Bachelor Thesis

Calculation of Peptide Persistence Length from Molecular Dynamics Simulations

Bestimmung der Persistenzlänge von Peptiden durch Molekulardynamiksimulationen

Verfasser: Cornelius Müthel
Abgabedatum: 30.07.2012
Erstgutachter/Betreuer: Prof. Dr. Helmut Grubmüller
Zweitgutachter: Prof. Dr. Jörg Enderlein

Contents

1 Introduction

The persistence length is a frequently used measure of polymer stiffness. The higher the persistence length of a polymer, the lower is its flexibility. A raw spaghetti in a pot of water, for example, has a very high persistence length. In contrast a cooked spaghetti in water has a low persistence length.

A more precise definition is given via correlation of chain segments. Two segments are correlated, if the motion of one segment influences the motion of the other. Two points on a raw spaghetti are highly correlated, because motion of one segment always means motion of all other segments. For the cooked spaghetti, in contrast, two points can move independently, if their distance is large enough. They are therefore less correlated. The averaged correlation of segments over the chain contour is expected to decrease exponentially. This exponential decay can be derived from the assumption, that the polymer performs Brownian motion in solvent. The persistence length is defined as the characteristic length, after which this correlation decreases to its e-th part.

Two common ways to measure the persistence length are important. First, it can be determined by directly observing the polymer's equilibrium motion. For example, one can mark certain points over polymer contour and extract persistence length from measured trajectory of these points. Second, one can pull at both ends of the polymer measuring the force-extension curve. This can be done by atomic force microscopy.

For most polymers, force-extension data from the latter method agrees well with the worm-like chain (WLC) model [1]. The WLC model describes a polymer as a continuous elastic rod having a constant length [2, p. 316 – 317]. For proteins, force-stretching measurements match well the predicted WLC behavior [3, 4]. Therefore proteins are often described as WLCs.

Proteins can fold to highly efficient nano machines, which play important roles in biological systems. Protein folding is a complex process, which includes hydrophobicity and side chain interactions. It is not yet understood in all detail. Therefore proteins are from a biophysical point of view especially interesting.

The complexity of protein folding makes proteins difficult to be described physically. The WLC model can not describe a folded protein, because it does not consider side chains or electrostatics. Therefore the persistence length of a folded protein can not be extracted from equilibrium measurements using the WLC model.

There is, however, a group of proteins, which do not adopt a unique fold under native conditions. These proteins perform their biological function from a non-folded conformation and are often called intrinsically disordered proteins (IDPs). We here ask, whether we can extract the persistence length of an IDP from equilibrium describing it as a WLC.

To address this question a 20-amino-acid peptide, which occurs in an IDP sequence, is investigated in this thesis. From equilibrium all-atom molecular dynamics (MD) simulations, the peptide's persistence length has been determined using the WLC model. The result will be compared to a recent measurement from force-stretching experiments [5]. Agreement of the two results would be a hint, that the peptide can be successfully described as a WLC. Disagreement would mean, that generally equilibrium measurements of this peptide using the WLC model do not yield a correct value for its persistence length.

As we expected marked deviations from the WLC model, we next asked, which properties of the peptide cause these deviations. To that aim, we considered long-range interactions, electrostatics and effects of side chains, which are not described by the WLC model.

Contