incorporates local internal friction, we show that the mean passage time for an incremental change of the end-to-end distance depends non-trivially on the solvent viscosity, exhibiting both linear as well as power-law behavior depending on the relative strength of internal and solvent viscosity, but always showing a non-zero folding time at vanishing solvent viscosity. The scaling form provided by the Rouse polymer model fits our simulation data quite well. Although protein kinetics is more complicated, involving also steric and HB effects, our calculation is a first step toward understanding the complex entanglement of internal and solvent friction effects and rationalizes the occurrence of power-laws for folding times as a function of solvent viscosity.

3.2 Simulation Details

Our explicit water molecular dynamics simulations use the GROMACS [17] simulation package version 4.5, Amber ff03 force field [30] and SPC/E water [29] model. The box size is 3.14 nm with 688 water molecules for Ala_8 and 3.18 nm with 711 water molecules for $(GlySer)_4$. Simulation times range up to 4 μs, pressure is set to 1 bar using a Parrinello-Rahman barostat [23], and temperature to 300 K using a velocity rescaling thermostat [22]. The heat bath in our MD simulations couples to the average velocity in the simulation box [22] and therefore is expected to modify the dynamics only negligibly, in contrast to standard Langevin simulations. The different water masses used are between $m/m_0 = 0.01$ and $m/m_0 = 9$, where m_0 denotes the standard water mass, and the integration time step has been adjusted accordingly (for further simulation details, see SI Text).

3.3 Results and Discussion

3.3.1 Simulations at varying water mass

To study the effects of solvent viscosity on peptide dynamics we consider two simple, 8-amino-acid peptides. The first, a $(\mathrm{GlySer})_4$ chain, is unstructured and has been widely studied experimentally [14], the second is the helix forming peptide Ala_8; both peptide termini are capped. By changing the mass of water in the simulations, the water viscosity varies between $\eta/\eta_0 = \sqrt{m/m_0} = 0.1$ and $\eta/\eta_0 = 3$.

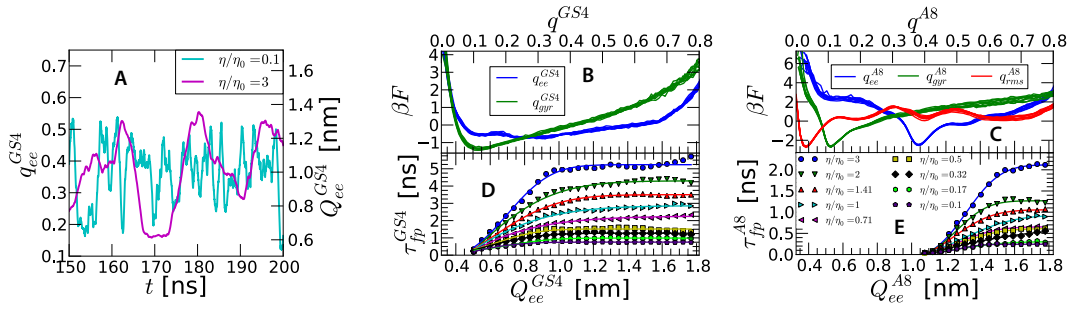

FIGURE 3.1: (A) Trajectories of the end-to-end distance $Q_{\mathrm{ee}}^{\mathrm{GS4}}$ for $(\mathrm{GlySer})_4$ for the lowest and the highest viscosities, $\eta/\eta_0 = 0.1$ and $\eta/\eta_0 = 3$. (B) Free energy $\beta F(Q)$ as a function of the end-to-end radius $Q_{\mathrm{ee}}^{\mathrm{GS4}}$ (blue lines) and radius of gyration $Q_{\mathrm{gyr}}^{\mathrm{GS4}}$ (green lines) for $(\mathrm{GlySer})_4$ for all different viscosities. (C) $\beta F(Q)$ for Ala_8, here in addition results as a function of the root-mean-square (RMS) deviation from the perfect helical state, q_{rms}^{Ala8} (red lines) are shown. (D) Mean-first-passage times $\tau_{\mathrm{fp}}^{\mathrm{GS4}}(Q_{\mathrm{ee}}, Q_{\mathrm{ee}}^f)$ in terms of the end-to-end distance for final position $Q_{\mathrm{ee}}^f = 0.5\mathrm{nm}$. (E) $\tau_{\mathrm{fp}}^{\mathrm{A8}}(Q_{\mathrm{ee}}, Q_{\mathrm{ee}}^f)$ for the final position $Q_{\mathrm{ee}}^f = 1.05\mathrm{nm}$.

In 3.1A we show the fluctuating end-to-end distance $Q_{\mathrm{ee}}^{\mathrm{GS4}}$ (defined as the distance between the terminal amino acids) for $(\mathrm{GlySer})_4$ for the lowest and the highest viscosities, $\eta/\eta_0 = 0.1$ and $\eta/\eta_0 = 3$, illustrating the clear difference in dynamics. In 3.1B we show the free energy $\beta F(Q) = -\ln P(Q)$, which follows from the averaged probability distribution $P(Q)$, as a function of $Q_{\mathrm{ee}}^{\mathrm{GS4}}$ (blue lines) and radius of gyration $Q_{\mathrm{gyr}}^{\mathrm{GS4}}$ (green lines) for $(\mathrm{GlySer})_4$ for all different viscosities. Note that q variables denote rescaled versions of the original Q, mapping the domain between the smallest and largest observed reaction coordinate (RC) values Q onto the interval $q = [0, 1]$, and that $P(q)$ is normalized as $\int_0^1 \mathrm{d}qP(q)$=1. 3.1C displays $\beta F(Q)$ for Ala_8, where in addition to q_{ee}^{Ala8} (blue lines), $q_{\mathrm{gyr}}^{\mathrm{Ala8}}$ (green lines) we show F as a function of the root-mean-square (RMS) deviation from the perfect helical state, $q_{\mathrm{rms}}^{\mathrm{Ala8}}$ (red lines). We note that the free energy profiles for each RC and different water viscosities for both peptides agree closely with each other, thus matching our expectation that solvent viscosity does not influence equilibrium properties and in addition showing that our simulations are well equilibrated. The $(\mathrm{GlySer})_4$ free energy profiles are rather broad and reflect the lack of ordered structure, while the Ala_8 profiles exhibit pronounced minima due to the α-helical state. A quantitative analysis of how solvent viscosity influences the peptide kinetics is possible by computing mean first passage times $\tau_{fp}(Q, Q^f)$, which measure how long it takes, starting from a given position Q along the RC, to reach the final state Q^f for the

first time. In 3.1D we show $\tau_{\mathrm{fp}}^{\mathrm{GS4}}(Q_{\mathrm{ee}}, Q_{\mathrm{ee}}^{f})$ for the end-to-end distance for the final position $Q_{\mathrm{ee}}^{f} = 0.5\mathrm{nm}$, in 3.1E we show $\tau_{\mathrm{fp}}^{\mathrm{A8}}(Q_{\mathrm{ee}}, Q_{\mathrm{ee}}^{f})$ for the final position $Q_{\mathrm{ee}}^{f} = 1.05\mathrm{nm}$. The final positions are in both cases chosen as convenient (i.e. statistically prominent) product states in a prototypical folding reaction starting from $Q_{\mathrm{ee}} > Q_{\mathrm{ee}}^{f}$. As already suggested by the trajectories in (A), the folding becomes slower with increasing viscosity. A well-equilibrated measure for the folding speed is the Boltzmann-averaged mean-first passage time

$$\bar{\tau}_{\mathrm{fp}}(Q^{f}) = \int_{Q^{f}}^{\infty} \mathrm{d}Q P(Q) \tau_{\mathrm{fp}}(Q, Q^{f}) / \int_{Q^{f}}^{\infty} \mathrm{d}Q P(Q). \tag{3.2}$$

If the solvent viscosity η were the only time scale in the folding kinetics, the ratio $\bar{\tau}_{\mathrm{fp}}/(\eta/\eta_0)$ would be independent of η/η_0. The results in 3.2A for the end-to-end distance of the $(\mathrm{GlySer})_4$ (blue) and Ala_8 peptides (red data points) exhibit a pronounced decrease of the rescaled folding time $\bar{\tau}_{fp}/(\eta/\eta_0)$ with η/η_0. Thus, the folding time $\bar{\tau}_{\mathrm{fp}}$ decreases sub-linearly with η/η_0 as $\eta/\eta_0 \to 0$, proving that finite internal friction exists both for the folding of $(\mathrm{GlySer})_4$ as well as Ala_8.

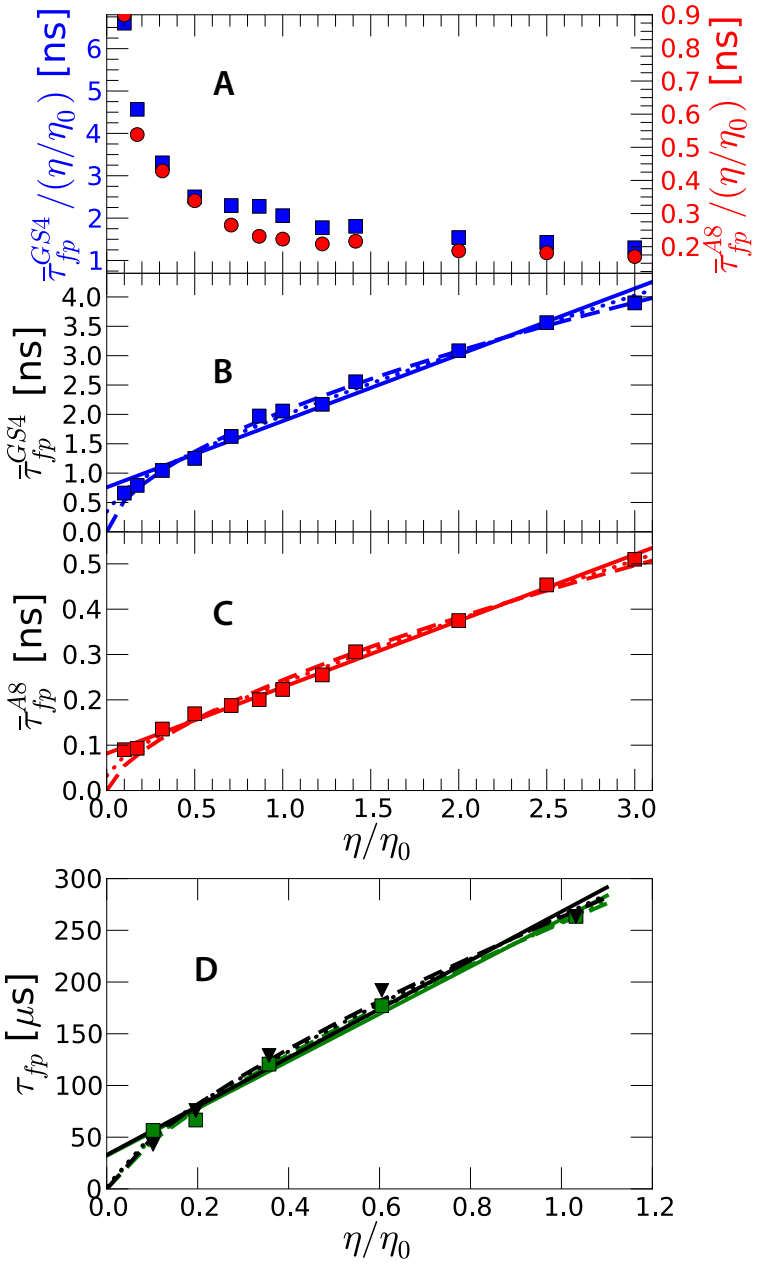

FIGURE 3.2: (A) Boltzmann-averaged mean-first passage time, Eq. 2, rescaled by the solvent viscosity, $\bar{\tau}_{\rm fp}(Q^f)/(\eta/\eta_0)$, for the end-to-end distance of the $(\text{GlySer})_4$ (blue) and Ala_8 peptides (red data points), as a function of η/η_0. (B) Averaged folding time $\bar{\tau}_{\rm fp}$ for $(\text{GlySer})_4$ as a function of the viscosity ratio η/η_0. (C) Analogous results for Ala_8. (D) Results for the averaged folding $\bar{\tau}_{\rm fp}^{\rm f}$ (green) and unfolding time $\bar{\tau}_{\rm fp}^{\rm u}$ (black) for the protein 1prb_{7-53} from a coarse-grained Go-type model. In B-D, fits according to 3.3 (solid), 3.4 (broken), and 3.7 (dotted lines, $N = 10$ in B and C and $N = 47$ in D) are shown. Note that in D, eqs. (3.4) and (3.7) are almost indistinguishable.

Testing the relevance of the peptide dynamics to proteins is challenging because of the computational demands, even for the fastest-folding proteins. We therefore use instead

a coarse-grained Go-type model [55] of the two-state 47-residue 3-helix bundle protein 1prb_{7-53} [56] with explicit solvent (details in SI). Remarkably, we find a very similar non-linear dependence of the folding and unfolding rates on viscosity, revealing a measurable internal friction even in this case. In 3.2B-D we plot the averaged folding time $\bar{\tau}_{\text{fp}}$ for $(\text{GlySer})_4$, Ala_8 and the coarse protein model as a function of the viscosity ratio η/η_0 and compare it with two previously suggested heuristic fitting functions [15]

$$\bar{\tau}_{\text{fp}} = \tau_{\text{int}} + \tau_{\text{wat}}\eta/\eta_0 \tag{3.3}$$

and

$$\bar{\tau}_{\text{fp}} = \tau_0(\eta/\eta_0)^{\alpha}. \tag{3.4}$$

While both forms give fits of essentially equivalent quality, shown as solid and broken lines, only the linear form 3.3 allows an easy separation of the folding time into an internal contribution τ_{int}, which would be the folding time in the hypothetical limit of vanishing solvent viscosity, and the water contribution τ_{wat}, which is the added folding time due to the presence of water. allows no such separation, rather introducing an exponent which turns out to be $\alpha = 0.59$ for $(\text{GlySer})_4$, $\alpha = 0.73$ for 1prb_{7-53}, and $\alpha = 0.66$ for Ala_8, quite close to experimental measurements for an α-helix forming peptide of $\alpha = 0.64$ [15], but leads to the non-intuitive extrapolation that folding occurs infinitely fast as solvent viscosity vanishes. We emphasize that both forms definitely assume that internal friction exists (otherwise we would have $\bar{\tau}_{\text{fp}} = \tau_0\eta/\eta_0$), but none is rigorously derived from a polymeric or protein model. This gap is filled by a Rouse-type calculation for the passage time of an ideal polymer chain with internal friction, results of which are shown as dotted lines in 3.2B-D.

3.3.2 Passage times from Rouse model including internal friction

In the standard Rouse model for polymer dynamics, N connected beads at positions $\mathbf{R}_n$ with $n = 1 \ldots N$ are subject to friction forces $-\xi_m \mathrm{d}\mathbf{R}_n/\mathrm{d}t$ that counteract bead motion proportional to the monomeric Stokes friction coefficient $\xi_m = 6\pi\eta a$, where a is the effective bead radius. The simplest mechanism for internal friction is bond friction that acts on the bond vector $\mathbf{R}_n - \mathbf{R}_{n-1}$ in the form of a force proportional to the bond stretching velocity, $-\xi_b \mathrm{d}(\mathbf{R}_n - \mathbf{R}_{n-1})/\mathrm{d}t$, where ξ_b is the bond friction coefficient that to leading order is independent of the solvent viscosity and only depends on dissipative processes within the polymer (e.g., due to dihedral barriers). Since each monomer participates in two bonds, the Langevin equation of motion takes in the continuum limit the symmetric form [51, 54]

$$\xi_m \mathrm{d}\mathbf{R}(n)/\mathrm{d}t = (\kappa + \xi_b \mathrm{d}/\mathrm{d}t)\mathrm{d}^2\mathbf{R}(n)/\mathrm{d}n^2 + \mathbf{f}(n) \tag{3.5}$$

where $\kappa \sim k_{\text{B}}T/a^2$ is a spring constant ensuring an equilibrium bond length $\sim a$ and $f(n)$ is a Gaussian random force. By normal-mode decomposition, 3.5 can be solved in closed form, and the final result for the auto-correlation function of the mean-squared

end-to-end distance $C_{\rm ee}(t) = \langle(\mathbf{R}_{\rm ee}(t) - \mathbf{R}_{\rm ee}(0))^2\rangle / \langle \mathbf{R}_{\rm ee}^2 \rangle$ with $\mathbf{R}_{\rm ee} = \mathbf{R}(N) - \mathbf{R}(0)$ is

$$C_{\rm ee}(t) = C_0 \sum_p^N \left(1 - {\rm e}^{-t/\tau_p}\right) / p^2. \tag{3.6}$$

Here C_0 is a normalization constant so that $C_{\rm ee}(\infty) = 1$, the sum runs over odd mode numbers p only, and the mode relaxation time is $\tau_p = N^2\tau_m/p^2 + \tau_b$ where we have defined the monomer relaxation time $\tau_m = \xi_m/(\pi^2\kappa)$ and the bond relaxation time $\tau_b = \xi_b/\kappa$. For large times, $C_{\rm ee}(t)$ approaches unity exponentially, $C_{\rm ee}(t) \sim 1 - \exp(-t/(N^2\tau_m + \tau_b))$, where $N^2\tau_m$ is the polymeric or Rouse relaxation time. For intermediate times, in the so-called Rouse regime, $C_{\rm ee}(t)$ grows as a power law, $C_{\rm ee}(t) \sim \sqrt{t/\tau_m}/N$. Most relevant for the present discussion based on the scenario of a protein diffusing in a 1D free energy landscape is the diffusive regime at short times, where we find three separate scaling ranges, depending on the relative strength of monomer and bond time scales: $C_{\rm ee}(t) \sim t/\tau_b$ for dominating internal friction $N^2\tau_m < \tau_b$, $C_{\rm ee}(t) \sim t/(N\sqrt{\tau_b\tau_m})$ for intermediate internal friction $\tau_m < \tau_b < N^2\tau_m$, and $C_{\rm ee}(t) \sim t/(N\tau_m)$ for negligible internal friction $\tau_b < \tau_m$ (see SI Text). To connect to our simulation results for folding times, we define a mean-passage time $\tau_{\rm mp}$ by the condition that $C_{\rm ee}$ has reached a certain threshold value, $C_{\rm ee}(\tau_{\rm mp}) \equiv C_{\rm ee}^*$, with $C_{\rm ee}^*$ chosen small enough so that one stays in the diffusive regime. A scaling function that contains all three regimes and accurately reproduces 3.6 is $\tau_{\rm mp}/C_{ee}^* = C_1\tau_b + C_2 N\sqrt{\tau_b\tau_m} + C_3 N\tau_m$ (see SI Text). The coefficients for $C_{\rm ee}^* = 0.01$ and $N = 10$ are $C_1 = 1$, $C_2 = 0.9$ and $C_3 = 2.1$. Assuming a linear relation between τ_m and η as $\tau_m = \tau_m^0\eta/\eta_0$, where τ_m^0 is defined as the monomer relaxation time at reference solvent viscosity $\eta = \eta_0$, we obtain

$$\tau_{\rm mp}/(C_{\rm ee}^*\tau_b) = C_1 + C_2 N\sqrt{\frac{\tau_m^0}{\tau_b}\frac{\eta}{\eta_0}} + C_3 N\frac{\tau_m^0}{\tau_b}\frac{\eta}{\eta_0}. \tag{3.7}$$

We note: i) 3.7 contains two fitting parameters, an overall time scale, here taken to be the bond relaxation time τ_b, and the ratio τ_m^0/τ_b, thus the number of fitting parameters is the same as in the previously discussed heuristic fitting forms eqs. (3.3) and (3.4). ii) 3.7 combines key aspects of eqs. (3.3) and (3.4), namely the folding time scale reaches a finite value for vanishing η, as in 3.3, and power law behavior for intermediate times is obtained, as in 3.4. The resulting fits according to 3.7 in 3.2B-D (dotted lines) are of the same quality as the other fits. iii) For large chain lengths N, the first term in 3.7 that is independent of solvent viscosity becomes negligible; this is interesting in light of the experimental observation that internal friction has been measured more often for fast-folding (i.e., small) protein or protein domains [35, 36], with only one exception [39].

The main point of our Rouse model calculation is to demonstrate the complexity of chain kinetics when the solvent viscosity is varied, even for the relatively simple case of local internal friction (mimicking dihedral barrier effects) as defined by 3.5. While internal and solvent friction effects are additive for each mode relaxation time τ_p, as assumed in 3.3, the mode mixing that takes place when a kinetic observable is calculated gives rise to power-law behavior, as assumed in 3.4. We now turn to more realistic friction scenarios including interactions and HB effects.

3.3.3 Locally resolved friction analysis

Some of the complexities observed in 3.2B-D could have to do with the fact that a folding time integrates over the friction landscape while the peptide moves from the starting configuration to the final configuration, particularly since the relative weight of internal friction should increase as one goes from open to more collapsed and hydrogen-bonded structures [37]. To look into this, we now resolve the friction profile locally. Assuming that a given RC evolves according to the Fokker-Planck equation [57]

$$\frac{\partial}{\partial t}P(Q,t) = \frac{\partial}{\partial Q}\frac{1}{\beta\xi(Q)}e^{-\beta F(Q)}\frac{\partial}{\partial Q}P(Q,t)e^{\beta F(Q)}, \tag{3.8}$$

where $P(Q,t)$ is the probability of having a configuration with RC value Q at time t, and $\xi(Q)$ is the friction profile. Defining the round-trip time $\tau_{\rm rt}(Q,Q^f) = \text{sign}(Q - Q^f)[\tau_{\rm fp}(Q,Q^f) + \tau_{\rm fp}(Q^f,Q)]$ as the time needed to start at Q, reach Q^f for the first time, start from Q_f again and reach back to Q for the first time, one finds [49] $\tau_{\rm rt}(Q,Q^f) = Z\int_{Q^f}^{Q} dQ'\,\beta\xi(Q')e^{\beta F(Q')}$, where $Z = \int_{Q^{\rm min}}^{Q^{\rm max}} {\rm d}Q e^{-\beta F(Q)}$ is the partition function which in our normalization is unity. The friction profile based on the round-trip time (see SI Text) reads [49]

$$\beta\xi(Q) = \frac{\partial\tau_{\rm rt}(Q,Q^f)/\partial Q}{Ze^{\beta F(Q)}}, \tag{3.9}$$

which in 3.3A is shown for different solvent viscosities for $(\text{GlySer})_4$ in terms of the rescaled end-to-end distance coordinate $q_{\rm ee}$. Indeed, the local friction increases with increasing η, and in addition ξ goes up for smaller $q_{\rm ee}$, i.e., when the peptide chain becomes more confined [37]. When comparing $\xi(q_{\rm ee})$ at normal water viscosity $\eta/\eta_0 = 1$ for $(\text{GlySer})_4$ (blue line) and Ala_8 (red line) in 3.3B, we see that Ala_8 shows a friction maximum at an intermediate value of $q_{\rm ee} \approx 0.4$. This can be compared with the number of intra-peptide HBs $N_{\rm HB}$ in 3.3C. HBs are defined according to the distance-angle criterion that the acceptor - donor - hydrogen angle θ should be smaller than $\theta = 30^o$ and the donor-acceptor distance smaller than 0.35 nm [58].

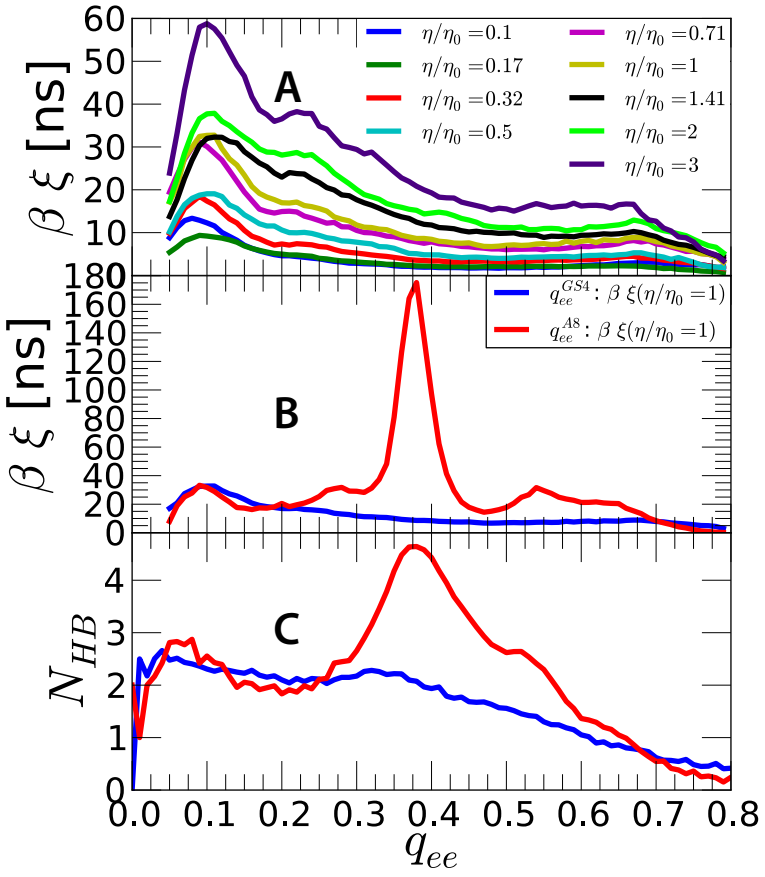

FIGURE 3.3: (A) Locally resolved friction profiles $\beta\xi(q_{ee})$ for different solvent viscosities for $(GlySer)_4$ as a function of the rescaled end-to-end distance q_{ee}. (B) $\beta\xi(q_{ee})$ at normal water viscosity $\eta/\eta_0 = 1$ for $(GlySer)_4$ (blue line) and Ala_8 (red line). (C) Number of intra-peptide HBs N_{HB} for $(GlySer)_4$ (blue line) and Ala_8 (red line).

Indeed, the maximum in ξ at $q_{ee} \approx 0.4$ for Ala_8 roughly matches the maximum in N_{HB}, corresponding to the helical state; the naive expectation (which will be rectified below) would be that the total number of HBs determines local friction.

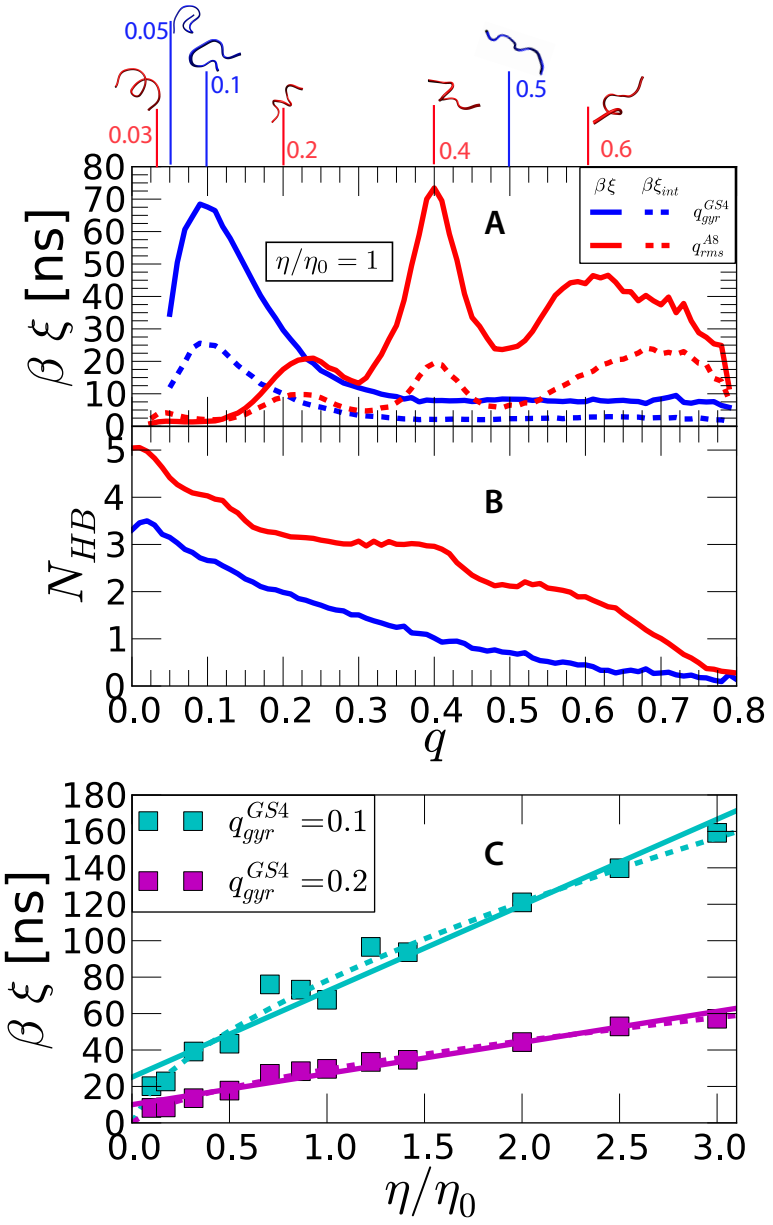

FIGURE 3.4: (A) Friction profiles $\beta\xi$ (solid lines) and internal contribution $\beta\xi_{\mathrm{int}}$ (broken lines) at $\eta/\eta_0 = 1$ for $(\mathrm{GlySer})_4$ as a function of the rescaled radius of gyration q_{gyr} (blue) and for Ala_8 as a function of the RMS from the perfect helical state q_{rms} (red). Representative simulation snapshots are shown at the top. (B) Corresponding number of intra-peptide hydrogen bonds. (C) $\beta\xi^{\mathrm{GS4}}$ for $q_{\mathrm{gyr}} = 0.1$ and $q_{\mathrm{gyr}} = 0.2$ as a function η/η_0. Solid and broken lines denote linear and power-law fits according to eqs. (3.3) and (3.4), respectively.

In fact, the end-to-end distance is not the most natural RC to characterize friction in our two model peptides. Therefore in 3.4 we show $\beta\xi$ as a function of the radius of gyration q_{gyr} for $(\mathrm{GlySer})_4$ and as a function of the RMS from the perfect helical state q_{rms} for Ala_8. For a disordered chain like $(\mathrm{GlySer})_4$, one would expect friction to be mainly due to unspecific intra-peptide HBs and increased in the collapsed state; indeed, ξ plotted versus q_{gyr} shows a pronounced maximum for small q_{gyr} in 3.4A (blue line) that is paralleled by a maximum in N_{HB} in 3.4B (blue line). Unexpectedly, for Ala_8, a

pronounced maximum in ξ appears at $q_{\mathrm{rms}} \approx 0.4$ (red line in 3.4A) and thus shifted away from the fully folded state (around $q_{\mathrm{rms}} \approx 0.1$) where N_{HB} has saturated to the maximum value. Scrutinizing N_{HB} for Ala_8 in 3.4B (red line) more closely, one sees that the maximum in ξ^{A8} around $q_{\mathrm{rms}} \approx 0.4$ correlates roughly with a sudden drop of N_{HB} around $q_{\mathrm{rms}} \approx 0.45$. A possible connection between friction and variations in N_{HB} is suggested by the invariance of the Fokker-Planck equation (3.8) under coordinate rescaling according to $\tilde{Q} = \tilde{Q}(Q)$ if the functions Ψ, F, ξ are simultaneously rescaled as $\tilde{\Psi} = \Psi/\tilde{Q}'$, $\tilde{F} = F + \beta^{-1} \ln \tilde{Q}'$, $\tilde{\xi} = \xi/(\tilde{Q}')^2$, where $\tilde{Q}' \equiv \mathrm{d}\tilde{Q}(Q)/\mathrm{d}Q$. Thus an arbitrary friction profile $\tilde{\xi}(\tilde{Q})$ can be designed, including the limiting simple case of a constant friction $\tilde{\xi}(\tilde{Q}) = \tilde{\xi}_0$, while the kinetics stays invariant. Since all observables (and in particular the HB number N_{HB}) are not modified by the rescaling $\tilde{Q} = \tilde{Q}(Q)$, an explanation of $\xi(Q)$ in terms of equilibrium observables is spurious. Thus, the friction profile can only be linked to derivatives, such as $\mathrm{d}N_{\mathrm{HB}}(Q)/\mathrm{d}Q$, which have similar transformation properties as $\xi(Q)$ itself. Indeed, when comparing ξ^{A8} in 3.4A (red line) with $N_{\mathrm{HB}}^{\mathrm{A8}}$ in 3.4B (red line), we see that maxima in ξ^{A8} approximately correlate with regions where $N_{\mathrm{HB}}^{\mathrm{A8}}$ changes pronouncedly with q_{rms}. A mechanistic interpretation of this would be that HB-related friction is particularly large when the number of HBs significantly *changes* along the RC, i.e., when additional HBs are created, most likely because formation of non-native HBs results in long-lived kinetic traps [39].

Finally, we check whether the local friction profile facilitates the separation into solvent and internal friction. To that end, in 3.4C we plot ξ^{GS4} for $q_{\mathrm{gyr}} = 0.1$ and $q_{\mathrm{gyr}} = 0.2$ as a function η/η_0. Similar to the mean first passage times shown in 3.2B, a linear fit as in 3.3 is possible (solid lines), but the data show clear signs of curvature, and a power-law fit according to 3.4 (broken lines) is equally accurate. Nevertheless, the broken lines in 3.4A show the internal friction profiles $\xi_{\mathrm{int}}(q)$ that are defined via the scaling form $\xi = \xi_{\mathrm{int}} + \xi_{\mathrm{wat}}\eta/\eta_0$ inspired by 3.3. $\xi_{\mathrm{int}}(q)$ largely parallels the total friction profiles $\xi(q)$ (solid lines), but the intuitive expectation that ξ_{int} should dominate in the folded (HB-rich) state while solvent friction, i.e., $\xi - \xi_{\mathrm{int}}$, should dominate in the unfolded (open) state is not borne out by the data. This might have to do with the lack of a rigorous recipe for the division between solvent and internal friction. Alternatively, this could mean that water indirectly also influences HB-induced internal friction, possibly because whenever an intra-peptide HB breaks, a water molecule penetrates and serves as an intermittent HB donor or acceptor.

3.4 Conclusion

Our simulations clearly show that internal friction effects exist but at the same time demonstrate that the quantitative separation into internal and solvent friction is not straightforward. This is true even when we locally resolve the friction profile along different RCs [48, 49] and although in simulations we can substantially reduce the solvent viscosity while making sure that free-energy folding profiles are not modified as viscosity changes. This has primarily to do with the lack of a simple but physically motivated definition of internal friction in terms of folding times or friction profiles. This complexity is corroborated by a simple Rouse-type model for the kinetics of a Gaussian

polymer chain including internal local friction, for which the calculated passage times show power-law and linear behavior as a function of solvent viscosity, depending on the internal friction strength and the polymer size N. It is the polymeric, multi-scale nature of the dynamics that gives rise to the power-law viscosity dependency of time scales in our model calculation, that turns out to be very similar to our simulations and to experimental observations [15]. In fact, an alternative explanation for power-law behavior based on the finite spatial range of spectroscopic probes has been previously given [59] and in reality both mechanisms will be entangled.

3.5 Supplementary Information

3.5.1 Reaction Coordinates

3.5.1.1 Definition

We project the simulation results onto three reaction coordinates: the end-to-end distance Q_{ee}, the radius of gyration Q_{gyr} and in the case of Ala_8 also the root mean-squared deviation (RMSD) from an ideal helix Q_{rms}.

End-to-end distance The end-to-end distance is calculated using the GROMACS tool `g_dist` and is defined as

$$Q_{ee} = |\mathbf{R}(1) - \mathbf{R}(8)|\,, \tag{3.10}$$

where we define $\mathbf{R}(1)$ and $\mathbf{R}(8)$ as the center of mass of the first and the last, i.e. $8th$ amino acid, omitting the endcaps (cf. simulation details).

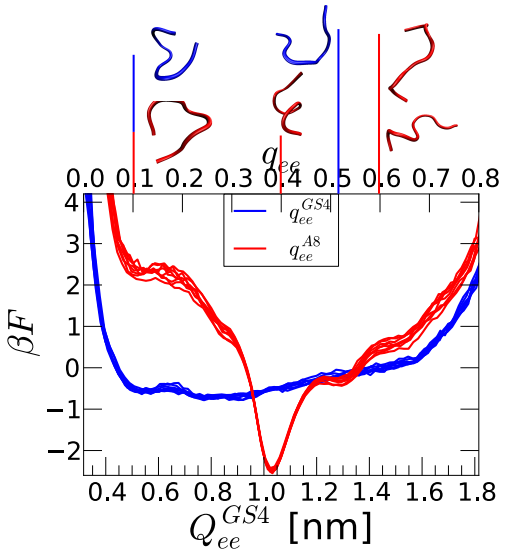

FIGURE 3.5: Free energy as a function of the rescaled end-to-end radius of $(\text{GS})_4$ (blue) and Ala_8 (red). Representative snapshots are shown above.

The corresponding free energy profile together with snapshots is shown in 3.5. In the case of the alanine helix, one can see quite clearly that the end-to-end distance is not a good reaction coordinate, as one point in the reaction coordinate represents very distinct peptide configurations, as visualized by the different snapshots for $q_{\text{ee}} = 0.6$, and thus cannot describe kinetic processes very accurately.

Radius of gyration The radius of gyration is calculated with the GROMACS tool `g_gyrate` and is defined as

$$Q_{\mathrm{gyr}} = \sqrt{\frac{\sum_i |\mathbf{r}_i|^2 \, m_i}{\sum_i m_i}}, \tag{3.11}$$

where m_i is the mass of the ith atom and $\mathbf{r}_i$ the position with respect to the peptide's center of mass.

Root mean squared deviation (RMSD) from ideal helix As explained before, neither the end-to-end distance nor the radius of gyration resolve the dynamics of the alanine peptide well. We thus consider a reaction coordinate that describes the deviation from an ideal helix and thus gives more insight into the helix dynamics. In order to calculate this reaction coordinate, we first define an ideal helix by

$$\mathbf{R}_i^{\mathrm{ideal}} = (r \, \cos(\phi_{\mathrm{tw}} \, i), r \, \sin(\phi_{\mathrm{tw}} \, i), d \, i), \tag{3.12}$$

where i denotes the ith C_α-atom, $r = 0.23\,\mathrm{nm}$ is the helical radius, $d = 0.15\,\mathrm{nm}$ is the rise per residue and $\phi_{\mathrm{tw}} = 1.745\,\mathrm{rad}$ is the twist per residue.

In a second step we minimize the mean-squared deviation between the C_α positions of the simulated helix and the ideal helix with respect to an overall translation and rotation of the simulated helix, using the functional

$$Q_{\mathrm{rms}} = \sqrt{\sum_{i=1}^{8} \left(\mathbf{R}_i^{\mathrm{ideal}} - \mathbf{R}_i^{\mathrm{sim}}\right)^2}, \tag{3.13}$$

where i again gives the ith C_α-atom and $\mathbf{R}_i^{\mathrm{ideal}}$ and $\mathbf{R}_i^{\mathrm{sim}}$ are the positions of the ith atom in the ideal and simulated alanine helix, respectively. The residual value Q_{rms} after optimizing is reported in the main text.

3.5.1.2 Rescaling

As the reaction coordinates produce different scales, we rescale them to enable a meaningful comparison. The reaction coordinates are rescaled to be in the range $[0, 1]$ according to:

$$q = \frac{Q - Q_{\mathrm{min}}}{Q_{\mathrm{max}} - Q_{\mathrm{min}}}, \tag{3.14}$$

where Q is one of the reaction coordinates above and Q_{max} and Q_{min} are the maximum and minimum of the respective reaction coordinate along all mass scalings. Note that the scalings are done for each peptide type individually.

3.5.2 Friction Coefficient and Round Trip Times

Although it is possible to calculate the friction coefficient from first passage times [60], this approach has certain drawbacks. A way to improve statistics is by calculating the *round trip time* for many targets q^f [49]. As we assume a Markovian process, the derivative $\partial\tau_{\mathrm{rt}}(q, q^f)/\partial q$ should not depend on q^f, i.e. the $\tau_{\mathrm{rt}}(q, q^f)$ are equal up to a constant. We can therefore average the obtained derivatives and thus drastically improve the statistics of the resulting friction profiles.

3.5.2.1 Definition

As already explained in the main text, we define the round-trip time as the time to reach q^f from q for the first time and then return to q

$$\tau_{\mathrm{rt}}(q, q^f) = \mathrm{sign}(q - q^f)[\tau_{\mathrm{fp}}(q, q^f) + \tau_{\mathrm{fp}}(q^f, q)] \tag{3.15}$$

In the top part of 3.6 we show round-trip times for different setups. The vertical coloured bars indicate the respective q^f. One can see that indeed different curves only differ by a constant, at least for a certain region around q^f. The solid lines show spline fits explained in the next section.

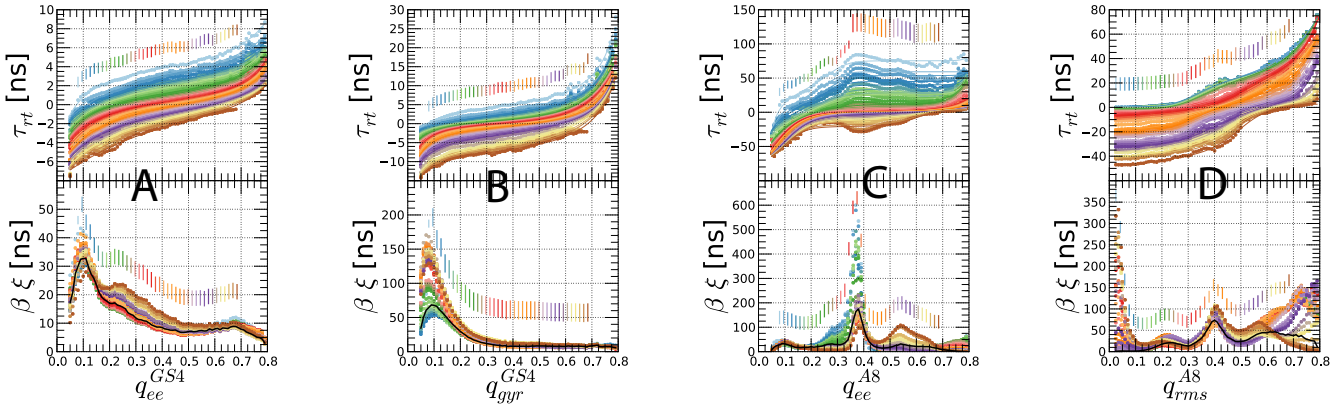

FIGURE 3.6: Round-trip times τ_{rt} and friction coefficient ξ calculated from round-trip time fits. Colours indicate different q^f (indicated by vertical bars). Solid black lines give the weighted average over different friction profiles as discussed in the text.
A) $q_{\mathrm{ee}}^{\mathrm{GS4}}(\eta/\eta_0 = 1)$ B) $q_{\mathrm{gyr}}^{\mathrm{GS4}}(\eta/\eta_0 = 1)$ C) $q_{\mathrm{ee}}^{\mathrm{A8}}(\eta/\eta_0 = 1)$ D) $q_{\mathrm{rms}}^{\mathrm{A8}}(\eta/\eta_0 = 1)$

3.5.2.2 Spline Fits

As a first step towards the calculation of the friction coefficient ξ we fit our round-trip times with a monotonic spline

$$\tau_{\text{rt, fit}}(q, q^f) = \tau_{\text{rt, fit}}(q_{\min}, q^f) + \int_{q_{\min}}^{q} \mathrm{d}q' \mathrm{e}^{W(q')}, \tag{3.16}$$

where $W(q')$ is an arbitrary function, which, as $\mathrm{e}^{W(q')} > 0$, via the integral ensures a monotonic rising function. $W(q')$ is then fitted by a base of $n_{\text{base}} = 79$ quintic B-splines. The number of base function n_{base} is chosen by $n_{\text{base}} = n_{\text{data}} + n_{\text{order}} - 2$ with $n_{\text{data}} = 75$ the number of data points and $n_{\text{order}} = 6$ the order of the quintic spline. The actual fitting is carried out using a standard least-squared scheme with the addition of a smooting term $\lambda \int_{q_{\min}}^{q_{\max}} \mathrm{d}q' (\partial W(q')/\partial q')^2$, which penalized roughness. This smoothing term is added to the sum of squared residues and then the combined sum is minimized. λ is chosen in such a way, that statistical noise is removed yet the curve is not overfitted, i.e. removing actual features in the curve by smoothing. Typical values were in the range of $\lambda = 10^{-6}\,\text{ns}^2$. The fitting routine is carried out using the Functional Data Analysis (FDA) package [61] of the `R` statistical software.

3.5.2.3 Calculation of ξ

As explained in the main text, we calculate the friction coefficient by

$$\beta\xi(q) = \frac{\partial\tau_{\text{rt}}(q, q^f)/\partial q}{Ze^{\beta F(q)}}. \tag{3.17}$$

The derivative of the round-trip time is straightforwardly calculated from the spline fits. As explained before, in order to improve statistics, we calculate τ_{rt} for several q^f and average the resulting derivative. Due to statistics, the round-trip time is most accurate in the neighbourhood of q^f and thus only derivatives in the finite range $q^f - q_0 < q < q^f + q_0$ contribute to the average, where we chose $q_0 = 0.15$. The corresponding friction profiles and the resulting average (solid black line) can be seen in 3.6 for several setups.

3.5.3 Derivation of Rouse Model

3.5.3.1 Model Description

We consider the usual Rouse model consisting of discrete beads interconnected by harmonic springs of stiffness κ. To include internal friction we add a force term opposing extensional motion of the bonds, namely $-\xi_b \mathrm{d}\left(\mathbf{R}_n - \mathbf{R}_{n-1}\right)/\mathrm{d}t$ [50, 53]. Therefore the total force on the nth bead due to internal friction reads:

$$\xi_b \mathrm{d}\left(\mathbf{R}_{n-1} - 2\mathbf{R}_n + \mathbf{R}_{n+1}\right)/\mathrm{d}t \tag{3.18}$$

Thus the dynamics of the nth bead are governed by the Langevin-equation

$$\xi_m \frac{\mathrm{d}\mathbf{R}_n}{\mathrm{d}t} = \left(\kappa + \xi_b \frac{\mathrm{d}}{\mathrm{d}t}\right)(\mathbf{R}_{n-1} - 2\mathbf{R}_n + \mathbf{R}_{n+1}) + \mathbf{f}_n \,, \tag{3.19}$$

where ξ_m is the usual Stokes friction and $\mathbf{f}_n$ is the stochastic contribution.
If we consider large N and investigate the dynamics of the chain only at large length scales, we can make a continuum approximation of 3.19, yielding

$$\xi_m \frac{\partial \mathbf{R}(n,t)}{\partial t} = \left(\kappa + \xi_b \frac{\partial}{\partial t}\right) \frac{\partial^2 \mathbf{R}(n,t)}{\partial n^2} + \mathbf{f}(n,t) \,, \tag{3.20}$$

where now the chain is described by a continuous space curve $\mathbf{R}(n,t)$. Under the assumption of free-end boundary conditions,

$$\frac{\partial \mathbf{R}(n,t)}{\partial n}\Big|_{n=0} = \frac{\partial \mathbf{R}(n,t)}{\partial n}\Big|_{n=N} = 0 \,, \tag{3.21}$$

3.20 has spatial eigenfunctions proportional to $\cos(p\pi n/N)$ with mode number p. Assuming the same boundary conditions for the stochastic forces $\mathbf{f}(n,t)$ as for $\mathbf{R}(n,t)$, we can expand both in terms of the cosine eigenfunctions by

$$\mathbf{R}(n,t) = \mathbf{X}_0(t) + 2\sum_{p=1}^{\infty} \mathbf{X}_p(t)\cos(p\pi n/N) \tag{3.22}$$

$$\mathbf{f}(n,t) = \frac{1}{N}\mathbf{f}_0(t) + \frac{1}{N}\sum_{p=1}^{\infty} \mathbf{f}_p(t)\cos(p\pi n/N) \,, \tag{3.23}$$

where the prefactors 2 and $1/N$ are included for convenience. These expansions can be inserted into the Langevin-equation 3.20, multiplied by $\cos(p'\pi n/N)$ and integrated over n,

$$\begin{aligned}
&\int_0^N \mathrm{d}n \cos(p'\pi\frac{n}{N})\xi_m \left(\frac{\partial \mathbf{X}_0(t)}{\partial t} + 2\sum_{p=1}^{\infty}\frac{\partial \mathbf{X}_p(t)}{\partial t}\cos(p\pi\frac{n}{N})\right) \\
&= \int_0^N \mathrm{d}n \cos(p'\pi n/N)\times \\
&\quad \times \left[\left(\kappa + \xi_b\frac{\partial}{\partial t}\right) 2\sum_{p=1}^{\infty}\mathbf{X}_p(t)\left(-\frac{p^2\pi^2}{N^2}\cos(p\pi n/N)\right)\right. \\
&\quad \left. + \left(\frac{1}{N}\mathbf{f}_0(t) + \frac{1}{N}\sum_{p=1}^{\infty}\mathbf{f}_p(t)\cos(p\pi n/N)\right)\right] .
\end{aligned} \tag{3.24}$$

Since

$$\int_0^N \mathrm{d}n \cos(p'\pi\frac{n}{N})\cos(p\pi\frac{n}{N}) = \delta_{p0}\delta_{pp'}N + \left(1-\delta_{p0}\right)\delta_{pp'}\frac{N}{2} \tag{3.25}$$

one obtains

$$\xi_p \frac{\mathrm{d}\mathbf{X}_p(t)}{\mathrm{d}t} = -\kappa_p \mathbf{X}_p(t) + \mathbf{f}_p(t)\,, \tag{3.26}$$

where

$$\kappa_p = \frac{2\pi^2 p^2 \kappa}{N} \quad \text{for } p \geq 0\,, \tag{3.27}$$

$$\begin{aligned} \xi_0 &= N\xi_m\,, \\ \xi_p &= 2N\xi_m + \frac{2\pi^2 p^2 \xi_b}{N} \quad \text{for } p > 0\,. \end{aligned} \tag{3.28}$$

With eqs. (3.27) and (3.28) we can calculate the mode relaxation times τ_p, yielding

$$\tau_p = \frac{\xi_p}{\kappa_p} = \frac{\tau_R}{p^2} + \tau_b\,, \tag{3.29}$$

where the Rouse time τ_R for a polymer with free ends is defined as

$$\tau_R = \frac{N^2\xi_m}{\pi^2\kappa} \tag{3.30}$$

and

$$\tau_b = \frac{\xi_b}{\kappa} \tag{3.31}$$

is the contribution of the internal friction to the mode relaxation times.

3.5.3.2 Noise Correlations

The inclusion of internal friction alters the noise correlation in real space with respect to the case without internal friction, where the correlations are known and trivial. In fact, internal friction leads to spatially correlated noise [53], meaning that the forces are no longer δ-correlated in space. Calculation of the correlations with incorporated internal friction in real space is rather involved, but if the dynamics are considered in terms of normal mode amplitudes as in 3.26, the calculation is feasible. To calculate the

correlation of the random normal mode amplitudes for a general space-correlation of the real space noise forces, we consider

$$\left\langle \mathbf{f}(n,t)\mathbf{f}^T(m,t')\right\rangle = 2k_\mathrm{B}T\mathbf{1}B(n,m)\delta(t-t')\,, \tag{3.32}$$

with some arbitrary function $B(n,m)$. Since we assume that the random forces $\mathbf{f}(n,t)$ can be expanded in a cosine basis, this should also hold for $B(n,m)$. Hence we can write

$$\begin{aligned} B(n,m) =& \frac{1}{N^2}B_{00} + \frac{2}{N^2}\sum_{k\geq 1}^{\infty} B_{k0}\cos\left(k\pi n/N\right) \\ &+ \frac{2}{N^2}\sum_{l\geq 1}^{\infty} B_{0l}\cos\left(l\pi m/N\right) \\ &+ \frac{4}{N^2}\sum_{k,l\geq 1}^{\infty} B_{kl}\cos\left(k\pi n/N\right)\cos\left(l\pi m/N\right)\,. \end{aligned} \tag{3.33}$$

Inserting this into 3.32 and multiplying this equation with $\cos\left(p'\pi n/N\right)\cos\left(q'\pi m/N\right)$, one obtains by integrating over n and m for the right hand side

$$\begin{aligned} 2k_\mathrm{B}T\mathbf{1}\delta(t-t')\int_0^N \mathrm{d}n\,\mathrm{d}m\,\cos\left(p'\pi\frac{n}{N}\right)\cos\left(q'\pi\frac{m}{N}\right)B(n,m) \\ = 2k_\mathrm{B}T\mathbf{1}\delta(t-t')B_{p'q'}\,. \end{aligned} \tag{3.34}$$

Using the expansion into normal modes, 3.23, the left hand side of 3.32 evaluates to

$$\begin{aligned} &\int_0^N \mathrm{d}n\,\mathrm{d}m\,\left\langle \mathbf{f}(n,t)\cos\left(p'\pi n/N\right)\mathbf{f}^T(m,t)\cos\left(q'\pi m/N\right)\right\rangle = \\ &\left\langle \mathbf{f}_{p'}(t)\left(\delta_{p'0}+\frac{1}{2}\left(1-\delta_{p'0}\right)\right)\mathbf{f}^T_{q'}(t)\left(\delta_{q'0}+\frac{1}{2}\left(1-\delta_{q'0}\right)\right)\right\rangle. \end{aligned} \tag{3.35}$$

3.32 can thus be rewritten as

$$\left\langle \mathbf{f}_p(t)\mathbf{f}_q(t')\right\rangle = \mathbf{1}\frac{2k_\mathrm{B}T\delta(t-t')B_{pq}}{\left(\delta_{p0}+\frac{1}{2}\left(1-\delta_{p0}\right)\right)\left(\delta_{q0}+\frac{1}{2}\left(1-\delta_{q0}\right)\right)}\,. \tag{3.36}$$

With the solution of the Langevin equation 3.26,

$$\mathbf{X}_p(t) = \mathbf{X}_p(t_0)e^{-(t-t_0)/\tau_p} + \int_{t_0}^{t}\mathrm{d}t'\,e^{-(t-t')/\tau_p}\,\frac{\mathbf{f}_p(t')}{\xi_p}\,, \tag{3.37}$$

one obtains in the limit $t_0 \to -\infty$ using 3.36

$$\left\langle \mathbf{X}_p(t)\mathbf{X}_q(t+\tau)\right\rangle = \frac{e^{-\tau/\tau_q}}{\xi_p\xi_q\left(1/\tau_p+1/\tau_q\right)}\mathbf{1} \times \frac{2k_\mathrm{B}TB_{pq}}{\left(\delta_{p0}+\frac{1}{2}\left(1-\delta_{p0}\right)\right)\left(\delta_{q0}+\frac{1}{2}\left(1-\delta_{q0}\right)\right)} . \tag{3.38}$$

For $p \geq 1$ the Langevin equation 3.26 of the normal mode amplitudes describes a particle in a harmonic potential with stiffness κ_p in each dimension. The corresponding Hamiltonian of this system is given by $\mathcal{H} = \kappa_p \mathbf{X}_p^2/2$. Using the standard Boltzmann probability density, the second moment $\left\langle \mathbf{X}_p^2\right\rangle_{eq}$ in equilibrium can simply be calculated via

$$\left\langle \mathbf{X}_p^2\right\rangle_{\mathrm{eq}} = \frac{\int_{-\infty}^{\infty}\mathbf{X}_p^2 e^{-\beta\frac{\kappa_p}{2}\mathbf{X}_p^2}\mathrm{d}^3X_p}{\int_{-\infty}^{\infty}e^{-\beta\frac{\kappa_p}{2}\mathbf{X}_p^2}\mathrm{d}^3X_p} . \tag{3.39}$$

One obtains

$$\left\langle \mathbf{X}_p^2\right\rangle_{\mathrm{eq}} = 3\frac{k_\mathrm{B}T}{\kappa_p} . \tag{3.40}$$

Since the Hamiltonian $\mathcal{H}$ does not include any cross terms connecting mode amplitudes with different mode numbers, one can conclude that

$$\left\langle \mathbf{X}_p\mathbf{X}_q^T\right\rangle_{\mathrm{eq}} = \mathbf{1}\frac{k_\mathrm{B}T}{\kappa_p}\delta_{pq} . \tag{3.41}$$

This constitutes the limit $\tau = 0$ of 3.38 and therefore B_{pq} has to be diagonal, $B_{pq} = B_p\delta_{pq}$ and furthermore

$$\tilde{B}_p \equiv \frac{B_p}{\left(\delta_{p0}+\frac{1}{4}\left(1-\delta_{p0}\right)\right)} = \xi_p . \tag{3.42}$$

Thus we have derived the correlation of the noise forces for mode numbers $p \geq 1$. Since the $p = 0$ mode describes the diffusion of the whole polymer,

$$\left\langle \mathbf{X}_0^2(t)\right\rangle = 6Dt + \left\langle \mathbf{X}_0^2(0)\right\rangle \tag{3.43}$$

has to be fulfilled, where $D = k_\mathrm{B}T/N\xi_m$ is the diffusion constant. Via the solution of the Langevin equation, 3.37, one obtains

$$\left\langle \mathbf{X}_0^2(t)\right\rangle = 6\,\frac{k_\mathrm{B}T}{\xi_0}\,t\,\frac{\tilde{B}_0}{\xi_0} + \left\langle \mathbf{X}_0^2(0)\right\rangle , \tag{3.44}$$

Hence $\tilde{B}_0 \stackrel{!}{=} \xi_0$ and therefore

$$\tilde{B}_p = \xi_p \quad \forall p \tag{3.45}$$

yielding

$$\left\langle \mathbf{f}_p(t)\mathbf{f}_q(t') \right\rangle = 2k_\mathrm{B}T\mathbf{1}\delta(t-t')\delta_{pq}\xi_p \; . \tag{3.46}$$

3.5.3.3 End-to-End Mean-Squared Displacement

The end-to-end vector of the chain in terms of mode amplitudes is given by

$$\mathbf{R}_\mathrm{ee}(t) \equiv \mathbf{R}(N,t) - \mathbf{R}(0,t) = -4 \sum_{p,\mathrm{odd}}^{\infty} \mathbf{X}_p(t) \; . \tag{3.47}$$

The end-to-end MSD is then given by

$$\begin{aligned} MSD_\mathrm{ee}(t) &= \left\langle \left(\mathbf{R}_\mathrm{ee}(t) - \mathbf{R}_\mathrm{ee}(0)\right)^2 \right\rangle \\ &= \left\langle \left(4 \sum_{p,\mathrm{odd}}^{\infty} \left(\mathbf{X}_p(0) - \mathbf{X}_p(t)\right) \right)^2 \right\rangle \\ &= 16 \sum_{p,q,\mathrm{odd}}^{\infty} \Big(\left\langle \mathbf{X}_p(0)\mathbf{X}_q(0) \right\rangle - \left\langle \mathbf{X}_p(0)\mathbf{X}_q(t) \right\rangle \\ &\qquad - \left\langle \mathbf{X}_p(t)\mathbf{X}_q(0) \right\rangle + \left\langle \mathbf{X}_p(t)\mathbf{X}_q(t) \right\rangle \Big) \; . \end{aligned} \tag{3.48}$$

To calculate this we need the correlation functions $\left\langle \mathbf{X}_p(t)\mathbf{X}_q(0) \right\rangle$ for $p > 0$. These can simply be evaluated by making use of the solution of the Langevin equation as given in 3.37:

$$\begin{aligned} \left\langle \mathbf{X}_p(t)\mathbf{X}_p(0) \right\rangle &= \left\langle \mathbf{X}_p^2(0) \right\rangle e^{-t/\tau_p} \\ &\qquad + \left\langle \mathbf{X}_p(0) \int_0^t \mathrm{d}t' \; e^{-(t-t')/\tau_p} \frac{\mathbf{f}_p(t')}{\xi_p} \right\rangle \\ &= \left\langle \mathbf{X}_p^2(0) \right\rangle e^{-t/\tau_p} \; . \end{aligned} \tag{3.49}$$

With 3.40 one obtains

$$\left\langle \mathbf{X}_p(t)\mathbf{X}_q(0) \right\rangle = 3\delta_{pq} \frac{k_\mathrm{B}T}{\kappa_p} e^{-t/\tau_p} \quad \text{for } p \geq 1 \; . \tag{3.50}$$

Finally one finds for the end-to-end MSD:

$$MSD_{ee}(t) = 96 \sum_{p,\text{odd}}^{\infty} \frac{k_B T}{\kappa_p} \left(1 - e^{-t/\tau_p}\right) . \tag{3.51}$$

3.5.3.4 Scaling Results for the MSD

The MSD of the end-to-end vector as given in 3.51 has a limiting value for $t \to \infty$

$$\begin{aligned} MSD_{\text{ee}}(t \to \infty) &= 96 \sum_{p,\text{odd}}^{\infty} \frac{k_B T}{\kappa_p} = 96 \frac{k_B T N}{2\pi^2 \kappa} \sum_{p,\text{odd}}^{\infty} \frac{1}{p^2} \\ &= \frac{6 k_B T N}{\kappa} , \end{aligned} \tag{3.52}$$

since $\sum_{p,\text{odd}}^{\infty} \frac{1}{p^2} = \pi^2/8$. Therefore the normalized version of the MSD, $C_{\text{ee}}(t) = \langle (\mathbf{R}_{\text{ee}}(t) - \mathbf{R}_{\text{ee}}(0))^2 \rangle / \langle \mathbf{R}_{\text{ee}}^2 \rangle$, reads

$$C_{ee}(t) = C_0 \sum_{p,\text{odd}}^{\infty} \frac{1}{p^2} \left(1 - e^{-t/\tau_p}\right) , \tag{3.53}$$

where $C_0 = 1/\sum_{p,\text{odd}}^{\infty} \frac{1}{p^2}$. Now we define the monomer relaxation time

$$\tau_m = \frac{\zeta_m}{\pi^2 \kappa} , \tag{3.54}$$

and with this write

$$C_{ee}(t) = C_0 \sum_{p,\text{odd}}^{\infty} \frac{1}{p^2} \left(1 - \exp\left(-\frac{t}{N^2 \tau_m / p^2 + \tau_b}\right)\right) . \tag{3.55}$$

In order to take into account the discreteness of the peptide chain, we introduce a mode number cutoff N. Therefore the following summations in the calculation of $C_{\text{ee}}(t)$ will be terminated at $p = N$.

The function $C_{\text{ee}}(t)$ as given in 3.55 has distinct scaling properties in the various time regimes:

3.5.4 Long Time Limit

In the long time limit characterized by $t > \tau_R + \tau_b$, the sum in 3.55 is dominated by its first term. Therefore we approximate

$$C_{\mathrm{ee}}(t) \sim 1 - e^{-t/(\tau_R+\tau_b)} \,. \tag{3.56}$$

This is a simple exponential decay to the limiting value of $C(t \to \infty) = 1$.

Short Time Limit For small t, $t < \tau_R/N^2 + \tau_b$, we can expand the exponential in 3.55 yielding

$$C_{\mathrm{ee}}(t) \sim \sum_{p=1,\,\mathrm{odd}}^{N} \frac{1}{p^2} \frac{t}{\tau_R/p^2 + \tau_b} = \frac{t}{\tau_R} \sum_{p=1,\,\mathrm{odd}}^{N} \frac{1}{1 + \frac{\tau_b}{\tau_R} p^2} \tag{3.57}$$

and distinguish three different cases:

1. $\frac{\tau_b}{\tau_R} N^2 < 1$: For $\tau_b \ll \tau_R$ the sum can be approximated by an integral,

$$C_{\mathrm{ee}}(t) \sim \frac{t}{\tau_R} \int_0^N \mathrm{d}p \, \frac{1}{1 + \frac{\tau_b}{\tau_R} p^2} \,. \tag{3.58}$$

   Since $\tau_b N^2/\tau_R < 1$, we can approximate the denominator of the integrand by 1, which yields

$$C_{ee}(t) \sim \frac{N}{\tau_R} t \,. \tag{3.59}$$

2. $\frac{\tau_b}{\tau_R} < 1 < \frac{\tau_b}{\tau_R} N^2$: Up to a mode number $\hat{p}$, defined by

$$\frac{\tau_b}{\tau_R} \hat{p}^2 = 1 \quad \to \quad \hat{p} = \sqrt{\frac{\tau_R}{\tau_b}} \,, \tag{3.60}$$

   the denominator in the integrand again can be replaced by 1. For $p > \hat{p}$ we neglect the 1 in the denominator. Thus one obtains

$$\begin{aligned} C_{\mathrm{ee}}(t) &\sim \frac{t}{\tau_R} \left(\left(\frac{\tau_R}{\tau_b} \right)^{1/2} + \int_{\hat{p}}^{N} \mathrm{d}p \, \frac{1}{\frac{\tau_b}{\tau_R} p^2} \right) \\ &= \frac{t}{\tau_R} \left(\left(\frac{\tau_R}{\tau_b} \right)^{1/2} + \left(\frac{\tau_R}{\tau_b} \right)^{1/2} - \frac{\tau_R}{\tau_b N} \right) \,. \end{aligned} \tag{3.61}$$

   This yields:

$$C_{\mathrm{ee}}(t) \sim \frac{2}{\sqrt{\tau_R \tau_b}} t \,. \tag{3.62}$$

3. $\tau_b > \tau_R$: We neglect the 1 in the denominator in 3.57 and therefore approximate $C_{\mathrm{ee}}(t)$ by

$$C_{\mathrm{ee}}(t) \sim \frac{t}{\tau_b} \sum_{p=1,\,\mathrm{odd}}^{N} \frac{1}{p^2} \approx \frac{t}{\tau_b}\left(1 - \frac{1}{N}\right) . \tag{3.63}$$

Intermediate Times For $\tau_R/N^2 + \tau_b < t < \tau_R + \tau_b$ we approximate the sum in 3.55 by an integral from 1 to N,

$$C_{\mathrm{ee}}(t) \sim \int_1^N \mathrm{d}p\, \frac{1}{p^2}\left(1 - e^{-t/\left(\tau_R/p^2 + \tau_b\right)}\right) . \tag{3.64}$$

Now we substitute

$$y = \frac{t}{\tau_R/p^2 + \tau_b} , \tag{3.65}$$

yielding

$$C_{\mathrm{ee}}(t) \sim \int_{y_1}^{y_2} \mathrm{d}y\, \frac{p}{2\tau_R}\frac{t}{y^2}\left(1 - e^{-y}\right) , \tag{3.66}$$

where

$$\begin{aligned} y1 &= \frac{t}{\tau_R + \tau_b} < 1 , \\ y2 &= \frac{t}{\tau_R/N^2 + \tau_b} > 1 . \end{aligned} \tag{3.67}$$

Using $p = \left(\tau_R/(t/y - \tau_b)\right)^{1/2}$ one obtains

$$\begin{aligned} C_{\mathrm{ee}}(t) &\sim \int_{y_1}^{y_2} \mathrm{d}y\, \left(\frac{\tau_R}{t/y - \tau_b}\right)^{1/2} \frac{t}{2\tau_R y^2}\left(1 - e^{-y}\right) \\ &= \frac{t^{1/2}}{2\tau_R^{1/2}} \int_{y_1}^{y_2} \mathrm{d}y\, \frac{\left(1 - \tau_b y/t\right)^{-1/2}}{y^{3/2}}\left(1 - e^{-y}\right) . \end{aligned} \tag{3.68}$$

To proceed further we split the integration according to

$$C_{\mathrm{ee}}(t) \sim \frac{t^{1/2}}{2\tau_R^{1/2}} \left(\underbrace{\int_{y_1}^{1} \mathrm{d}y \, \frac{(1-\tau_b y/t)^{-1/2}}{y^{3/2}} \left(1-e^{-y}\right)}_{I_1} + \underbrace{\int_{1}^{y_2} \mathrm{d}y \, \frac{(1-\tau_b y/t)^{-1/2}}{y^{3/2}} \left(1-e^{-y}\right)}_{I_2} \right) . \tag{3.69}$$

Since $\tau_b/t < 1$ and $y_1 < 1$ we can expand the square-root and the exponential in I_1, yielding

$$\begin{aligned} I_1 &\simeq \int_{y_1}^{1} \mathrm{d}y \, y^{-3/2} \left(1 + \frac{1}{2}\frac{\tau_b y}{t} + \mathcal{O}(y^2)\right) \left(y - \frac{y^2}{2} + \dots\right) \\ &= \int_{y_1}^{1} \mathrm{d}y \, y^{-1/2} \left(1 + \frac{1}{2}\frac{\tau_b y}{t} + \mathcal{O}(y^2)\right) \left(1 - \frac{y}{2} + \dots\right) \\ &= 2 - 2y_1^{1/2} + \frac{\tau_b}{2t}\frac{2}{3}\left(1 - y_1^{3/2}\right) - \frac{1}{3}\left(1 - y_1^{3/2}\right) \\ &\simeq 2 - 2y_1^{1/2} \, . \end{aligned} \tag{3.70}$$

To evaluate the second integral I_2 we first of all split again the integration

$$I_2 = \underbrace{\int_{1}^{y_2} \mathrm{d}y \, \frac{(1-\tau_b y/t)^{-1/2}}{y^{3/2}}}_{I_3} - \underbrace{\int_{1}^{y_2} \mathrm{d}y \, \frac{(1-\tau_b y/t)^{-1/2}}{y^{3/2}} e^{-y}}_{I_4} \, . \tag{3.71}$$

In I_3 we substitute $z = \tau_b y/t$ resulting in

$$I_3 = \left(\frac{\tau_b}{t}\right)^{1/2} \int_{z_1}^{z_2} \mathrm{d}z \, z^{-3/2} \, (1-z)^{-1/2} \, , \tag{3.72}$$

where

$$\begin{aligned} z_2 &= \frac{\tau_b}{t} y_2 = \frac{\tau_b}{\tau_b + \tau_R/N^2} < 1 \, , \\ z_1 &= \frac{\tau_b}{t} \ll z_2 < 1 \, . \end{aligned} \tag{3.73}$$

Thus we can expand the square-root and obtain

$$
\begin{aligned}
I_3 &\simeq \left(\frac{\tau_b}{t}\right)^{1/2} \int_{z_1}^{z_2} \mathrm{d}z\, z^{-3/2} \left(1 + \frac{1}{2}z + \dots\right) \\
&= \left(\frac{\tau_b}{t}\right)^{1/2} \left[2\left(z_1^{-1/2} - z_2^{-1/2}\right) + \left(z_2^{1/2} - z_1^{1/2}\right) + \dots\right] \\
&= const. - 2\left(\frac{\tau_b}{t}\right)^{1/2} \left(\frac{\tau_b + \tau_R/N^2}{\tau_b}\right)^{1/2} + \dots \,. \qquad (3.74)
\end{aligned}
$$

The last integral I_4 is evaluated by partial integration

$$
\begin{aligned}
I_4 &= \int_1^{y_2} \mathrm{d}y\, \frac{(1 - \tau_b y/t)^{-1/2}}{y^{3/2}} e^{-y} \\
&= -y^{-3/2} (1 - \tau_b y/t)^{-1/2} e^{-y} \Big|_1^{y_2} \\
&\quad - \int_1^{y_2} \mathrm{d}y \left(\frac{3}{2} y^{-5/2} (1 - \tau_b y/t)^{-1/2} \right. \\
&\qquad\qquad \left. - \frac{1}{2}\frac{\tau_b}{t} y^{-3/2} (1 - \tau_b y/t)^{-3/2}\right) e^{-y} \\
&\simeq (1 - \tau_b/t)^{-1/2} e^{-1} + \mathcal{O}(e^{-y}) \,. \qquad (3.75)
\end{aligned}
$$

Hence we obtain for $C_{\mathrm{ee}}(t)$

$$
C_{\mathrm{ee}}(t) \simeq \frac{t^{1/2}}{2\tau_R^{1/2}} \left(const. + y_1^{1/2} + y_2^{-1/2} + \dots\right) \,. \qquad (3.76)
$$

Combining all results for the different time regimes we have the complete scaling behaviour.

Summary of Scaling Behaviour

1. Negligible internal friction $\frac{\tau_b}{\tau_R} N^2 = \frac{\tau_b}{\tau_m} < 1$:

$$
C_{\mathrm{ee}}(t) \sim \begin{cases} \frac{Nt}{\tau_R} = \frac{t}{N\tau_m} & \text{for } t < \tau_R/N^2 = \tau_m \\ \left(\frac{t}{\tau_R}\right)^{1/2} & \text{for } \tau_R/N^2 < t < \tau_R \\ 1 - \exp\left(-\frac{t}{\tau_R}\right) & \text{for } t > \tau_R \end{cases} \qquad (3.77)
$$

2. Intermediate internal friction $\frac{\tau_b}{\tau_R} < 1 < \frac{\tau_b}{\tau_R} N^2$:

$$C_{\mathrm{ee}}(t) \sim \begin{cases} \frac{t}{(\tau_R \tau_b)^{1/2}} & \text{for } t < \tau_R/N^2 + \tau_b \approx \tau_b \\ \left(\frac{t}{\tau_R}\right)^{1/2} & \text{for } \tau_b < t < \tau_R \\ 1 - \exp\left(-\frac{t}{\tau_R + \tau_b}\right) & \text{for } t > \tau_R \end{cases} \tag{3.78}$$

3. Dominating internal friction $\frac{\tau_b}{\tau_R} > 1$:

$$C_{\mathrm{ee}}(t) \sim \begin{cases} \frac{t}{\tau_b}\left(1 - \frac{1}{N}\right) & \text{for } t < \tau_b \\ 1 - \exp\left(-\frac{t}{\tau_b}\right) & \text{for } t > \tau_b \end{cases} \tag{3.79}$$

3.5.5 Passage Times from Rouse Model including Internal Friction

As defined in the main text, the mean-passage time τ_{mp} arising from the end-to-end MSD is given by the inversion of

$$C_{\mathrm{ee}}(\tau_{\mathrm{mp}}) = C^*_{\mathrm{ee}} \tag{3.80}$$

in the short time regime.

3.5.5.1 Scaling Results for the Passage Time

Since we considered the short time limit of $C_{\mathrm{ee}}(t)$, we can obtain the dependence of τ_{mp} on τ_m simply via the scaling behaviour of the end-to-end MSD in the short time regime:

$$\begin{aligned} \text{negligible } \tau_b: \quad & C_{ee}(t) \sim \frac{t}{N\tau_m} \stackrel{!}{=} C^*_{\mathrm{ee}} \\ & \Rightarrow \quad \tau_{mp} \sim C^*_{\mathrm{ee}} N \tau_m \,, \\ \text{intermediate } \tau_b: \quad & C_{ee}(t) \sim \frac{t}{N\left(\tau_m \tau_b\right)^{1/2}} \stackrel{!}{=} C^*_{\mathrm{ee}} \\ & \Rightarrow \quad \tau_{mp} \sim C^*_{\mathrm{ee}} N \left(\tau_m \tau_b\right)^{1/2} \,, \\ \text{dominating } \tau_b: \quad & C_{ee}(t) \sim \frac{t}{\tau_b} \stackrel{!}{=} C^*_{\mathrm{ee}} \\ & \Rightarrow \quad \tau_{mp} \sim C^*_{\mathrm{ee}} \tau_b \,. \end{aligned} \tag{3.81}$$

3.5.5.2 Effective Power Law Exponents

To investigate the scaling of τ_{mp} in the three different viscosity regimes, we show the slope of τ_{mp} versus the solvent viscosity η in a log-log-plot of τ_{mp}, namely $\mathrm{d}\log(\tau_{\mathrm{mp}}/\tau_b)/\mathrm{d}\log(\tau_m^0\eta/\tau_b\eta_0)$. This slope gives the effective power law exponent α of the viscosity dependence of τ_{mp} as shown in 3.7. The results are consistent with the prediction of the theory in 3.81, one can easily distinguish the three different regimes with slopes of zero, 1/2, and unity from the graph. In between the asymptotic scaling regimes, the slope is continuously shifting. As defined in the main text, τ_m^0 is the monomer relaxation time at reference solvent viscosity $\eta = \eta_0$, meaning that $\tau_m = \tau_m^0\eta/\eta_0$.

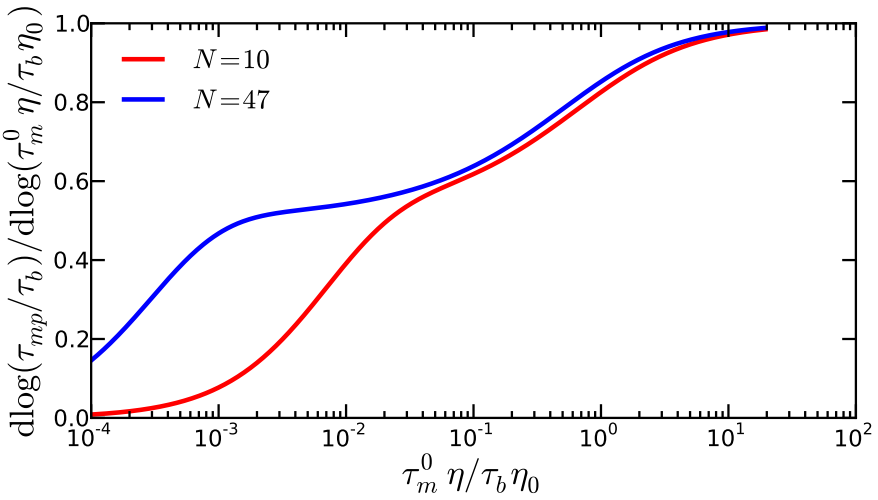

FIGURE 3.7: Effective power law exponent α of the viscosity dependence of τ_{mp} as a function of viscosity η for $N = 10$ (red), $N = 47$ (blue) and $C_{ee}^* = 0.01$. Consistent with the scaling predictions in 3.81, the effective exponent vanishes for vanishing solvent viscosity indicating a dominance of the constant part of τ_{mp} given by τ_b, the passage time due to internal friction. For intermediate solvent viscosities the curve flattens around $\alpha \approx 0.5$, indicative of the intermediate scaling range with a $\sqrt{\eta}$-dependence. Increasing η leads to $\alpha \to 1$, since internal friction effects become negligible compared to solvent friction effects and one obtains a linear behaviour $\sim \eta$. For small N the intermediate scaling regime is very narrow and one observes rather a continuously shifting exponent.

3.5.5.3 Fit of the Passage Time

Consider the scaling behaviour of τ_{mp} as given in 3.81. In order to apply our theory to experimental or simulation results for mean-passage times in a convenient fashion, we now construct a closed-form scaling function that exhibits all three viscosity regimes and also accurately reproduces the entire function in 3.55. Such a function can simply be constructed by addition and reads

$$\frac{\tau_{\mathrm{mp}}^{\mathrm{Fit}}}{C_{\mathrm{ee}}^{*}} = C_1\tau_b + C_2 N\sqrt{\tau_m\tau_b} + C_3 N\tau_m \tag{3.82}$$

$$= C_1\tau_b + C_2 N\sqrt{\tau_b\tau_m^0\frac{\eta}{\eta_0}} + C_3 N\tau_m^0\frac{\eta}{\eta_0} \tag{3.83}$$

where τ_m^0 is the proportionality constant relating τ_m and η/η_0, $\tau_m = \tau_m^0\,\eta/\eta_0$. The amplitudes C_i are a priori unknown and we determine them via a fit to numerical results for τ_{mp}. In order to make sure that we recover the exact result for vanishing solvent viscosity, we fix the fit parameter C_1 by the limiting value of $\tau_{\mathrm{mp}}/\tau_b$ at $\tau_m = 0$. So we are left with two fit parameters C_2 and C_3. The results of this fit for $N = 10, 47$ are shown in 3.8 and parameter values are given for $N = 10, 20, 47, 100$ in 3.1.

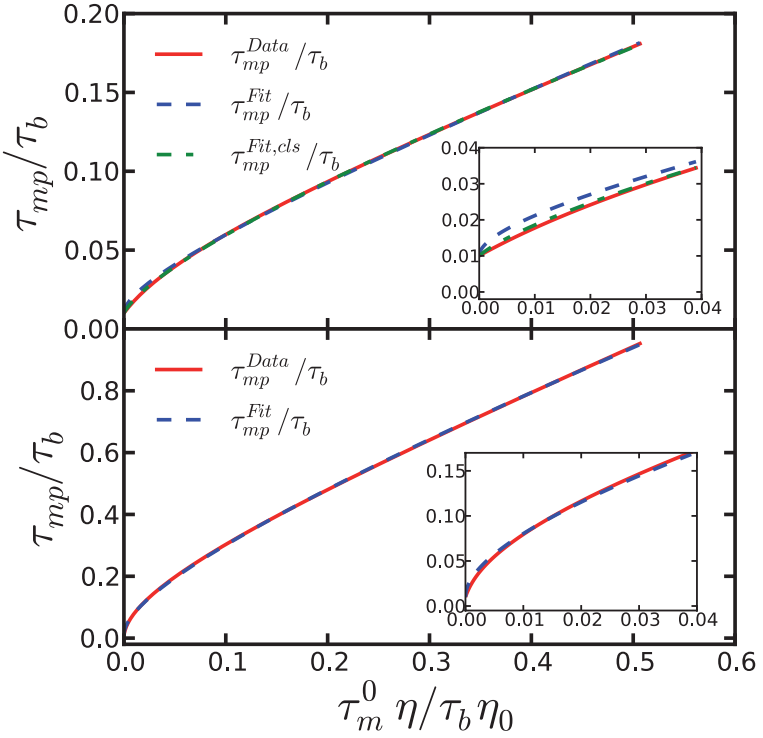

FIGURE 3.8: Fit of the Rouse result for passage time τ_{mp} versus solvent viscosity η obtained via numerical inversion of $C_{ee}(\tau_{mp}) = C_{ee}^{*}$ for mode number cutoff $N = 10$ (upper panel), $N = 47$ (lower panel) and $C_{ee}^{*} = 0.01$. The fit functions used are defined in eqs. (3.82) and (3.87) and the values of the fit parameters are given in Table 3.1. As clearly visible in the inset, the fit function including the correction to leading scaling yields a better representation of the numerically determined curve τ_{mp}^{Data} in the small viscosity regime.

The results for the fit parameters do not differ much, as one would expect, since we treat the leading N-dependence explicitly. For both N shown in 3.8 the overall behaviour of the numerically exact result for τ_{mp}, as defined in 3.55, namely $\tau_{\mathrm{mp}}^{\mathrm{Data}}$, is well described by the fit function. But for $N = 10$ in the regime of small solvent viscosities η there are clear deviations of $\tau_{\mathrm{mp}}^{\mathrm{Fit}}$ from $\tau_{\mathrm{mp}}^{\mathrm{Data}}$. Furthermore we varied C_{ee}^{*} for fixed $N = 10$ and performed fits. As shown in 3.1, the parameter values for different values of C_{ee}^{*} are

very similar, the small deviations for $C^*_{ee} = 0.1$ are reasonable since $C_{ee}(t)$ in that range is not entirely in the diffusive short-time regime. Given the numerical values of the amplitudes C_i, one is now able to fit 3.83 to experimental or simulation results of mean-passage times. Then there are only two fit parameters, namely the mode relaxation time due to internal friction τ_b and the proportionality constant τ^0_m.

TABLE 3.1: Values of the fit parameter C_i in eqs. (3.82) and (3.87)

N	C^*_{ee}		C_1	C_2	C_3	C_4
10	0.1	τ^{Fit}_{mp}	1.0	0.8	3.12	-
		$\tau^{Fit,cls}_{mp}$	1.0	1.35	3.36	2.16
	0.01	τ^{Fit}_{mp}	1.0	0.91	2.11	-
		$\tau^{Fit,cls}_{mp}$	1.0	1.67	1.92	1.59
	0.001	τ^{Fit}_{mp}	1.0	0.91	2.03	-
		$\tau^{Fit,cls}_{mp}$	1.0	1.68	1.81	1.56
20	0.01	τ^{Fit}_{mp}	1.0	1.15	2.01	-
		$\tau^{Fit,cls}_{mp}$	1.0	1.35	1.96	3.18
47	0.01	τ^{Fit}_{mp}	1.0	1.28	2.14	-
		$\tau^{Fit,cls}_{mp}$	1.0	1.28	2.14	13304187.15
100	0.01	τ^{Fit}_{mp}	1.0	1.25	2.85	-
		$\tau^{Fit,cls}_{mp}$	1.0	1.25	2.85	11403389.0

3.5.5.4 Correction to leading scaling

As already mentioned in the preceding section, the fit function given in 3.82 does not correctly reproduce the theoretical results in the small viscosity regime. To improve the quality of the description we now reconsider the behaviour of τ_{mp} in this regime and calculate the correction to leading scaling. Therefore we reconsider 3.57, which can be rearranged to

$$C_{\mathrm{ee}}(t) \approx \frac{t}{\tau_b} \sum_{p=1,\,\mathrm{odd}}^{N} \frac{1}{p^2} \frac{1}{N^2\tau_m/\tau_b p^2 + 1} \,. \tag{3.84}$$

In the limit $\tau_m/\tau_b \to 0$ we can make a Taylor expansion, yielding

$$\begin{aligned} C_{\mathrm{ee}}(t) &\approx \frac{t}{\tau_b} \sum_{p=1,\,\mathrm{odd}}^{N} \frac{1}{p^2} \left(1 - \frac{N^2\tau_m}{\tau_b p^2} + \left(\frac{N^2\tau_m}{\tau_b p^2}\right)^2 - \dots\right) \\ &= \frac{t}{\tau_b} \sum_{p=1,\,\mathrm{odd}}^{N} \left(\frac{1}{p^2} - \frac{N^2\tau_m}{\tau_b}\frac{1}{p^4} + \dots\right) \stackrel{!}{=} C^*_{\mathrm{ee}} \,. \end{aligned} \tag{3.85}$$

We assume $\sum_{p=1,\,\mathrm{odd}}^{N} 1/p^2 \approx \sum_{p=1,\,\mathrm{odd}}^{N} 1/p^4 \approx 1$ and obtain by inversion of 3.85

$$\tau_{\mathrm{mp}} \approx \tau_b C^*_{\mathrm{ee}} \left(1 - \frac{N^2 \tau_m}{\tau_b}\right)^{-1} \approx C^*_{\mathrm{ee}} \tau_b + C^*_{\mathrm{ee}} N^2 \tau_m \,, \tag{3.86}$$

where we made an expansion up to first order in τ_m/τ_b. Thus the correction to the leading scaling is linear in τ_m and we can define an improved fit function $\tau_{\mathrm{mp}}^{\mathrm{Fit,cls}}$ including the correction to leading scaling as

$$\frac{\tau_{\mathrm{mp}}^{\mathrm{Fit,cls}}}{C^*_{\mathrm{ee}}} = C_1 \tau_b + \tau_b \left(\frac{1}{C_2 N \sqrt{\frac{\tau_m}{\tau_b}} + C_3 N \frac{\tau_m}{\tau_b}} + \frac{1}{C_4 N^2 \frac{\tau_m}{\tau_b}} \right)^{-1} \tag{3.87}$$

$$= C_1 \tau_b + \left(\frac{1}{C_2 N \sqrt{\tau_b \tau_m^0 \frac{\eta}{\eta_0}} + C_3 N \tau_m^0 \frac{\eta}{\eta_0}} + \frac{1}{C_4 N^2 \tau_m^0 \frac{\eta}{\eta_0}} \right)^{-1} \tag{3.88}$$

Analogously to the preceding section we performed a fit of 3.87 to the numerical results for τ_{mp}. The outcome can again be seen for $N = 10, 47$ in 3.8 and for $N = 10, 20, 47, 100$ in 3.8. For large N the parameter C_4 has large values indicating that the correction to leading scaling is less important, since large C_4 values give negligible weight to the correction term. Once the parameter values are determined, 3.88 can be used as an alternative fit function for experimental or simulation results for mean-passage times that would be more accurate in the limit of vanishing solvent viscosity.

3.5.6 Fits of folding times as a function of viscosity

Troughout the main text two different ways of fitting simulation data are employed: In the first way, the averaged mean first passage time data are fitted to the models defined in (Eqs. (3), (4) and (7)) in the main text:

$$\text{Linear model: } \bar{\tau}_{\mathrm{fp}} = \tau_{\mathrm{int}} + \tau_{\mathrm{wat}}\, \eta/\eta_0 \tag{3.89}$$

$$\text{Power law model: } \bar{\tau}_{\mathrm{fp}} = \tau_0 \left(\eta/\eta_0\right)^{\alpha} \tag{3.90}$$

$$\text{Rouse model:} \tag{3.91}$$

$$\frac{\bar{\tau}_{\mathrm{fp}}}{C^*_{\mathrm{ee}}\, \tau_b} = C_1 + C_2\, N \sqrt{\frac{\tau_m^0}{\tau_b} \frac{\eta}{\eta_0}} + C_2\, N\, \frac{\tau_m^0}{\tau_b} \frac{\eta}{\eta_0}$$

In the second way, the local friction coefficient is fitted to a viscosity dependent model prediction. Here we only use the linear model and fit the friction data for different values of the reaction coordinate, thereby obtaining the internal friction profile ξ_{int},

$$\xi(Q, \eta/\eta_0) = \xi_{\mathrm{int}}(Q) + \xi_{\mathrm{wat}}(Q)\, \eta/\eta_0. \tag{3.92}$$

At first sight, both ways of fitting, the first being based on folding times, the second being based on friction, seem equivalent. This of course is only true, if one can make sure that the free energy landscape is not modified by changes in the solvent viscosity, which experimentally is an issue, as discussed in the main text. If viscogenic agents affect both the solvent viscosity and the free energy profile, a discussion of folding times in terms of internal friction alone (and neglecting the variations of the free energy profile) would be highly inconsistent. From a more general point of view, the friction is an input parameter whereas the folding time is a derived quantity, which would also suggest to center the discussion of internal friction effects on the friction itself. In support of this, we note that the friction profile typically varies locally as a function of the reaction coordinate; since the folding process integrates over a finite segment along the reaction coordinate, the relative weights of internal and solvent friction are thereby smeared out in the resulting folding time. Taking all arguments together, the full friction profile, in combination with a theoretical model, is necessary to unambiguously disentangle internal from solvent friction effects.

In our Rouse model calculation, we have extracted the internal friction contribution from the mean-passage time in the diffusive short-time limit, where friction coefficient and folding time are trivially related. So the issue as to whether friction or folding time is a better target for fitting does not occur on that level. We have checked in an explicit calculation that in the Rouse model, the local friction coefficient does not vary along the end-to-end reaction coordinate, showing that the Rouse model in that respect is less complex than the simulation model.

All fits are executed using a least-square scheme, i.e. the residual sum-of-squares $RSS = \sum_{i=1}^{n} \epsilon_j^2$ is minimized, where $\epsilon_j = \bar{\tau}_{\mathrm{fp,j}} - \bar{\tau}_{\mathrm{fp}}$ is the residual. In the case of linear fits an Ordinary Least Squares (OLS) scheme is used, for the other functional forms nonlinear least squares schemes using the Gauss-Newton algorithm have to be applied. All calculations are carried out using the `R` statistics package. The resulting parameters of the fits for eqs. (3.89) to (3.91) are shown 3.2.

TABLE 3.2: Fit parameters for linear, Rouse model and power law fit (cf. main text for definition) as well as corresponding residual standard error estimates $\hat{\sigma}$ according to 3.93

	τ_{int}	τ_{wat}	$C^*_{\mathrm{ee}}\tau_b$	$\frac{\tau_m^0}{\tau_b}$	τ_0	α	$\hat{\sigma}_{\mathrm{linear}}$	$\hat{\sigma}_{\mathrm{rouse}}$	$\hat{\sigma}_{\mathrm{power}}$
$Q^{\mathrm{GS4}}_{\mathrm{ee}}$	0.75 ns	1.12 ns	0.34 ns	0.09	2.05 ns	0.58	0.166 ns^2	0.093 ns^2	0.081 ns^2
$Q^{\mathrm{GS4}}_{\mathrm{gyr}}$	0.50 ns	0.67 ns	0.24 ns	0.07	1.28 ns	0.55	0.137 ns^2	0.091 ns^2	0.068 ns^2
$Q^{\mathrm{A8}}_{\mathrm{ee}}$	0.08 ns	0.14 ns	0.03 ns	0.13	0.24 ns	0.64	0.010 ns^2	0.011 ns^2	0.019 ns^2
$Q^{\mathrm{A8}}_{\mathrm{gyr}}$	0.36 ns	0.77 ns	0.13 ns	0.18	1.21 ns	0.67	0.352 ns^2	0.345 ns^2	0.344 ns^2
$Q^{\mathrm{A8}}_{\mathrm{rms}}$	0.46 ns	2.60 ns	0.07 ns	1.40	3.14 ns	0.86	0.477 ns^2	0.485 ns^2	0.497 ns^2
$Q^{\mathrm{1prb7}}_{\mathrm{fold}}$	32.1 μs	228 μs	1.85 μs	0.83	257 μs	0.75	8.58 $\mu\mathrm{s}^2$	7.51 $\mu\mathrm{s}^2$	7.93 $\mu\mathrm{s}^2$
$Q^{\mathrm{1prb7}}_{\mathrm{unfold}}$	32.8 μs	235 μs	2.29 μs	0.65	262 μs	0.73	16.2 $\mu\mathrm{s}^2$	11.2 $\mu\mathrm{s}^2$	8.51 $\mu\mathrm{s}^2$

As a simple measure to quantify goodness-of-fit we show the residual standard error $\hat{\sigma}$ for the models defined above in 3.2:

$$\hat{\sigma} = \sqrt{\frac{RSS}{n-2}}, \tag{3.93}$$

where n is the number of data points being regressed.

3.5.7 Hydrogen Bonds

The number of hydrogen bonds is determined using the `g_hbond` tool provided by GROMACS. Here OH and NH groups are considered to be donors and O and N to be acceptors. Hydrogen bonds are then determined using a cutoff radius of $r_{\text{hbond}} = 0.35\,\text{nm}$ between donor and acceptor and a cutoff $\alpha_{\text{hbond}} = 30°$ for the donor–hydrogen–acceptor angle [58].

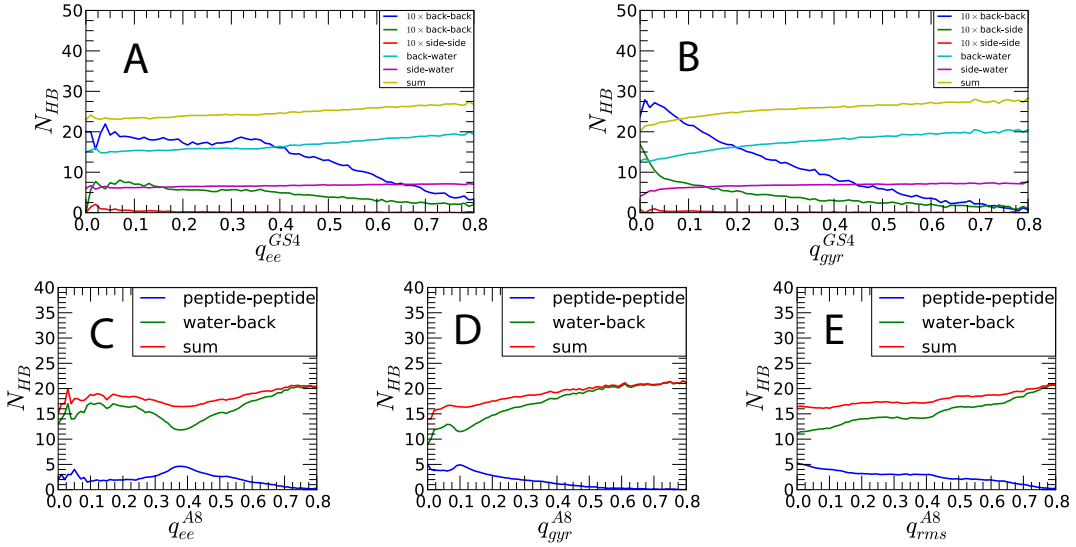

FIGURE 3.9: Number of hydrogen bonds between different subgroups of the system (sum means the sum of all peptide-peptide and peptide-water hydrogen bonds) as a function of different reacation coordinates:
A) $q_{\text{ee}}^{\text{GS4}}$ B) $q_{\text{gyr}}^{\text{GS4}}$ C) $q_{\text{ee}}^{\text{A8}}$ D) $q_{\text{gyr}}^{\text{A8}}$ E) $q_{\text{rms}}^{\text{A8}}$

3.9 shows the resulting hydrogen bond profiles between water and various parts of the peptide along the different reaction coordinates. The sum of intra-peptide and peptide-water hydrogen bonds is almost constant, which shows that upon unfolding water more or less compensates for the loss of possible intrapeptide hydrogen bonds. For GS_4 the total number of hydrogen bonds is dominated by water-peptide hydrogen bonds, and shows no prominent features. The slight decrease of the total N_{HB} at small reaction coordinate values $q_{\text{gyr}}^{\text{GS4}}$ in 3.9B indicates that very compact structures are unable to efficiently hydrogen bond. The decomposition into the various contributions shows that although the number of side-chain backbone hydrogen bonds goes up for decreasing values of $q_{\text{gyr}}^{\text{GS4}}$, this increase is not strong enough to make up for the loss of

water-peptide hydrogen bonds. In the corresponding figure for the alanine helix one can see, as expected, that the number of peptide-peptide hydrogen bonds is highest for perfectly helical structures, which is appreciated from the plot in 3.9E as a function of $q_{\mathrm{rms}}^{\mathrm{A8}}$. Opening helix turns means breaking intra-peptide hydrogen bonds and thus exposing them to the water. Interestingly, the total number of water hydrogen bonds goes down upon helix formation, i.e. more peptide-water HBs are broken than peptide-peptide HBs are created when the helix forms. This simply shows that water is a very versatile and strong hydrogen bond forming agent.

3.6 Details of Atomistic Polypeptide Simulations

Our simulation are carried out using the GROMACS 4.5 [17] molecular dynamics package using the Amber ff03 force field [30] and the SPC/E [29] water model. The box size is triclinic with vectors (3.18 nm,0. nm,0. nm), (0. nm,3.18 nm,0. nm) and (1.59 nm,1.59 nm, 2.25 nm) for GS_4 and (3.14,0. nm,0. nm), (0. nm,3.14 nm,0. nm) and (1.59 nm, 1.59 nm, 2.25 nm) for Ala_8. The box is filled with 711 water molecules for the glycine-serine chain and with 688 water molecules for the alanine helix. Detailed information about running time and time steps is shown in 3.3. We simulate in the NPT ensemble, where the temperature is coupled to a heat bath of $T = 300\,\mathrm{K}$ by a v-rescale thermostat [22] and the pressure is set to 1 bar using a Parinello-Rahman barostat [23] where the bulk compressibility is $\kappa = 4.5 \times 10^{-5}\,\mathrm{bar}^{-1}$. The Particle-Mesh Ewald method [21] is used for the long-ranged Coulomb interactions. The cutoff distance for non-bonded Coulomb and Lennard-Jones interactions is set to 0.9 nm. Every 20 steps the neigbour list for non-bonded interactions is updated. All bonds are constrained using the LINCS algorithm [62].

In order to keep the system electrically neutral and to eliminate spurious end effects, we cap the N-terminus of the peptides with an acetyl (ACE) group and the C-terminus with an N-methyl amide (NME).

TABLE 3.3: Simulation time steps and trajectory resolution (identical for both peptides) and simulation length for both peptides

m/m_0	time step [fs]	trajectory resolution [ps]	duration $(GS)_4$ [μs]	duration Ala_8 [μs]
0.01	0.25	5	0.5	0.5
0.03	0.5	2.5	0.5	0.75
0.1	1	5	0.5	1
0.25	1	5	0.5	1
0.5	2	10	1	2
0.75	2	10	1	2
1	2	10	1	3
1.5	2	10	1	4
2	2	10	2	4
4	2	10	2	3
6.25	2	10	2	4
9	2	10	2	4

3.6.1 Coarse-grained protein folding model

3.6.1.1 Modifications of "Martini" force field

As the basis for our protein folding model we use the "Martini" coarse-grained force-field for both the solvent [63] and protein [64]. This model represents the protein residues by a backbone bead and a number of side-chain beads, which differ according to residue type. The solvent is represented by a "coarse water" particle that represents a blob of approximately four waters. The model uses a standard force field energy function, with harmonic bonds and angles, and nonbonded interactions described using 12-6 Lennard-Jones and electrostatic interactions. The balance of protein-solvent, protein-protein and solvent-solvent interactions in the Martini force field maintains a compact structure for the protein, but does not preserve the native fold. To maintain the native structure using this potential, additional energy terms such as an elastic network model or fixed torsion angles have been used [65]. Since such approaches are clearly not applicable to folding, we have instead adapted the protein force-field to create a type of Gō model. Specifically, we made the following modifications: (i) all charges were set to zero; (ii) pairs of beads closer than 9 Å in the native structure were treated with the standard Martini Lennard-Jones parameters; (iii) bead pairs further than 9 Å in the native structure were treated by a repulsive potential of the form $V(r) = (\sigma_{\text{rep}}/r)^{12}$ with $\sigma_{\text{rep}} = 7.2$ Å; (iv) Cα-Cα equilibrium bond lengths were set to 3.8 Å to be consistent with experimental protein structures (v) Cα-Cα-Cα-Cα torsion potentials were introduced for the backbone, taken from the work by Karanicolas and Brooks [66]; (vi) the spring constants for bonds and angles were set to 1.0×10^4 kJ mol^{-1} nm^{-2} and 6.0×10^2 kJ mol^{-1} rad^{-2} respectively, the latter to avoid colinearity of atoms involved in torsion angles. Note that all protein-solvent and solvent-solvent pair interactions were unchanged. Tools for transforming the original Martini model into a Gō-like model are available on request from the authors.

3.6.1.2 Simulation protocols

We constructed a Gō-type model of the 47-residue 3-helix bundle protein $1prb_{7-53}$ studied by Gai and co-workers [67] as described above, based on the experimental structure [68] (3.10 (A), (B)). Simulations were run with the Gromacs 4.0.5 simulation package [69, 70], using parameters similar to those for the standard Martini force field, modified as described above. The folded protein was solvated in a truncated octahedron simulation cell with an initial distance between nearest faces of 6.5 nm. All bond lengths were left unconstrained, except for the constraints defined within side-chain rings in MARTINI. A time step was chosen for each solvent mass such that energy was conserved in the absence of a thermostat: this resulted in time steps of 3 fs for a mass of 2 a.m.u., 5 fs for a mass of 8 a.m.u., 10 fs for a mass of 24 a.m.u. and 15 fs for the larger solvent masses of 72 and 216 amu (where the protein dynamics limits the time step). The reference mass $m_0 = 72$ amu, corresponding to four water molecules. After a short equilibration at constant pressure of 1 bar (Parinello-Rahman barostat [23]) and temperature of 320 K simulations were run at constant volume at a temperature of 320 K, with a Nosé-Hoover thermostat [71], so as to minimally interfere with the dynamics.

3.6.1.3 Folding Times

Folding times were calculated from the average lifetime of the protein in the unfolded state, and unfolding times from the average lifetime in the folded state. Folding was monitored using the fraction of native contacts Q (3.10 D): a folding event was counted when Q first reached a value of 0.85, starting from unfolded, while an unfolding event was counted when Q first reached 0.44, starting from folded.

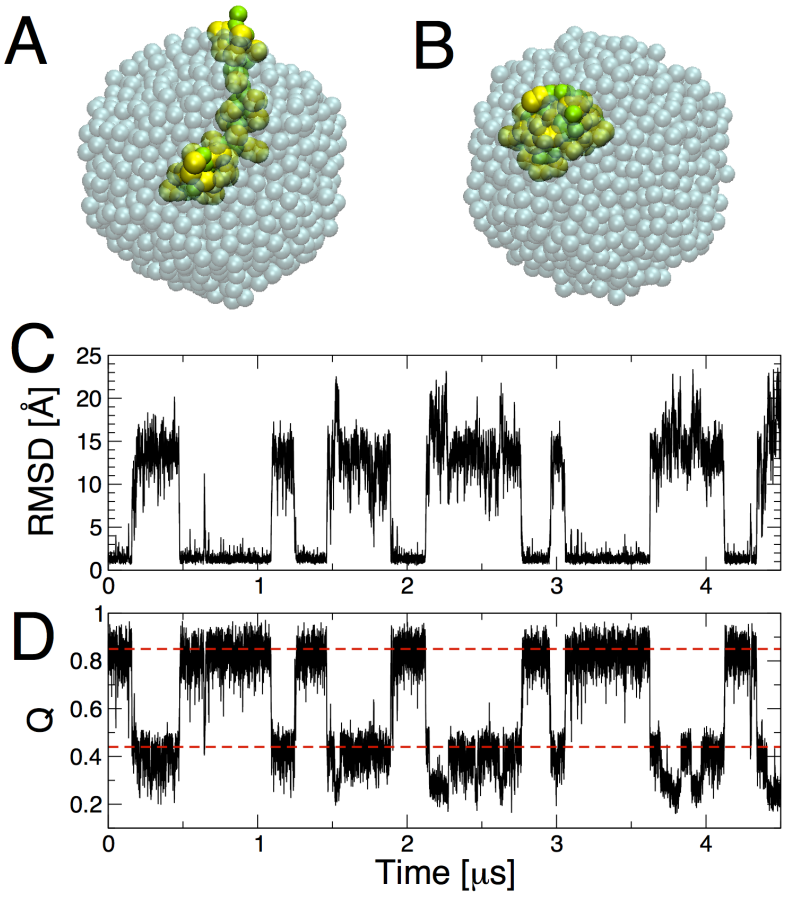

FIGURE 3.10: Coarse-grained protein folding model. Snapshots of (A) unfolded and (B) folded conformations of the protein in explicit solvent; (C) RMSD from native structure and (D) fraction of native contacts, showing two-state folding. Q cut-offs for defining transitions between unfolded ($Q = 0.44$) and folded ($Q = 0.85$) are shown by broken red lines.

Chapter 4

Unfolding and folding internal friction of β-hairpins is smaller than of α-helices

By the forced unfolding of polyglutamine and polyalanine homopeptides in competing α-helix and β-hairpin secondary structures, we disentangle equilibrium free-energetics from non-equilibrium dissipative effects. We find that α-helices are characterized by larger friction or dissipation upon unfolding, regardless of whether they are free-energetically preferred over β-hairpins or not. Our analysis, based on MD simulations for atomistic peptide models with explicit water, suggests that this difference is related to the internal friction and mostly caused by the different number of intra-peptide hydrogen bonds in the α-helix and β-hairpin states.

4.1 Introduction

The folded structure a protein adopts under in-vitro equilibrium conditions is determined by the lowest free energy state [72]. This insight forms the starting point for theoretical as well as experimental studies employing the concept of a protein folding free-energy landscape. At the same time, since the early studies on the T and R forms of hemoglobin [73], it has been appreciated that for many proteins the native state is not unique and conformational diversity is important to achieve proper function. This revised view, according to which the native state is represented by an entire ensemble of conformers, is essential to understand a number of biological processes such as enzymatic catalysis [74], protein-protein recognition [75], and signal transduction [76], to name a few.

While protein conformational diversity is vital in many situations, it gives rise to detrimental health effects when proteins misfold into states that are toxic or exhibit loss of function. A number of neurodegenerative diseases such as Parkinson's, Alzheimer's,

Huntington's are associated with the conversion of normally soluble proteins into insoluble filaments, so-called amyloid fibrils, enriched in β-sheet structures [77]. Whereas the easily observable filamentous protein aggregates had initially been believed to be pathogenic, more recent studies indicate that abnormal soluble protein monomers or small oligomeric assemblies might be the main culprit [78].

A subclass of these human diseases are caused by homo-amino acid repeats [79]. Interestingly, mostly polyalanine (polyAla) and polyglutamine (polyGln) tracts are known to be prone to mutations that expand the homopeptide lengths and, beyond specific thresholds, lead to devastating inheritable diseases. [80] Similarly to the other neurodegenerative diseases, the molecular mechanisms that cause their toxicity upon expansion have remained largely unknown, [81] as do the normal molecular roles of polyAla and polyGln tracts. [82, 83]

The nine known polyGln disorders, including Huntington's disease [84], are late-onset neurodegenerative diseases with a typical onset length threshold of about 30–40 glutamines. [84, 85] The known polyAla diseases comprise congenital developmental disorders as well as progressive late-onset neurological diseases. [86] In the disease-associated proteins the normal polyAla lengths are 10–20; the disease-causing increases range from a single alanine to 11. [86] There is some evidence that polyAla expansions are even more toxic than polyGln expansions of similar length. [87]

Of particular importance for the present study is the experimental observation that the in-vitro aggregation kinetics of polyGln peptides is determined by an unfavorable monomer folding transition [88]. This stands in contrast to conventional nucleation models for polymer and filament growth and means that the study of the conformational interconversion kinetics of single homopeptides is relevant and interesting.

As a matter of fact, even the characterization of the equilibrium ensemble of single homopeptide conformers, involving the study of the structural and thermodynamic features of monomers, represents a major challenge with contemporary simulation technology. In extensive solvent-explicit MD simulations, the stability of different folding states of a single Gln_{40} homopeptide chain in the pathological length range was investigated [89]. Different structures, among them the conventional α-helix and β-strand states, but also more exotic ones such as β-sheet-stack and steric-zipper states were tested against each other and the relative stability was assessed from the speed with which different initially prescribed structures became unstable during the course of a simulation. While some structures could clearly be excluded from being relevant for the aggregation process, α-helix and β-hairpin structures showed high stabilities. In a similar simulation study for polyAla homopeptides the free-energetic stability of the α-helix and β-hairpin states was studied with particular stress on solvent effects [90].

Since the process of amyloid formation and fiber elongation typically occurs far from equilibrium, it transpires that the free energies of different peptide-monomer conformations are not the only decisive factors determining the likelihood with which a particular conformer is incorporated into a fibrillar structure. To give a simple example, let us assume that the α-helix and β-hairpin states (defined over certain basins in a suitably chosen peptide configurational space representation) of a particular peptide have

identical free energies compared to the disordered state. In that case the rates at which the α-helix and β-hairpin states form from the disordered ensemble are solely determined by the respective configurational diffusivities (i.e. the prefactors of the Kramers rates) and will in general not be identical. As a matter of fact, these diffusivities sensitively determine the probability with which a certain conformer is built into a fibrillar construct if (and only if) the fibril formation process occurs far from equilibrium.

This observation forms the conceptual starting point of the present work. To have a framework simple enough to allow for complete simulations but also for novel concepts to be developed and tested, we study two different relatively short homopolymers, namely polyalanine Ala_{15} and polyglutamine Gln_{15}, that are known to form competing α and β structures [89, 90]. The choice of peptides is motivated by their relevance for many neurodegenerative diseases, as discussed above. To disentangle free energetic and dissipative effects, we introduce a computational framework where we unfold different ordered structures by pulling on the peptide ends with prescribed speed. By extrapolation of the unfolding work to the limit of vanishing pulling speed, we estimate equilibrium free-energy differences between different peptide secondary structures. From the dependence of the unfolding work on the unfolding speed we extract the dissipative work and from that the effective friction coefficient for unfolding different folded states. Since on the linear-response level the friction for unfolding is the same as the friction for folding along the same path, our results thus hold for both folding and unfolding reactions.

While there are clearly more efficient simulation techniques for obtaining free energy differences and entire free energy landscapes of proteins, our method is uniquely suited to establish and compare the friction work associated with the unfolding of different secondary structures. While the free energetic weight of the α-helix and β-hairpin states differ considerably for Ala_{15} and Gln_{15} sequences, we find robustly that the unfolding friction of the β-hairpin state is considerably lower than for the α-helix state. Our interpretation of this finding in terms of the underlying hydrogen-bonding pattern in both states suggests that this finding should hold generally true also for other sequences. Note that β-hairpins experimentally typically show slower folding times than α-helix forming sequences. This is not in contradiction to our predictions which specifically concern the friction contribution. Interestingly, in recent work a β-hairpin sequence was designed that folds as fast as α-helices of comparable size [91], suggesting that it should experimentally be feasible to study β and α forming sequences that show similar or even identical free energies.

In a non-equilibrium situation, for example when β conformers are sequestered into kinetically arrested fibrils while the reverse process of fibril breakup is kinetically hindered, the folding friction can become the determining factor for the probability of a certain conformer to be found in fibrils. This simple argument illustrates why in situations far from equilibrium, it can be important to consider both free-energetic as well as dissipative properties of different peptide conformations.

4.2 Simulation Details

We study two homopeptides, polyalanine Ala_{15} and polyglutamine Gln_{15}, each 15 amino acids long. Except for Gln_{15} in the β starting configuration, all peptides are capped with ACE and NME. Some additional simulations are run for the shorter peptide Ala_8.

For our simulations we use the Gromacs 4.5 [17] molecular dynamics simulation package with the Amber03 [30] forcefield and the explicit SPC/E [29] water model. After prior equilibration for 100 ps in the NPT ensemble, production runs are performed in the NVT ensemble, with the temperature kept at $T = 300\,\mathrm{K}$ by the v-rescale thermostat [22]. The box size in pulling direction is 8 nm and in lateral directions $3\,\mathrm{nm} \times 4\,\mathrm{nm}$ for β-hairpins and $3\,\mathrm{nm} \times 3\,\mathrm{nm}$ for α-helices, resulting in about 3,000 water molecules. Electrostatics are accounted for using PME [21] with a cutoff of 0.9 nm.

The free energy landscape for Ala_8 in Fig. 4.6b has been obtained from a 1 μs equilibrium simulation, whereas the free energy landscapes for Ala_{15} in Figs. 4.2c and 4.6b have been obtained using the replica exchange method (1 μs each for 30 temperatures ranging from 300 K to 452 K).

For pulling simulations at constant pulling velocity v we use the *direction_periodic* mode of Gromacs' *pull code*, which adds a time-dependent potential acting between the center-of-masses of the two pulled groups,

$$U_{\mathrm{str}}(t) = \frac{k}{2}\Big(l_x(t) - (l_x(0) - v\,t)\Big)^2, \tag{4.1}$$

where k is the force constant, l_x the distance between the pulled groups along the x-direction, and t the simulation time. Force constants are varied between $k = 50\,\mathrm{kJ/(mol\,nm^2)}$ for low pulling speeds to $k = 150\,\mathrm{kJ/(mol\,nm^2)}$ for fast pulling speeds. The pulled groups are the first and last amino-acid residues of the peptide (or capping groups, if present).

The α and β states are separated by a large free energy barrier related to a torsional angle in the turn of hairpins in our particular setup, therefore the final states after a pulling simulation starting from α and β states are typically not the same [92]. In order to meaningfully compare the unfolding works of the α and β states, we also perform relaxation simulations by applying the time-dependent potential

$$U_{\mathrm{rel}}(t) = \frac{k}{2}\Big(l_x(t) - (l_x(0) + F_{init}/k + v\,t)\Big)^2, \tag{4.2}$$

where $l_x(0)$ is the final extension and F_{init} is the final force measured in a pulling simulation. By construction, the relaxation simulation starts with the final extension from the pulling simulation, by the term F_{init} the externally applied force at the beginning of the relaxation simulation is the same as at the end of the pulling simulation.

Similar to our previous work [93], we decompose the total dissipative friction work into the internal and solvent friction contributions by varying the water mass. To do so, we also perform simulations with modified water masses of $m/m_0 = 0.1$ and 10.

Our primary reaction coordinate is the distance along the x-axis l_x between the center of masses of the first and last aminoacids. In addition, we also consider the end-to-end distance R_{ee}, which is the separation in 3D space between the first and last aminoacids.

For monitoring the "folded" and "unfolded" ensembles we calculate further quantities, namely the solvent-accessible surface area, the number of intra-peptide and intra-backbone hydrogen bonds and the fraction of folded contacts Q_{aa}, which is defined as [94]

$$Q_{aa} = \frac{1}{N_{aa}} \sum_{i>j} \frac{1}{1 + \exp(\kappa\,(r_{ij} - \lambda\, r_{ij}^0))}. \tag{4.3}$$

The sum runs over all N_{aa} pairs of folded contacts, defined by two heavy atoms i and j being separated by at least three intermediate heavy atoms, i.e. $|i-j| > 3$, and a sub-threshold distance $r_{ij}^0 \leq 4.5\,\text{Å}$ in the folded state. The parameter κ controls the steepness of the crossover (here $\kappa = 5\,\text{Å}^{-1}$) and λ accounts for fluctuations in the folded state (here $\lambda = 1.5$). A value of $Q_{aa} = 1$ corresponds to all folded contacts being intact.

For the comparison of the Ala_8 and Ala_{15} free-energy landscapes we use the root-mean squared deviation from an ideal helical structure, Q_{rms}, based on the C_α atom positions; for this, the ideal helix radius and the ideal rise per residue are assumed to be 0.23 nm and 0.15 nm, respectively. As shown in our previous work [93], Q_{rms} is well suited to distinguish different conformations of an alanine α-helix.

The number of hydrogen bonds is calculated by the `g_hbond`-tool of Gromacs. Accordingly, a hydrogen bond between a donor (OH or NH) and an acceptor (O and N) exists if the bond angle is smaller than 30° and the distance between donor and acceptor atoms is less than 3.5 Å (cf. Luzar and Chandler [58]). For the system including ACE/NME caps, the hydrogen bond number does not include those caps. The average number of hydrogen bonds are calculated in 200 ns (50 ns for the alanine α-helix due to fast unfolding) equilibrium simulation. As the alanine β-hairpin is quite unstable and unfolds very quickly, in our analysis we use the backbone-only hydrogen bond number N_{HB}^{bb} of the Gln_{15} β-hairpin instead, motivated by the fact that N_{HB}^{bb} for the Gln_{15} and Ala_{15} α-structures are very similar and given by $N_{HB}^{bb} = 9.2$.

4.3 Results and discussion

4.3.1 Unfolding work

The time-dependent external stretching force acting between the two terminal pulled groups is according to Eq. 4.1 given by

$$F_{str}(t) = k \cdot \Big(l_x(t) - (l_x(0) - v\,t) \Big). \tag{4.4}$$

In Fig. 4.1b we plot the stretching force $F_{str}(t)$ for simulations of Ala_{15} starting in an initial α-helical structure, averaged over ten different simulation runs and smoothed over neighboring data points, as a function of the terminal end-group separation $l_x(t)$ for five

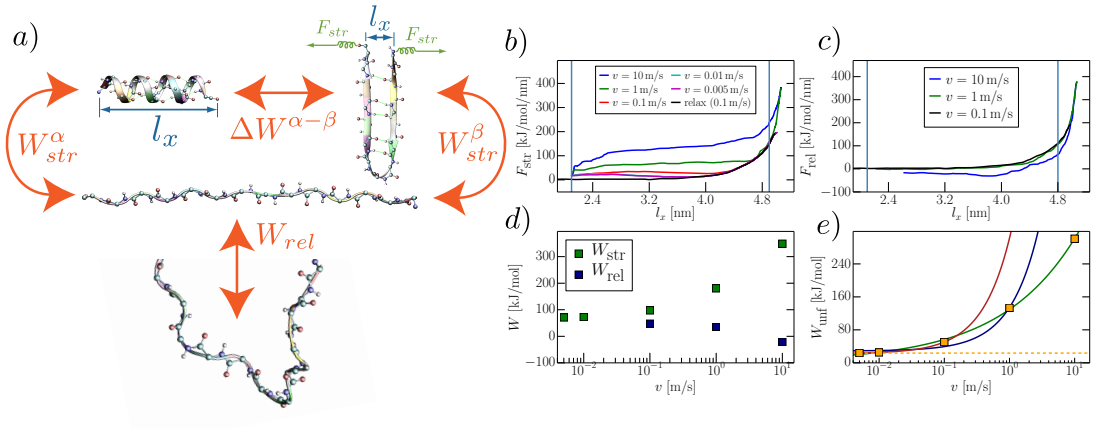

FIGURE 4.1: a) Snapshots of the Ala_{15} chain in the different representative states and schematic illustration of the pulling and relaxation simulation setup. b) Stretching force extension curves for Ala_{15} for different pulling velocities starting in the α-helical state. c) Relaxation force extension curves for Ala_{15} for different relaxation velocities starting in the fully stretched state. d) Stretching and relaxation works vs. pulling and relaxation speed v for Ala_{15} in the α-helical state. e) Unfolding work $W_{unf} = W_{str} - W_{rel}$ for Ala_{15} in the α-helical state defined in Eq. 4.7 vs. pulling speed v. The lines show different fits described in the text.

different stretching velocities. For the largest pulling velocity considered, $v = 10\,\mathrm{m/s}$, the stretching force is of the order $F_{str} \sim 100\,\mathrm{pN}$ for a whole range of intermediate separations. When v is reduced by a factor of 1,000 to $v = 0.01\,\mathrm{m/s}$ the measured force decreases considerably, which shows that the main contribution to the stretching force at large velocities is due to friction and thus is of dissipative nature. The steep force rise at large separations $l_x > 4.8\,\mathrm{nm}$ is due to backbone stretching and seen in all curves, regardless of the pulling velocity.

The stretching work performed by the external potential is given by the integral of the mean stretching force over the stretching separation

$$W_{str} = \int_{l_i}^{l_f} F_{str}(l_x)\,\mathrm{d}l_x. \tag{4.5}$$

The integration is carried out by a trapezoidal scheme with the lower and upper integration limits l_i and l_f indicated by vertical lines in Fig. 4.1b. In Fig. 4.1d we show the stretching work W_{str} for the Ala_{15} α-helix as a function of the pulling velocity (green squares), W_{str} is seen to saturate at a value of about $W_{str} \simeq 80\,\mathrm{kJ/mol}$ for velocities below $v \simeq 0.01\,\mathrm{m/s}$. In this limit, this stretching work is the free energy difference between the two thermodynamic states defined by the start and end configuration ensembles, corresponding to the α-helical and the strongly stretched states, respectively, as schematically shown in Fig. 4.1a. Most of this work corresponds to the elastic stretching of the peptidic backbone, an effect that we had scrutinized previously [95] and which we are not interested in here. To meaningfully subtract the elastic backbone contribution, we in Fig. 4.1c show force-extension curves obtained in the relaxation protocol using the time-dependent external potential $U_{rel}(t)$ defined in Eq. 4.2. Here we start in the fully stretched state and slowly decrease the separation between the peptide terminal groups. While for the stretching protocol data in Fig. 4.1b the external

potential performs work on the peptide, in the relaxation protocol the peptide chains performs work on the external potential. The resulting velocity dependence is quite different from the stretching-protocol data in Fig. 4.1b, for the largest relaxation velocity $v = 10\,\mathrm{m/s}$ the measured force in Fig. 4.1c is in fact smaller than the force at smaller velocities, in other words, dissipative effects in the relaxation protocol reduce the work extracted from the pre-stretched peptide chain, in agreement with the second law of thermodynamics. Note that the difference between the two relaxation force trajectories obtained for $v = 1\,\mathrm{m/s}$ and $v = 0.1\,\mathrm{m/s}$ in Fig. 4.1c is much smaller than the corresponding difference between the two stretching force trajectories obtained for $v = 1\,\mathrm{m/s}$ and $v = 0.1\,\mathrm{m/s}$ in Fig. 4.1b. This means that dissipative effects are less important in the relaxation from the stretched peptide state than in the forced unfolding of the α-helical state.

The interpretation of the asymmetry between pulling and relaxation transformations is straightforward in terms of the two-dimensional force trajectories as a function of the folded contact fraction Q_{aa} and the end-to-end radius R_{ee} shown in Fig. 4.2a: While the trajectories for pulling speed $v = 0.1\,\mathrm{m/s}$ (shown in red, all ten pulling trajectories are plotted on top of each other with the start and end positions denoted by blue symbols) connect the α-helical state, defined by $Q_{aa} \simeq 1$, with the fully stretched state defined by $R_{ee} \simeq 5\,\mathrm{nm}$, the relaxation trajectories at the same speed of $v = 0.1\,\mathrm{m/s}$ (shown in green, endpoints are shown by yellow triangles) do not bring the system back to the α-helical state. They rather correspond to the relaxation into the disordered ensemble with an unconstrained end-to-end distance distribution (to obtain this, we intentionally perform the pulling and relaxation simulation with a force along a fixed direction and not along the chain end-to-end group direction). In Fig. 4.2b the pulling and relaxation trajectories are shown for speeds of $v = 1\,\mathrm{m/s}$, here the ensemble of states after relaxation is even further away from the helical state. Our results in Fig. 4.2a demonstrate that the refolding time exceeds the allocated time in the relaxation simulations at $v = 0.1\,\mathrm{m/s}$ and leads to a disordered ensemble, this is further corroborated by comparison with the equilibrium distribution obtained by replica exchange simulations shown in Fig. 4.2c. The relatively large relaxation speed used in the simulations is not dictated by computer time considerations, but is rather chosen purposely and important for our further analysis, as will be explained further below.

We note at this point that the disordered ensemble that is created by the relaxation simulation clearly depends on the relaxation speed, as seen by comparing Figs. 4.2a and b. The relaxation work, defined as

$$W_{\mathrm{rel}} = \int_{l_i}^{l_f} F_{\mathrm{rel}}(l_x)\,\mathrm{d}l_x, \tag{4.6}$$

and for an Ala_{15} chain shown in Fig. 4.1d (blue squares), is smaller than the corresponding stretching work (green squares).

We define the pulling velocity-dependent unfolding work by

$$W_{\mathrm{unf}}(v) = W_{\mathrm{str}}(v) - W_{\mathrm{rel}}(v = 0.1\,\mathrm{m/s}), \tag{4.7}$$

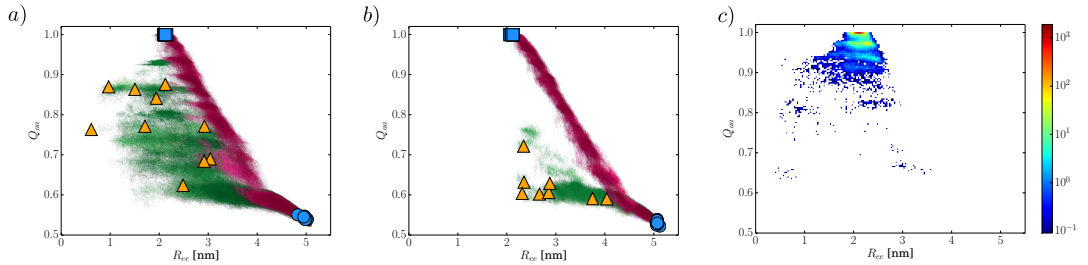

FIGURE 4.2: Ten different Ala_{15} stretching trajectories starting in the α-helical state (red, initial states: blue squares, final states: blue circles) and relaxation trajectories (green, final states: orange triangles) shown as a function of the fraction of folded contacts Q_{aa} and the end-to-end distance R_{ee} for identical pulling and relaxation speeds of a) $v = 0.1\,\mathrm{m/s}$ and b) $v = 1\,\mathrm{m/s}$. c) Free energy landscape of Ala_{15} obtained using replica-exchange simulations.

where we subtract the relaxation work at a fixed relaxation speed of $v = 0.1\,\mathrm{m/s}$. The precise relaxation speed used in the above definition is not crucial as long as it is fast enough not to allow the peptide chain to relax back into the folded state. As discussed before, the value of the relaxation speed defines the unfolded ensemble that one compares the folded state with. As defined in Eq. 4.7, $W_{\mathrm{unf}}(v)$ is the work needed to bring the chain from the ordered initial state to a particular unfolded ensemble and contains both equilibrium as well as dissipative contributions. The equilibrium unfolding work follows from an extrapolation of $W_{\mathrm{unf}}(v)$ to small v, although we hasten to add that in the strict limit $v \to 0$ the folded ensemble will not remain stable during the initial period of the pulling simulations. Clearly, if both pulling and relaxation transformation were performed at vanishing speeds, the pulling and the relaxing curves would perfectly superimpose and the unfolding work would disappear. So the present simulation methodology rests on a scale separation of the unfolding and refolding times on the one hand and the chain relaxation time on the other hand: both pulling and relaxation speeds are chosen fast enough such that the chain does not spontaneously unfold/refold upon pulling/relaxation, yet they are slow enough such that we enter the linear-response regime where the friction force is proportional to pulling speed and the extracted reconfigurational mobilities allow to extract the equilibrium diffusivities that govern peptide folding and unfolding.

In Fig. 4.1e we show the unfolding work $W_{\mathrm{unf}}(v)$ for the Ala_{15} α-helix as a function of the pulling velocity v (yellow squares). In order to separate the equilibrium unfolding work $W_{\mathrm{unf}}(v = 0)$ from the dissipative contribution $W_{\mathrm{diss}}(v) = W_{\mathrm{unf}}(v) - W_{\mathrm{unf}}(0)$ we fit the data to the form

$$W_{\mathrm{unf}}(v) = W_{\mathrm{unf}}(0) + W_{\mathrm{diss}}(v) = W_{\mathrm{unf}}(0) + \gamma\, L\, v^{\alpha} \tag{4.8}$$

where $L = l_f - l_i$ is the pulling distance and γ denotes the friction coefficient. In the linear-response regime the dissipative work should scale linear with the velocity [96] and thus $\alpha = 1$. The unrestricted fit to the entire data set in Fig. 4.1e (denoted by a green line) yields an exponent $\alpha = 0.37$ and describes the data over the whole velocity range quite well. But since we are mostly interested in the low-velocity linear-response regime, where we know that dissipation scales linearly in the velocity, an exponent

differing from unity is not very meaningful and a fit with fixed $\alpha = 1$ over a restricted velocity range is more useful, as will be discussed later.

A simple scaling argument helps us to estimate at which pulling velocities linear response is expected to break down [96]: From the scaling relation $F^* = k_B T/a_0$ for the mean thermal force acting on a fluctuating bond of range a_0, we find that for a typical length scale $a_0 \simeq 2 \times 10^{-10}$ m, which has been shown to roughly describe the range over which a hydrogen bond acts, the typical force above which non-linear effects set in is of the order of $F^* \simeq 20$ pN [96]. Indeed, a linear fit with $\alpha = 1$ over the data excluding only the high-velocity data point for $v = 10$ m/s (blue line in Fig. 4.1e) does not work very well, reflecting that the friction force for $v = 1$ m/s (which is the difference between the force for the smallest velocity $v = 0.01$ m/s and $v = 1$ m/s in Fig. 4.1a) is of the order of 50 pN and thus exceeds our simple scaling estimate $F^* \simeq 20$ pN. The linear fit with $\alpha = 1$ over the data excluding the data points for both $v = 10$ m/s and $v = 1$ m/s in Fig. 4.1e (red line) yields the equilibrium work of $W_{\rm unf}(v \to 0) \simeq 22.5$ kJ/mol, denoted by a horizontal broken line. The connection of this result with an equilibrium definition of the unfolding free energy will be discussed further below. Similar results for Ala_{15} in the β-hairpin state and for Gln_{15} are shown in Fig. 4.3a, the resulting equilibrium unfolding works are listed in Table 4.1.

4.3.2 Unfolding friction of β-hairpin versus α-helix

In Fig. 4.3b we show the difference of the unfolding work between an α-helix and a β-hairpin, defined as

$$\Delta W^{\alpha-\beta}(v) = W^{\alpha}_{\rm unf}(v) - W^{\beta}_{\rm unf}(v), \tag{4.9}$$

both for Gln_{15} as well as Ala_{15}, as a function of the pulling velocity. Simulation data are denoted by symbols, solid lines show the difference of the fitting curves according to Eq. 4.8 from Fig. 4.3a and are not fitted to the actual differential data in Fig. 4.3b. The good agreement with the data up to (and including) $v = 0.1$ m/s in Fig. 4.3b demonstrates that even the differences between unfolding works of α and β secondary structures can be described by a linear viscous law. Note also that the difference of the unfolding works $\Delta W^{\alpha-\beta}(v)$ is independent of the definition of the unfolded ensemble via the relaxation speed, which was primarily needed in order to be able to come up with unfolding free works $W^{\alpha}_{\rm unf}(v)$ and $W^{\beta}_{\rm unf}(v)$ that can be compared with alternative estimates based on equilibrium simulations. In that sense, our construction of the unfolding work difference $\Delta W^{\alpha-\beta}(v)$ is similar to a Born cycle and schematically shown in Fig. 4.1a.

By extrapolation to $v = 0$ m/s the equilibrium free energy difference between α and β secondary structures is estimated. For Ala_{15} we obtain $\Delta W^{\alpha-\beta}(v = 0) = 4.3$ kJ/mol, suggesting that the α-helical structure is slightly favored, for Gln_{15} we obtain $\Delta W^{\alpha-\beta} = 0$ kJ/mol within our numerical accuracy, so α and β structures are degenerate for polyglutamine. In previous simulation studies of secondary structure formation of Ala_{12} the β structure was found to be slightly favored with respect to the α helical state [90],

but we note that the precise free energy difference depends on the ensemble definitions and also on the used protein and water force fields [97].

We repeat that the main point of the present work is not to obtain free energy differences between folded structures, as there are techniques better suited for this, but rather to disentangle equilibrium from dissipative effects and in particular to obtain the dissipative contribution to the folding and unfolding kinetics for different secondary structures. In this respect, Fig. 4.3b reveals that $\Delta W^{\alpha-\beta}(v)$ for both Gln_{15} and Ala_{15} increases with v, meaning that the dissipative unfolding contribution of an α-helical structure is higher than of a β-hairpin structure, rather independent of the equilibrium free energy difference. Note that this finding, which is the main result of our work, is independent of details of the data fitting and therefore should be rather robust. Clearly, the dissipative unfolding work contains - among other contributions – a part due to hydrodynamic friction owing to the motion of the unfolded peptide section through the aqueous solvent by the action of the externally applied potential. Interestingly, the dissipative asymmetry between α and β structures seen in Fig. 4.3b cannot be due to hydrodynamic effects, as the terminal groups of the α-helices are moved over a shorter separation range ($l_i = 2.1\,\mathrm{nm}$ to $l_f = 4.8\,\mathrm{nm}$ for Ala_{15} and from $l_i = 1.8\,\mathrm{nm}$ to $l_f = 4.9\,\mathrm{nm}$ for Gln_{15}) than the end-groups of the β-hairpins (which are pulled from $l_i = 0.5\,\mathrm{nm}$ to $l_f = 4.8\,\mathrm{nm}$, see Table 4.1). This shows that the hydrodynamic drag contribution (which will be estimated in more detail further below) is in fact larger for β-hairpins, so the difference seen in Fig. 4.3b must be due to internal friction effects.

In Fig. 4.3c we present the difference between the unfolding works in the α-helical and β-hairpin states rescaled by the respective average intra-peptide hydrogen bond numbers N_{HB}^{α} and N_{HB}^{β} given in Table 4.1, defined as

$$\Delta W_{\mathrm{HB}}^{\alpha-\beta}(v) = W_{\mathrm{unf}}^{\alpha}(v)/N_{\mathrm{HB}}^{\alpha} - W_{\mathrm{unf}}^{\beta}(v)/N_{\mathrm{HB}}^{\beta}. \tag{4.10}$$

Except for the extreme pulling speed $v = 10\,\mathrm{m/s}$, the values of $\Delta W_{\mathrm{HB}}^{\alpha-\beta}(v)$ are quite close to zero (and much smaller than if we would simply divide $\Delta W^{\alpha-\beta}(v)$ by N_{HB}), suggesting that the difference of the dissipative unfolding contribution between α and β states is related to the different number of hydrogen bonds in the two states.

This reasoning is very much in line with recent work showing that the sliding friction force F_{f} of peptide bundles and of surface-adsorbed peptides in the viscous regime is proportional to the number of hydrogen bonds N_{HB} and the sliding velocity v and given by [96, 98]

$$F_{\mathrm{f}} = \gamma_{\mathrm{HB}}\, v\, N_{\mathrm{HB}}. \tag{4.11}$$

The friction coefficient per hydrogen bond was determined to be $\gamma_{\mathrm{HB}} \simeq 10^{-8}\,\mathrm{kg/s}$ for peptides sliding on a planar hydrophilic surface [96]. For peptide bundles it ranges from $\gamma_{\mathrm{HB}} \simeq 10^{-11}\,\mathrm{kg/s}$ up to $\gamma_{\mathrm{HB}} \simeq 10^{-6}\,\mathrm{kg/s}$ depending on the aggregation number of the peptide bundles [98]. Connecting this with our present findings, the larger friction when unfolding an α-helix can thus be rationalized by the larger number of hydrogen bonds per monomer present in the folded α-structure.

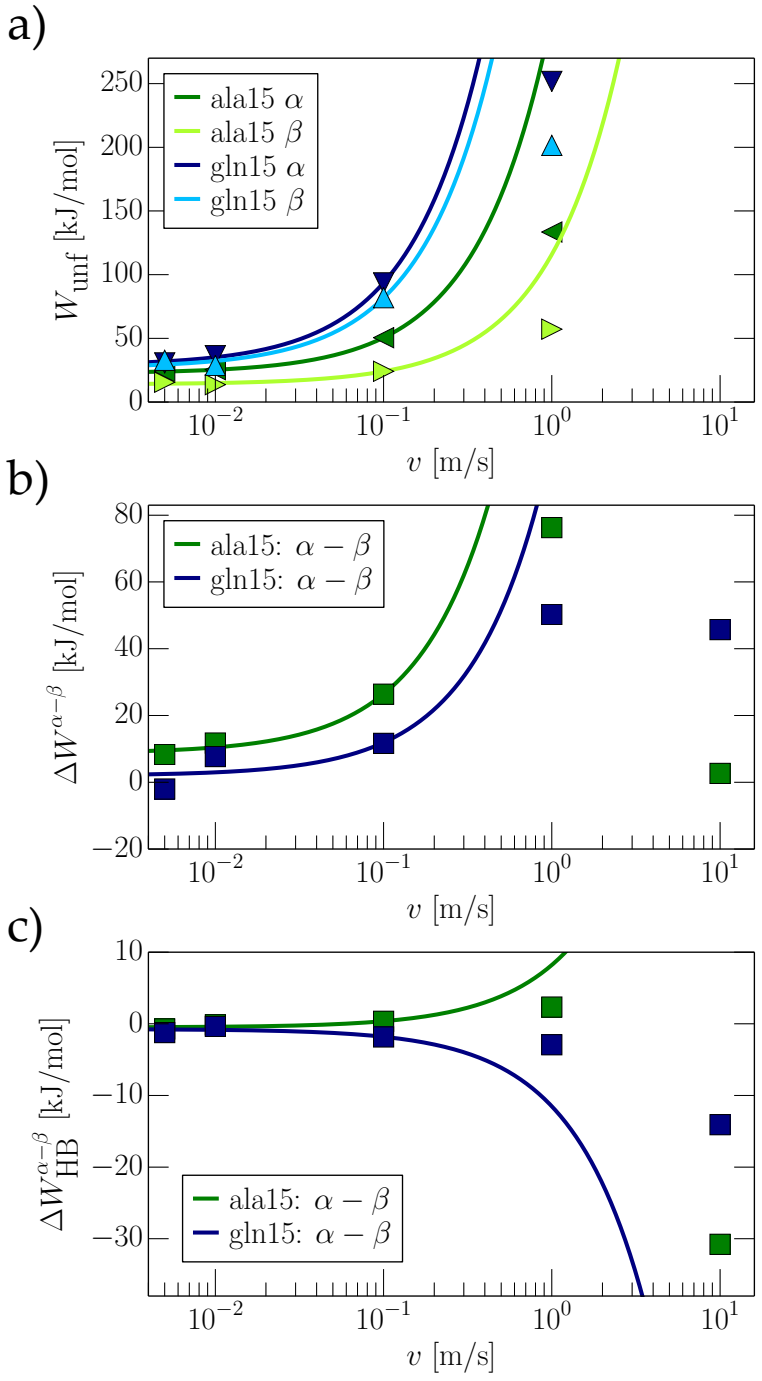

FIGURE 4.3: a) Unfolding work W_{unf} defined in Eq. 4.7 for Ala_{15} and Gln_{15} in the α-helical and β-hairpin states vs. pulling speed v. The lines show linear fits excluding the data points for $v = 10\mathrm{m/s}$ and $v = 1\,\mathrm{m/s}$. b) Unfolding work difference $\Delta W^{\alpha-\beta}(v)$ between the α and β structures as a function of the pulling speed, v, defined in Eq. 4.9. c) Work difference $\Delta W^{\alpha-\beta}_{HB}(v)$ between α and β structures rescaled by the respective number of hydrogen bonds according to Eq. 4.10. The lines in b) and c) are based on the linear fits presented in a).

Based on the viscous friction law Eq. 4.11 we can now derive a simple expression for the dissipative work upon unfolding defined in Eq. 4.8. For this we assume the friction

force F_f to be linear in the velocity v, $F_\mathrm{f} = \gamma\, v$, as valid in the linear-response regime for low pulling velocities, thereby defining the friction coefficient γ. Furthermore assuming the friction force to be constant over the entire pulling process, which seems reasonable since the α and β structures are homogeneous and unzipped locally at the pulled ends, we obtain for the dissipative work

$$W_\mathrm{diss}(v) = W_\mathrm{unf}(v) - W_\mathrm{unf}(0) = \int_{l_i}^{l_f} \mathrm{d}l_x\, F_\mathrm{f}(l_x) \simeq \gamma\, L\, v, \tag{4.12}$$

with $L = l_f - l_i$ being the integration limits denoted in Fig. 4.1 by vertical lines and listed in Table 4.1 for the different structures used. The friction coefficients γ following from the linear fits to the unfolding work in Fig. 4.3a are shown in Table 4.1.

It is instructive to compare these friction coefficients with what one would expect based on hydrodynamic drag. For this consider the schematic picture Fig. 4.4b of a β-hairpin that is terminally zipped open. The frictional force acting on each pulled strand with a transient length l_x is given by $F_\mathrm{f} \simeq v\ l_x/(\mu_\mathrm{hyd}\ b)$, here b is the contour length per amino-acid and μ_hyd denotes the hydrodynamic mobility per amino acid. In our previous simulations [96] we have estimated $1/\mu_\mathrm{hyd} \simeq 10^{-12}\,\mathrm{kg/s}$ in good agreement with experiments. The hydrodynamic dissipative work in the pulling process reads

$$W_\mathrm{hyd}(v) = 2\int_0^{L/2} \mathrm{d}l_x\, F_\mathrm{f} \simeq \frac{L^2\, v}{4\, b\, \mu_{hyd}}, \tag{4.13}$$

where the pre factor of two accounts for the fact that two strands are symmetrically pulled from the hairpin. Approximating $L/b \simeq 15$ we obtain $W_\mathrm{hyd}/(L\ v) \simeq 4 \times 10^{-12}\,\mathrm{kg/s}$ which is at least one order of magnitude smaller than the friction coefficients derived from the simulated dissipated works presented in Table 4.1. We conclude that while hydrodynamic friction effects are present in the simulations, and can play an important role for larger monomer numbers, they are overwhelmed by dissipative friction effects of non-hydrodynamic origin.

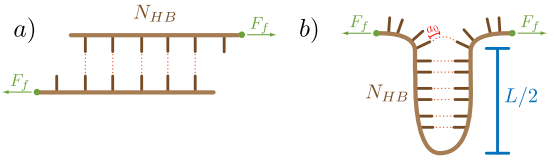

FIGURE 4.4: a) Friction scenario where shearing forces are applied on the opposing peptide strands in a β-strand structure and all N_HB hydrogen bonds are simultanenously sheared. b) Friction scenario where unzipping forces are applied on a β-hairpin and the hydrogen bonds rupture one by one.

To make the notion of hydrogen bond friction more concrete, we now estimate the friction per hydrogen bond from our data. The hydrogen-bond friction law in Eq. 4.11 was derived under the assumption that N_HB hydrogen bonds are sheared simultaneously, valid when an adsorbed peptide slides laterally over a planar surface [96] or when a peptide slides axially with respect to a co-axial bundle of peptides [98], as shown schematically in Fig. 4.4a. In the present case, the α-helix and β-hairpin secondary structures are unzipped at the ends, and it seems reasonable that only the outmost

terminal hydrogen bond is put under stress and thus the hydrogen bonds rupture one by one. Assuming that each hydrogen bond has a range of a_0 over which it acts and produces frictional resistance, the effective number of hydrogen bonds that participate in the frictional stress production on average is $N_{\mathrm{HB}}^{\mathrm{eff}} = N_{\mathrm{HB}}\, a_0/L$. This follows from the following reasoning, shown schematically in Fig. 4.4b: To break a length a_0 of the secondary structure, one needs to pull this much of the peptide through the region where the structure breaking takes place (open end of the β-hairpin); as the hydrogen bond density along the peptide is N_{HB}/L, one ends up breaking $a_0 N_{\mathrm{HB}}/L$ hydrogen bonds. Replacing N_{HB} by $N_{\mathrm{HB}}^{\mathrm{eff}}$ in Eq. 4.11 we obtain the modified friction law for the sequential unzipping of hydrogen bonded secondary structures

$$F_{\mathrm{f}} = \gamma_{\mathrm{HB}}\, v\, N_{\mathrm{HB}} \frac{a_0}{L}, \tag{4.14}$$

where N_{HB} is the total number of hydrogen bonds stabilizing a secondary structure consisting of a peptide strand of contour length L, while a_0 is the range over which each hydrogen bond contributes to the friction. Comparing Eq. 4.14 with the definition of the friction coefficient, $F_{\mathrm{f}} = \gamma\, v$, we obtain

$$\gamma_{\mathrm{HB}} = \frac{L\, \gamma}{a_0\, N_{\mathrm{HB}}}. \tag{4.15}$$

Using a hydrogen-bond range $a_0 = 0.2\,\mathrm{nm}$, in agreement with our previous result based on the velocity-dependent crossover of the hydrogen bond friction [96], the results for γ_{HB} for Ala_{15} and Gln_{15} in the α-helical and β-hairpin states are shown in Table 4.1. Compared to the hydrogen-bond friction coefficients in peptide bundles [98], we see that the obtained values for γ_{HB} roughly correspond to bundles formed from two to three peptides, which seems realistic in terms of the packing density and the water accessibility in the α-helical and β-hairpin states. The pronounced difference between polyalanine and polyglutamine can be rationalized by the fact that γ_{HB} includes the effects of varying hydrogen bond strength and bond-breakage cooperativity, which are influenced by steric details and the hydrophobic environment and thus depend on the amino-acid side-chain architecture. This is in line with the large spread in γ_{HB} we have observed in our previous work on friction in peptide bundles of different aggregation levels [98]. Apart from this sequence dependence, there seems to be very little additional influence of the secondary structure type, by this we mean that the difference between γ_{HB} for the α-helix and β-hairpin structures for either polyalanine or polyglutamine is rather small. We thus conclude that the dissipative work in different secondary structures can be satisfactorily explained by the concept of hydrogen bond friction. By the same reasoning, we expect the friction coefficients for different sequences to be different, but the comparison of friction coefficients for α-helical and β-hairpin states of identical or very similar sequences should show the same characteristic ratio we observe for polyalanine and polyglutamine sequences.

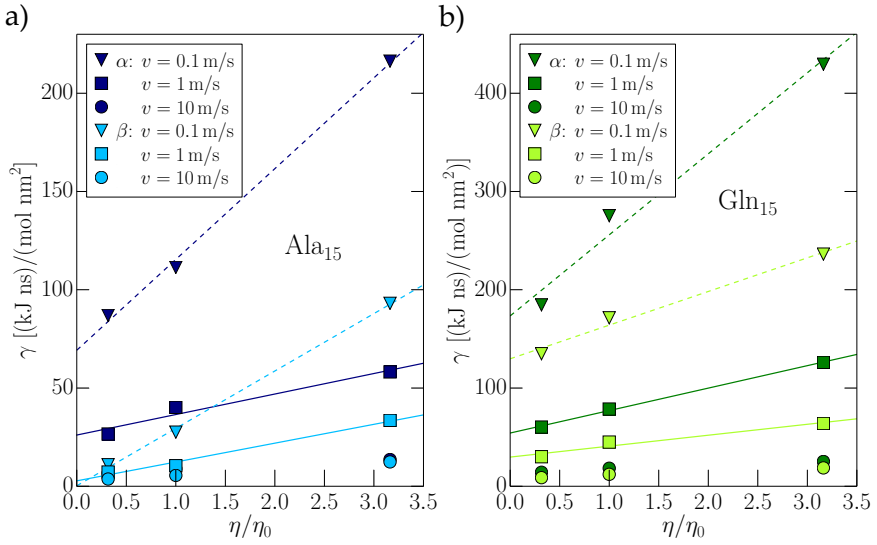

FIGURE 4.5: Friction coefficient as defined in Eq. 4.16 as a function of the rescaled solvent viscosity η/η_0 for different pulling speeds for a) Ala_{15} and b) Gln_{15}. The straight lines denote linear fits according to Eq. 4.17.

4.3.3 Dependence on solvent viscosity: internal versus solvation versus hydrodynamic friction

By changing the mass m of the water molecules in our MD simulations we can vary the solvent viscosity η without changing the free energetics [93]. Denoting the normal water mass by m_0, we performed additional stretching simulations of Gln_{15} and Ala_{15} starting in the α-helix and β-hairpin states for $m/m_0 = 1/10$ and $m/m_0 = 10$, giving rise to water viscosities $\eta/\eta_0 = 1/\sqrt{10}$ and $\eta/\eta_0 = \sqrt{10}$, respectively, where η_0 is the unmodified water viscosity. In particular the simulations at reduced water mass are very time consuming as they require substantially reduced simulation time steps, so we did not perform additional relaxation simulations at modified water mass. Rather, in this section we define the friction coefficient as

$$\gamma = \frac{W_{\text{str}}(v) - W_{\text{str}}(0)}{L\, v}, \tag{4.16}$$

where the stretching work at vanishing pulling speed $W_{\text{str}}(0)$ is determined by linear extrapolation for the data with standard water mass from the last section.

In Fig. 4.5a and b we show the friction coefficient γ as a function of η/η_0 for Ala_{15} and Gln_{15} for α-helix and β-hairpin states for a few different velocities. We fit the data with the heuristic linear law

$$\gamma(v) = \gamma_{\text{int}} + \gamma_\eta \eta/\eta_0, \tag{4.17}$$

where γ_{int} is the internal friction in the hypothetical limit of vanishing solvent viscosity, while the coefficient γ_η describes the solvent-related contribution to the total friction coefficient [15, 93]. Focusing the attention first on the Ala_{15} data in Fig. 4.5a we note that the slopes for α-helix (dark blue) and β-hairpin data (in light blue) are quite similar for the slowest velocity $v = 0.1\text{m/s}$, while the α-helix data are shifted upwards with respect

to the β-hairpin data. This would mean the helix structure has a higher internal friction than the hairpin, while the solvent contribution is roughly similar. The data for Gln_{15} in Fig. 4.5b exhibit a similar behavior, although the slopes of the curves in the α and β states differ somewhat. Nevertheless, the same trend is observed, and we conclude that the different friction coefficients seen for α and β states are primarily caused by a difference in the internal friction. This is in line with our interpretation of the dissipative effects in terms of the internal peptide-peptide hydrogen bonds. Since non-linear effects dominate the data for larger velocities, we only show those data without discussion.

4.3.4 Comparison with the equilibrium free energy landscape

In this section we compare the unfolding free energy, which in the previous sections has been estimated by extrapolation to vanishing pulling velocity, with the standard estimate based on the equilibrium folding free energy landscape. In this section, we in addition show simulations for a shorter Ala_8 peptide, as equilibration is less of an issue for this chain length.

If one is interested in comparing the folded and unfolded states, it is clearly of importance that both ensembles are reasonably sampled. In Fig. 4.6a we compare the velocity-dependent unfolding works $W_{\mathrm{unf}}(v)$ of the Ala_8 and Ala_{15} peptides. The lines show linear fits and the extrapolated equilibrium works $W_{\mathrm{unf}}(v = 0\,\mathrm{m/s})$ are quoted directly in the plot. In Fig. 4.6b the free energy landscape of the Ala_8 and Ala_{15} peptides are shown as a function of the rms deviation from the ideal helical state Q_{rms}. We adopt the simplest possible definition of the folded and unfolded ensemble and assume the peptide to be folded for $Q_{\mathrm{rms}} < Q^*_{\mathrm{rms}}$ and unfolded for $Q_{\mathrm{rms}} > Q^*_{\mathrm{rms}}$.

The unfolding free energy, i.e. the free energy difference between the folded and unfolded states, thus follows as

$$\Delta F = k_B\, T\, \ln \frac{\int_{-\infty}^{Q^*_{\mathrm{rms}}} P(Q_{\mathrm{rms}})\, \mathrm{d}Q_{\mathrm{rms}}}{\int_{Q^*_{\mathrm{rms}}}^{\infty} P(Q_{\mathrm{rms}})\, \mathrm{d}Q_{\mathrm{rms}}}. \tag{4.18}$$

Now assuming $\Delta F = W_{\mathrm{unf}}(v = 0\,\mathrm{m/s}) = 2.3\,\mathrm{kJ/mol}$ for Ala_8, as follows from the data in Fig. 4.6a, we obtain $Q^*_{\mathrm{rms}} = 0.26\,\mathrm{nm}$ for the folding free energy curve for Ala_8. We indicate this transition value by green and blue colors of the free energy data in the folded and unfolded ensembles in Fig. 4.6b, respectively. In Fig. 4.6c we show the distribution at the end of the relaxation simulations at a relaxation speed of $v = 0.1\,\mathrm{m/s}$ of 100 separate simulations of Ala_8, calculated as averages over the last 1 ns of these relaxation simulations (turquoise bars). We see that except a few relaxation simulations that basically refolded and are characterized by small values of Q_{rms}, most relaxation simulations indeed end in the unfolded free energy basin. We conclude that for Ala_8 the unfolding free energy estimates based on non-equilibrium simulations are consistent with the free energy landscape. In the non-equilibrium simulations, the unfolding free energy depends on the choice of the relaxation speed, in the free-energy landscape scenario the unfolding free energy depends on the arbitrarily chosen position of the division between the folded and unfolded basins.

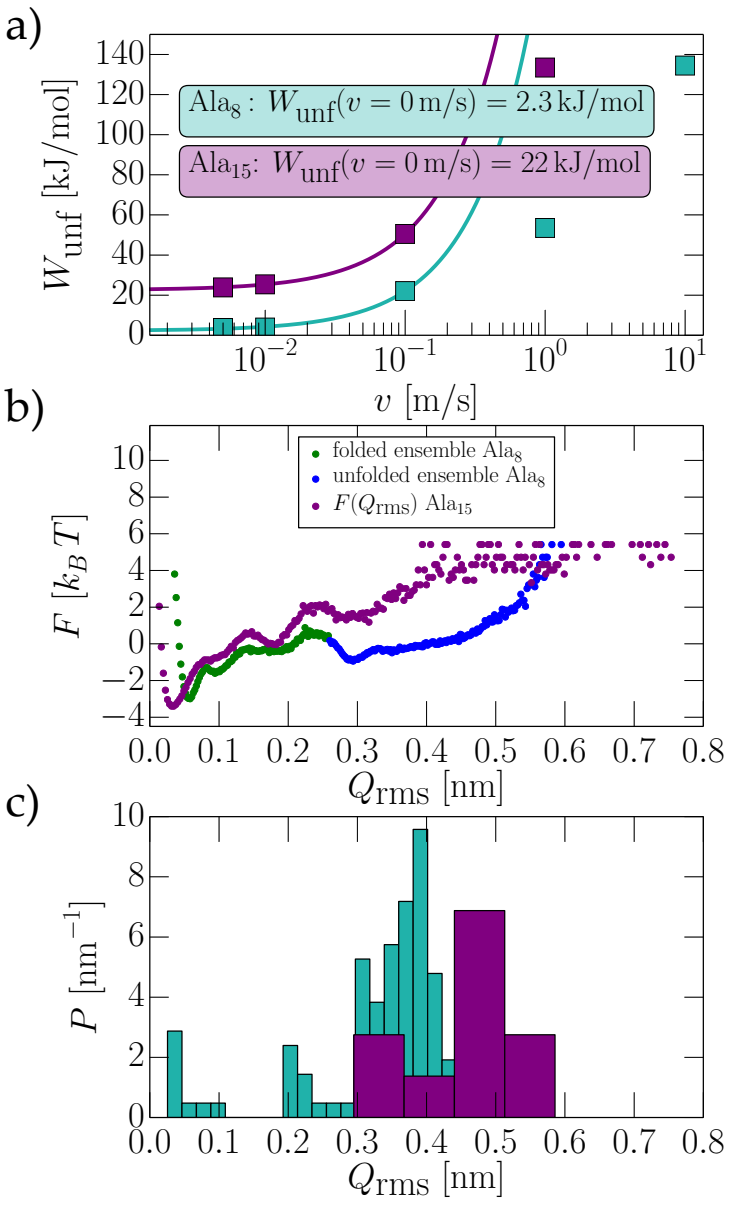

FIGURE 4.6: a) Unfolding work W_{unf}, defined in Eq. 4.7, of Ala_8 (turquoise squares) and Ala_{15} (purple squares) in the α-helical state as a function of the pulling speed v, lines show linear fits according to Eq. 4.8. b) Free energy landscapes for Ala_8 and Ala_{15} in terms of the deviation from the ideal helix state Q_{rms}. The division between the folded (green) and unfolded (blue) domains for Ala_8 is chosen such that the equilibrium folding free energy coincides with the extrapolated non-equilibrium unfolding work (see text). c) Probability distribution of the disordered ensemble from relaxation simulations with relaxation speed $v = 0.1\mathrm{m/s}$, obtained from the mean Q_{rms} over the last 1 ns of the relaxation trajectories. Results from 10 simulation trajectories for Ala_{15} (purple bars, corresponding to the yellow triangles in Fig. 4.2a) are compared with results from 100 simulation trajectories for Ala_8 (turquoise bars).

For Ala_{15} in the α-state, the transition state would be predicted to be $Q^*_{rms} = 0.71\,nm$ based on the extrapolated unfolding work $\Delta F = W_{unf}(v = 0\,m/s) = 23\,kJ/mol$, which is far in the unfolded regime where the free energy landscape is badly sampled, as can be seen in Fig. 4.6b. This is consistent with Fig. 4.6c where one sees that most of the relaxation runs of Ala_{15} indeed do end up in the unfolded and badly sampled part of the free energy landscape (purple bars). We conclude that while for Ala_{15} chains equilibration problems in the equilibrium simulations prevent a quantitative comparison of the non-equilibrium unfolding work with corresponding equilibrium results, for Ala_8 the two complementary approaches based on non-equilibrium pulling simulations and equilibrium free-energy landscape simulations lead to consistent results.

TABLE 4.1: Pulling/relaxation distance L for the different homopeptides and different secondary structures. Initial and final terminal group separations l_i and l_f. Unfolding work $W_{unf}(v = 0\,m/s)$ obtained from Eq. 4.8. Friction coefficients γ and γ_{HB} defined in Eq. 4.12 and Eq. 4.15. Total number of intra-peptide hydrogen bonds N_{HB} obtained from equilibrium simulations. * Number of hydrogen bonds for Ala_{15} in the β structure is based on the number of backbone-backbone hydrogen bonds of β-Gln15 , see text. N^{bb}_{HB} denotes the backbone-only hydrogen bonds.

	L [nm]	$l_i - l_f$ [nm]	$W_{unf}(0)$ [kJ/mol]	γ [10^{-10} kg/s]	N_{HB}	N^{bb}_{HB}	γ_{HB} [10^{-10} kg/s]
Ala15 α	2.7	2.1 – 4.8	23	1.7	9.2	9.2	2.5
Ala15 β	4.3	0.5 – 4.8	14	0.4	4.7*	4.7*	1.8
Gln15 α	3.1	1.8 – 4.9	29	3.5	13.6	9.2	4.0
Gln15 β	4.3	0.5 – 4.8	27	2.1	9.4	4.7	4.9

4.4 Conclusions

We pull with prescribed velocities on homopeptides that are folded into different secondary structures and by a velocity-dependent analysis decompose the total work of unfolding into its equilibrium and dissipative (frictional) contributions. The frictional dissipation is dominated by non-hydrodynamic effects and shown to correlate well with the total number of intra-peptide hydrogen bonds that stabilize the secondary structure. In agreement with the different number of hydrogen bonds that stabilize α-helices and β-hairpins, we find the friction of an α-helix to be larger than of a β-hairpin, irrespective of the equilibrium free energy difference. This finding is in qualitative agreement with recent conclusions from single-molecule experiments [99].

Let us demonstrate, with a simple thought experiment, the consequences these differences in friction coefficients could have on folding behavior. Assume an iso-free energetic situation, i.e., that the competing α-helix and β-hairpin states as well as the disordered state are all connected along a constant free energy surface. The times required for folding/unfolding the secondary structures then follow from the one-dimensional diffusion law, $t = 2\,L^2/D$, with the diffusion constant given by $D = k_B\,T/\gamma$. The length L here is the length (given in Table 4.1) over which the peptide needs to diffuse when folding/unfolding is taking place. For the Gln_{15} α-helix one obtains $t = 1.6\,\mu s$, for

the Gln_{15} β-hairpin 1.8 μs; for Ala_{15} we obtain 600 ns for the α-helix, and 360 ns for the β-hairpin.

This insight that, free energies permitting, folding to β-hairpins could actually be faster than to α-helices is particularly interesting in light of the prominent role played by β-sheet structures in neurodegenerative diseases, but is also relevant in other situations when α-helices convert to β-sheet structures [100]. In a non-equilibrium situation, i.e. when β-hairpin structures are sequestered into kinetically arrested fibrillar structures, the faster formation rate of β-hairpins could lead to the proliferation of β-rich fibrils, even when free-energetics by themselves would not prefer β structures. Actually, physiological chaperone rescue systems that are based on equilibrium free energy differences would not be of much help in this situation, since they are not susceptible to the faster formation rate of β structures. Whether this is related to the universal role of β-structures in neurodegenerative diseases is at this point nothing but an interesting speculation.

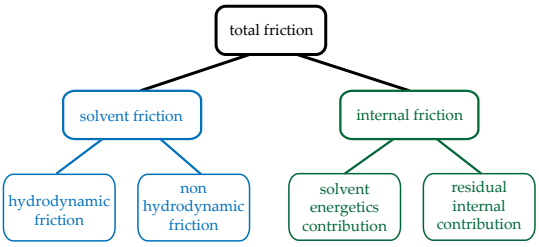

FIGURE 4.7: Schematic decomposition of the unfolding or folding friction into different contributions: At the upper level, the friction work splits into the internal friction (obtained in the hypothetical limit of vanishing solvent viscosity $\eta = 0$) and the solvent friction. The solvent friction is operationally subdivided into a hydrodynamic contribution and a contribution that cannot be described by the standard continuum hydrodynamic equations. The internal friction is operationally subdivided into a contribution that accounts for the equilibrium solvent influence on the internal friction, for example via long-ranged dielectric effects, and the residual internal frictional contribution.

Our results also shed light on a typical MD simulation protocol that probes the stability of competing protein folding structures: In a standard setup, one starts simulations in different states and watches which state becomes unstable first. Clearly, also in this situation, the unfolding rate results from the convolution of the free energy difference and the unfolding diffusivity or mobility. To put it in more drastic terms: In a situation where the free energy of two different conformations are identical, the rate at which a certain conformation unfolds will be solely determined by the diffusivity, and a discussion in terms of free-energetic stability alone would lead to a spurious interpretation. In such a situation, a separation into equilibrium and dissipative effects would be needed.

On a more fundamental level, we show by our simulations at varying solvent mass how the total friction of a peptide transformation can be formally decomposed into an internal friction part, obtained in the hypothetical limit of vanishing solvent viscosity, and the remaining contribution, which is due to the finite solvent viscosity. The solvent friction can be further decomposed into hydrodynamic friction, caused by the hydrodynamic drag of peptide parts when moving through solvent, which can be described

by the Navier-Stokes equation, and effects that can not be described by continuum hydrodynamic equations. Note that this distinction is purely operational but is useful when mechanisms of how solvent degrees of freedom influence folding rates shall be delineated and discussed. As an example of a non-hydrodynamic solvent friction effect we could think of hydration water that dynamically forms hydrogen bonds with peptide surface groups and thereby influences folding or unfolding rates. It should be clear that such effects could in principle be taken into account in generalized hydrodynamic descriptions where the hydration layer kinetics is described by modified hydrodynamic boundary conditions involving slip effects and increased surface viscosities, as was recently done in the context of electrohydrodynamic effects [101]. On the other hand, the internal friction contribution, which we can think of as being caused by viscous effects within the peptide due to intra-peptide hydrogen bonding but also salt-bridging as well as hydrophobic effects, will again be influenced by equilibrium solvent properties. One example of such an equilibrium influence would be the dielectric constant of water, which certainly will influence the Coulombic interactions within a protein and thereby also the kinetics of salt-bridge and hydrogen-bond formation. We therefore can decompose the internal friction into a part that is influenced by solvent equilibrium properties and a residual internal contribution (which again is a purely operational decomposition). We summarize the various contributions to the friction of protein folding in the schematic drawing in Fig. 4.7.

In our simulations, we have initiated the unfolding of the α-helix and β-hairpin states by pulling the peptides at their ends, thereby dictating a certain path in the multi-dimensional free energy landscape. Clearly, the actual equilibrium unfolding pathways of β-hairpins and α-helices are more varied, as recent experiments and modeling approaches show [102, 103]. While in the limit of vanishing pulling speed the prescribed pathway should not influence the results in a drastic manner, we admit that the actual pulling speeds employed in our simulation are still relatively large, so we point out that our results should be interpreted with care. One advantage of pulling at the peptide ends is that this scenario can be directly probed in experiments where peptide chains are linked to magnetic or optical tweezers. Likewise, while for the sequences used by us hydrogen bonds are dominant, for other sequences that are stabilized by strong hydrophobic bonds or salt bridges these additional interactions could also play an important role and constitute alternative mechanisms for internal friction, as was recently suggested [104].

In our simulations we used the Amber03 peptide forcefield in combination with the SPC/E water model. As a matter of fact, the choice of water model plays a rather minor role [105, 106], while different peptide force-fields were recently shown to sensitively influence the stabilities of competing folding states of a single Gln_{40} homopeptide chain [89]. This is not of concern to us, since our main focus in this paper is not on free energies. However, it remains to be tested to what degree the friction difference between α-helix and β-hairpin states is robust again force-field variations.

Chapter 5

Conclusions

In this thesis we investigate the effects of internal friction on peptide kinetics. In the first part we use mass scaling of the solvent (i.e. water) to prove that internal friction does indeed exist. If internal friction would not exist, the mean first passage times would scale according to Kramers law and not deviate for low viscosities. The possibility to investigate the viscous effects at lower viscosities than water's without manipulating the free energy landscape (e.g. by increasing temperature) is a unique feature of this method. The seperation into internal and solvent viscous effects is however not trivial and no simple model is available that can explain the experiments on that matter. The two most prominent heuristic models for internal friction are a linear model and a power law model. While the power law model fits nicely to some experiments and also to the GlySer model peptide in our simulations, it must be wrong for vanishing viscosities, since it would lead to infinitely fast folding speeds. In our work we present a simple model based on a Rouse chain that includes both the linear form and a form that can resemble a power law, thus including both models. The Rouse model includes friction by introducing a damping element between the monomers. While this does not account for more complex contributions to internal friction, such as hydrogen bonds, it exhibits a complex functional forms for the viscosity dependence of the internal timescale and can motivate the existence of the diverse findings in experiments. Using a locally resolved friction analysis we can also link the existence of internal friction to the number of hydrogen bonds. By the simple argument that by proper rescaling a friction profile can always be flattened whereas the hydrogen profile cannot, we argue that internal friction is however not linked to the number of hydrogen bonds but to their variation along the reaction coordinate. In this work we developed a tool for studying the influence of viscosity and quantifying internal frictions for small peptides. Considering the fast increase in computational power, simulations of larger proteins will become feasible and the method developed here can be readily applied in the future.

As the experiments by Jas et. al. [15] show, the functional form of the viscosity dependence between an α-helix and a β-hairpin differs, suggesting internal friction plays an important role in the formation of secondary structure. The transition from helical to hairpin state is particularly important in the context of neurodegenerative diseases. We developed a method to obtain unfolding work differences between these secondary

structures by forced unfolding via pulling at different finite speeds. By extrapolation to vanishing speed the unfolding work difference gives the free energetic difference of the two states. While this method is not particularly good for obtaining free energy values, it is uniquely suited for investigating dissipative effects and thus the differences in internal friction between secondary structures. Our choice of the homopeptides consisting of alanine and glutamine is motivated by their importance in neurodegenerative diseases. Our results show that the friction is always higher for the helical state. By comparing this with the number of hydrogen bonds in each state we show that the characteristic hydrogen bond number of the secondary structure is the major reason for the difference in internal friction. Considering that protein folding typically takes place in a non-equilibrium environment, the folding friction can become the determining factor for the probability of a certain conformer and it is thus important to not consider free energetics alone but also dissipative effects.

List of Publications

This thesis is based on the following papers, which have been published in peer-reviewed journals:

[i] **Julius C. F. Schulz**, Lennart Schmidt, Robert B. Best, Joachim Dzubiella, and Roland R. Netz. Peptide chain dynamics in light and heavy water: zooming in on internal friction. *Journal of the American Chemical Society*. **134.14** (2012): 6273–6279.

[ii] **Julius C. F. Schulz**, Markus S. Miettinen, and Roland R Netz. Unfolding and folding internal friction of β-hairpins is smaller than of α-helices. *Journal of Chemical Physics B*. **119:13**, 4565–74 (2015).

The following publication is not discussed in the present thesis:

[iii] **Klaus F. Rinne**, Julius C. F. Schulz, and Roland R. Netz. Dielectric Spectrum of Eight Residual Alanine: Impact of Peptide Secondary Structure and Hydration Shell Water. *accepted for publication in J. Chem. Phys.*

Bibliography

[1] Mulder, G. J. On the composition of some animal substances. *Journal für praktische Chemie* **16**, 15 (1839).

[2] Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662–666 (1958).

[3] Dill, K. A. & MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **338**, 1042–1046 (2012).

[4] Hofmeister, F. Ueber den Bau des Eiweißmoleküls. *Naturwiss. Rundschau* **17**, 529–545 (1902).

[5] Fischer, E. Autoreferat. *Chem. Ztg.* **26** (1902).

[6] Astbury, W. & Woods, H. t. X-ray studies of the structure of hair, wool, and related fibres. ii. the molecular structure and elastic properties of hair keratin. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 333–394 (1934).

[7] Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* **37**, 205–211 (1951).

[8] Bernstein, F. C. *et al.* The protein data bank. *European Journal of Biochemistry* **80**, 319–324 (1977).

[9] Gerstein, M. & Levitt, M. A structural census of the current population of protein sequences. *Proceedings of the National Academy of Sciences* **94**, 11911–11916 (1997).

[10] Kauzmann, W. Some factors in the interpretation of protein denaturation. *Advances in protein chemistry* **14**, 1–63 (1959).

[11] DeBrunner, J. & Munck, E. Mossbauer spectroscopy in biological systems: Proceedings of a meeting held in allerton house. *Monticello, Illinois* (1969).

[12] Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284–303 (1940).

[13] Ansari, A., Jones, C. M., Henry, E. R., Hofrichter, J. & Eaton, W. A. The role of solvent viscosity in the dynamics of protein conformational changes. *Science* **256**, 1796–1798 (1992).

[14] Bieri, O. *et al.* The speed limit for protein folding measured by tripletâĂŞtriplet energy transfer. *Proc. Natl. Acad. Sci. USA* **96**, 9597–9601 (1999).

[15] Jas, G. S., Eaton, W. A. & Hofrichter, J. Effect of viscosity on the kinetics of α-helix and β-hairpin formation. *J. Phys. Chem. B* **105**, 261–272 (2001).

[16] Manke, C. W. & Williams, M. C. Internal viscosity of polymers and the role of solvent resistance. *Macromolecules* **18**, 2045–2051 (1985).

[17] Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory. Computation.* **4**, 435–447 (2008).

[18] Miyamoto, S. & Kollman, P. A. Settle: an analytical version of the shake and rattle algorithm for rigid water models. *Journal of computational chemistry* **13**, 952–962 (1992).

[19] Hess, B., Bekker, H., Berendsen, H. J., Fraaije, J. G. *et al.* Lincs: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **18**, 1463–1472 (1997).

[20] Ewald, P. P. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik* **369**, 253–287 (1921).

[21] Darden, T., York, D. & Pedersen, L. Particle mesh ewald: An n log(n) method for ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

[22] Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

[23] Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

[24] Zwanzig, R. Time-correlation functions and transport coefficients in statistical mechanics. *Annu. Rev. Phys. Chem.* **16**, 67–102 (1965).

[25] Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys. Rev. Lett.* **56**, 930 (1986).

[26] Ashkin, A. & Dziedzic, J. Optical trapping and manipulation of viruses and bacteria. *Science* **235**, 1517–1520 (1987).

[27] Isralewitz, B., Gao, M. & Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Chem. Biol.* **11**, 224–230 (2001).

[28] User:Itub. Simple water models (2007). URL `https://en.wikipedia.org/wiki/File:Water_models.svg`. File: `Water_models.svg`.

[29] Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).

[30] Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012 (2003).

[31] Best, R. B., Buchete, N.-V. & Hummer, G. Are current molecular dynamics force fields too helical? *Biophy. J.* **95**, L07–L09 (2008).

[32] Tsong, T. Y. & Baldwin, R. L. Effects of solvent viscosity and different guanidine salts on the kinetics of ribonuclease a chain folding. *Biopolymers* **17**, 1669–1678 (1978).

[33] Matthews, C. Role of diffusion in the folding of the alpha subunit of tryptophan synthase from escherichia coli. *Biochemistry* **29**, 2149–2154 (1990).

[34] Plaxco, K. & Baker, D. Limited internal friction in the rate-limiting step of a two-state protein folding reaction. *Proc. Natl. Acad. Sci. USA* **95**, 13591–13596 (1998).

[35] Qiu, L. & Hagen, S. J. A limiting speed for protein folding at low solvent viscosity. *J. Am. Chem. Soc.* **126**, 3398–3399 (2004).

[36] Pabit, S. A., Roder, H. & Hagen, S. J. Internal friction controls the speed of protein folding from a compact configuration. *Biochemistry* **43**, 12532–12538 (2004).

[37] Cellmer, T., Henry, E. R., Hofrichter, J. & Eaton, W. A. Measuring landscape roughness in ultrafast folding kinetics. *Proc. Natl. Acad. Sci. USA* **105**, 18320–18325 (2008).

[38] Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Protein* **21**, 167–195 (1995).

[39] B G Wensley, e. a. Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* **463**, 685–688 (2010).

[40] Lapidus, L. J., Steinbach, P. J., Eaton, W. A., Szabo, A. & Hofrichter, J. Effects of chain stiffness on the dynamics of loop formation in polypeptides. appendix: Testing a 1-dimensional diffusion model for peptide dynamics. *J. Phys. Chem. B* **106**, 11628–11640 (2002).

[41] Nettels, D., Gopich, I., Hoffmann, A. & Schuler, B. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. USA* **104**, 2655–2660 (2007).

[42] Klimov, D. K. & Thirumalai, D. Viscosity dependence of the folding rates of proteins. *Phys. Rev. Lett.* **79**, 317–320 (1997).

[43] Zagrovic, B. & Pande, V. Solvent viscosity dependence of the folding rate of a small protein: Distributed computing study. *J. Comput. Chemistry.* **24**, 1432–1436 (2003).

[44] Best, R. B. & Hummer, G. Diffusive model of protein folding dynamics with kramers turnover in rate. *Phys. Rev. Lett.* **96**, 228104 (2006).

[45] Pastor, R. W. & Karplus, M. Parametrization of the friction constant for stochastic simulations of polymers. *J. Phys. Chem.* **92**, 2636–2641 (1988).

[46] Nguyen, P. H. Replica exchange simulation method using temperature and solvent viscosity. *J. Chem. Phys.* **132**, 144109 (2010).

[47] Lin, I. & Tuckerman, M. E. Enhanced conformational sampling of peptides via reduced side-chain and solvent masses. *J. Phys. Chem. B* **114**, 15935–15940 (2010).

[48] Best, R. B. & Hummer, G. Coordinate-dependent diffusion in protein folding. *Proc. Natl. Acad. Sci. USA* **107**, 1088–1093 (2010).

[49] Hinczewski, M., von Hansen, Y., Dzubiella, J. & Netz, R. R. How the diffusivity profile reduces the arbitrariness of protein folding free energies. *J. Chem. Phys.* **132**, 245103 (2010).

[50] Bazua, E. R. & Williams, M. C. Molecular formulation of the internal viscosity in polymer dynamics, and stress symmetry. *J. Chem. Phys.* **59**, 2858 (1973).

[51] de Gennes, P. G. Origin of internal viscosities in dilute polymer solutions. *J. Chem. Phys.* **66**, 5825 (1977).

[52] Alexander-Katz, A., Wada, H. & Netz, R. R. Internal friction and nonequilibrium unfolding of polymeric globules. *Phys. Rev. Lett.* **103**, 028102 (2009).

[53] Khatri, B. S., Kawakami, M., Byrne, K., Smith, D. A. & McLeish, T. C. B. Entropy and barrier-controlled fluctuations determine conformational viscoelasticity of single biomolecules. *Biophys. J.* **92**, 1825–1835 (2007).

[54] Khatri, B. S. & McLeish, T. C. B. Rouse model with internal friction: A coarse grained framework for single biopolymer dynamics. *Macromolecules* **40**, 6770–6777 (2007).

[55] Monticelli, L. *et al.* The martini coarse-grained force field: extension to proteins. *J. Chem. Theor. Comp.* **4**, 819–834 (2008).

[56] Wang, T., Zhu, Y. & Gai, F. Guiding the search for a protein's maximum rate of folding. *J. Phys. Chem. B* **108**, 3694–3697 (2004).

[57] Bryngelson, J. D. & Wolynes, P. G. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.* **93**, 6902–6915 (1989).

[58] Luzar, A. & Chandler, D. Hydrogen-bond kinetics in liquid water. *Nature* **379**, 55–57 (1996).

[59] Cheng, R. R., Uzawa, T., Plaxco, K. W. & Makarov, D. E. The rate of intramolecular loop formation in dna and polypeptides: The absence of the diffusion-controlled limit and fractional power-law viscosity dependence. *J. Phys. Chem. B* **113**, 14026–14034 (2009).

[60] Weiss, G. H. *First Passage Time Problems in Chemical Physics*, 1–18 (John Wiley & Sons, Inc., 2007).

[61] Ramsay, J. O., Hooker, G. & Graves, S. *Functional Data Analysis with R and MATLAB* (Springer, New York, 2009).

[62] Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. Lincs: A linear constraint solver for molecular simulations. *J. Comp. Chem.* **18**, 1463–1472 (2007).

[63] Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).

[64] Monticelli, L. *et al.* The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theor. Comp.* **4**, 819–834 (2008).

[65] Periole, X., Cavalli, M., Marrink, S. J. & Mark, A. E. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics and intermolecular recognition. *J. Chem. Theory Comput.* **5**, 2531–2543 (2009).

[66] Karanicolas, J. & Brooks, C. L. The origins of asymmetry in the folding transition states of protein L and protein G. *Prot. Sci.* **11**, 2351–2361 (2002).

[67] Wang, T., Zhu, Y. & Gai, F. Folding of a three-helix bundle at the folding speed limit. *J. Phys. Chem. B* **108**, 3694–3697 (2004).

[68] M U Johansson, M. d. C., Wikström, M., Forsén, S., Drakenberg, T. & Björck, L. Solution structure of the albumin-binding ga module: a versatile bacterial protein domain. *J. Mol. Biol.* **316**, 859–865 (2002).

[69] Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: a message passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* **91**, 43–56 (1995).

[70] Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317 (2001).

[71] Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50**, 1055–1076 (1983).

[72] Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).

[73] Perutz, M. F., Ladner, J. E., Simon, S. R. & Ho, C. Influence of globin structure on the state of the heme. i. human deoxyhemoglobin. *Biochemistry* **13**, 2163–2173 (1974).

[74] Henzler-Wildman, K. A. *et al.* Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838–844 (2007).

[75] Lange, O. F. *et al.* Recognition dynamics up to microseconds revealed from an rdc-derived ubiquitin ensemble in solution. *Science* **320**, 1471–1475 (2008).

[76] Smock, R. G. & Gierasch, L. M. Sending signals dynamically. *Science* **324**, 198–203 (2009).

[77] Selkoe, D. J. Folding proteins in fatal ways. *Nature* **426**, 900–904 (2003).

[78] Ross, C. A. & Poirier, M. A. What is the role of protein aggregation in neurodegeneration? *Nat. Rev. Mol. Cell Biol.* **6**, 891–898 (2005).

[79] Kubrycht, J., Sigler, K., Souček, P. & Hudeček, J. Structures composing protein domains. *Biochimie* **95**, 1511–1524 (2013).

[80] Almeida, B., Fernandes, S., Abreu, I. A. & Macedo-Ribeiro, S. Trinucleotide repeats: a structural perspective. *Front. Neurol.* **4**, 76 (2013).

[81] Winter, R., Liebold, J. & Schwarz, E. The unresolved puzzle why alanine extensions cause disease. *Biol. Chem.* **394**, 951–963 (2013).

[82] Brown, L. Y. & Brown, S. A. Alanine tracts: the expanding story of human illness and trinucleotide repeats. *Trends Genet.* **20**, 51–58 (2004).

[83] Wetzel, R. Physical chemistry of polyglutamine: Intriguing tales of a monotonous sequence. *J. Mol. Biol.* **421**, 466–490 (2012).

[84] Orr, H. T. & Zoghbi, H. Y. Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* **30**, 576–621 (2007).

[85] Scherzinger, E. *et al.* Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: Implications for Huntington's disease pathology. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4604–4609 (1999).

[86] Albrecht, A. & Mundlos, S. The other trinucleotide repeat: polyalanine expansion disorders. *Curr. Opin. Genet. Dev.* **15**, 285–293 (2005).

[87] Latouche, M. *et al.* Polyglutamine and polyalanine expansions in ataxin7 result in different types of aggregation and levels of toxicity. *Mol. Cell. Neurosci.* **31**, 438–445 (2006).

[88] Chen, S., Ferrone, F. A. & Wetzel, R. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11884–11889 (2002).

[89] Miettinen, M. S., Knecht, V., Monticelli, L. & Ignatova, Z. Assessing polyglutamine conformation in the nucleating event by molecular dynamics simulations. *J. Phys. Chem. B* **116**, 10259–10265 (2012).

[90] Levy, Y., Jortner, J. & Becker, O. M. Solvent effects on the energy landscapes and folding kinetics of polyalanine. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2188–2193 (2001).

[91] Davis, C. M., Xiao, S., Raleigh, D. P. & Dyer, R. B. Raising the speed limit for β-hairpin formation. *J. Am. Chem. Soc.* **134**, 14476–14482 (2012).

[92] Best, R. B. & Mittal, J. Microscopic events in β-hairpin folding from alternative unfolded ensembles. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11087–11092 (2011).

[93] Schulz, J. C., Schmidt, L., Best, R. B., Dzubiella, J. & Netz, R. R. Peptide chain dynamics in light and heavy water: zooming in on internal friction. *J. Am. Chem. Soc.* **134**, 6273–6279 (2012).

[94] Best, R. B. & Mittal, J. Free-energy landscape of the gb1 hairpin in all-atom explicit solvent simulations with different force fields: Similarities and differences. *Proteins: Struct., Funct., Bioinf* **79**, 1318–1328 (2011).

[95] Hugel, T., Rief, M., Seitz, M., Gaub, H. E. & Netz, R. R. Highly stretched single polymers: atomic-force-microscope experiments versus ab-initio theory. *Phys. Rev. Lett.* **94**, 048301 (2005).

[96] Erbas, A., Horinek, D. & Netz, R. R. Viscous friction of hydrogen-bonded matter. *J. Am. Chem. Soc.* **134**, 623–630 (2011).

[97] Best, R. B. & Mittal, J. Balance between α and β structures in ab initio protein folding. *J. Phys. Chem. B* **114**, 8790–8798 (2010).

[98] Erbaş, A. & Netz, R. R. Confinement-dependent friction in peptide bundles. *Biophys. J.* **104**, 1285–1295 (2013).

[99] Chung, H. S. & Eaton, W. A. Single-molecule fluorescence probes dynamics of barrier crossing. *Nature* **502**, 685–688 (2013).

[100] Zhmurov, A. *et al.* Mechanical transition from α-helical coiled coils to β-sheets in fibrin (ogen). *J. Am. Chem. Soc.* **134**, 20396–20402 (2012).

[101] Bonthuis, D. J. & Netz, R. R. Beyond the continuum: How molecular solvent structure affects electrostatics and hydrodynamics at solid–electrolyte interfaces. *J. Phys. Chem. B* **117**, 11397–11413 (2013).

[102] Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. Folding dynamics and mechanism of β-hairpin formation. *Nature* **390**, 196–199 (1997).

[103] De Sancho, D. & Best, R. B. What is the time scale for α-helix nucleation? *J. Am. Chem. Soc.* **133**, 6809–6816 (2011).

[104] de Sancho, D., Sirur, A. & Best, R. B. Molecular origins of internal friction effects on protein-folding rates. *Nat. Commun.* **5**, 4307 (2014).

[105] Hegefeld, W. A., Chen, S.-E., DeLeon, K. Y., Kuczera, K. & Jas, G. S. Helix formation in a pentapeptide: Experiment and force-field dependent dynamics. *J. Phys. Chem. A* **114**, 12391–12402 (2010).

[106] Hu, Z. & Jiang, J. Assessment of biomolecular force fields for molecular dynamics simulations in a protein crystal. *J. Comp. Chem.* **31**, 371–380 (2010).

Abstract

Protein dynamics are significantly influenced by frictional effects, not only from the surrounding solvent but also due to interactions inside the protein chain itself. Experimentally, such "internal friction" has been investigated by studying folding or binding kinetics at varying solvent viscosity; however the molecular origin of these effects is hard to pinpoint. In this thesis, we studied the effects of internal friction by both equilibrium and non-equilibrium molecular dynamics simulations.

We developed a mechanism using scaled solvent mass to probe more than two orders of magnitude in viscosity without altering the free energy landscape of the protein. Our method is especially suited to investigate peptide kinetics near vanishing viscosities that are not reachable experimentally. While previous experimental studies have suggested different functional forms for the viscosity dependence, our findings suggest that solvent and internal friction effects are intrinsically entangled. This finding is rationalized by calculation of the polymer end-to-end distance dynamics from a Rouse model that includes internal friction. While this simple Rouse model does not include effects such as hydrogend bonds, an analysis of the local friction profile along different reaction coordinates suggests a connection between friction and the formation of hydrogen bonds upon folding.

Since hydrogen bonding is a major factor for the determination of secondary structure, internal friction can help in understanding the folding process. As the secondary structure is of vast importance for the biological function of a protein, misfolding is thought to be the explanation for many diseases including neurogedegenerative ones such as Huntington's or Parkinon's disease. By the forced unfolding of polyglutamine and polyalanine homopeptides in competing α-helix and β-hairpin secondary structures, we disentangle equilibrium free-energetics from non-equilibrium dissipative effects. We find that α-helices are characterized by larger friction or dissipation upon unfolding, regardless of whether they are free-energetically preferred over β-hairpins or not. Our analysis, based on MD simulations for atomistic peptide models with explicit water, suggests that this difference is related to the internal friction and mostly caused by the different characteristic number of intra-peptide hydrogen bonds in the α-helix and β-hairpin states, which is higher for the α-helical state.

Kurzfassung

Die Proteindynamik wird entscheidend von Reibungseffekten beeinflußt, die nicht nur vom umgebenden Lösungsmittel herrühren, sondern auch durch die Wechselwirkungen innerhalb der Proteinkette bestimmt werden. Experimentell ist diese interne Reibung durch Faltungs- oder Bindungskinetik bei verschiedenen Lösungsmittelviskositäten untersucht worden; allerdings ist die molekulare Ursache dieser Effekte nicht gut verstanden. In dieser Arbeit haben wir die Effekte von interner Reibung sowohl mit Gleichgewichts- als auch Nichtgleichgewichts-Molekulardynamiksimulationen genauer untersucht.

Wir haben eine Methode entwickelt, um mittels skalierter Lösungsmittelmasse die Viskosität zwei Größenordnungen zu variieren ohne dabei die Freie-Energie-Landschaft des Proteins zu verändern. Unsere Methode ist besonders geeignet, um die Peptidkinetik bei nahezu verschwindenden Viskositäten zu untersuchen, die experimentell nicht zugänglich sind. Experimentelle Studien haben unterschiedliche funktionelle Abhängigkeiten von der Viskosität vorgeschlagen, wobei unsere eigenen Ergebnisse nahelegen, dass die internen Reibungseffekte und die des Lösungsmittels intrinsisch miteinander verknüpft sind. Dieses Ergebnis unterlegen wir mit der Berechnung der End-zu-End-Entfernungs-Dynamik von Polymeren mittels eines Rouse-Modells, welches interne Reibungseffekte beinhaltet. Während dieses Modell allerdings keine Effekte wie Wasserstoffbrückenbindungen enthält, zeigt eine Analyse des lokalen Reibungsprofils entlang verschiedener Reaktionskoordinaten, dass eine Verbindung zwischen Reibung und der Bildung von Wasserstoffbrückenbindungen besteht.

Da die Bildung von Wasserstoffbrückenbindungen allerdings ein entscheidender Faktor bei der Bildung von Sekundärstrukturen ist, kann interne Reibung zum Verständnis des Faltungsprozesses beitragen. Da die Sekundärstruktur von großer Bedeutung für die biologische Funktion eines Proteins ist, wird eine Mißfaltung als Erklärung für viele Krankheiten betrachtet, darunter auch verschiedene neurodegenerative Krankheiten wie Chorea Huntington oder Morbus Parkinson. Mittels erzwungener Entfaltung von Polyglutamin- und Polyalanin-Homopeptiden in miteinander konkurrierenden α-Helix- und β-Schleifen-Strukturen können wir die Gleichgewichts-Freie-Energie-Effekte von den dissipativen Nichtgleichgewichts-Effekten trennen. Unsere Ergebnisse zeigen, dass α-Helices duch eine größere Reibung bzw. Dissipation bei der Faltung charakterisiert sind, unabhängig, ob sie, gemessen an der freien Energie, gegenüber der β-Schleife bevorzugt sind. Basierend auf Molekulardynamiksimulationen von atomistischen Peptidmodellen, legen unsere Analysen nahe, dass dieser Unterschied mit der internen Reibung zusammenhängt und vor allem durch eine unterschiedliche, charakteristische Anzahl von Intra-Peptid-Wasserstoffbückenbindungen des α-helikalen und des β-Schleifen-Zustands bestimmt ist, wobei die Anzahl größer für den helikalen Zustand ist.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertationsschrift mit dem Titel

Internal Friction in Peptide Kinetics

selbständig angefertigt und hier für keine anderen als die angegebenen Hilfsmittel verwendet habe. Die Arbeit ist weder in einem früheren Promotionsverfahren angenommen noch als ungenügend beurteilt worden.

Berlin, 28. April 2015

Lebenslauf

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.

Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während meiner Promotion unterstützt haben. Mein größter Dank gilt meinem Doktorvater, Herrn Prof. Dr. Roland Netz, von dem ich viel lernen durfte, der mir große Freiheit bei meiner Forschung ließ und mich auf die interessanten Themen dieser Dissertation ausmerksam machte. Mein Dank gilt an dieser Stelle auch Prof. Dr. Joachim Dzubiella, bei dem ich, noch in München, meine Dissertation beginnen durfte.

Für die erfolgreiche Zusammenarbeit bei den Teilprojekten dieser Arbeit danke ich den Mitautoren meiner Veröffentlichungen Lennart Schmidt, Joachim Dzubiella, Robert Best und Markus Miettienen. Ich danke dem HPC Cluster der ZEDAT und Jens Dreger vom Fachbereich Physik für die Bereitstellung der Rechenkapazität, ohne die diese Arbeit nicht möglich gewesen wäre.

Allen aktuellen und ehemaligen Mitgliedern der Arbeitsgruppe Netz danke ich für die tolle Atmosphäre sowohl bei der wissenschaftlichen Arbeit als auch abseits des Arbeitsalltags. Besonders hervorheben möchte ich an dieser Stelle meine Kollegen Klaus Rinne und Alexander Schlaich, die mir an unzähligen Stellen bei den kleinen Problemen des Wissenschaftsalltags geholfen haben. Weiterhin möchte ich mich bei Kai von Krbek für das Korrigieren meiner Arbeit und seine Motivationshilfe bedanken.

Vor allem aber danke ich meinen Eltern und meinen Geschwistern Reimar und Lena von Herzen für die Unterstützung und das Verständnis für meine wissenschaftliche Begeisterung.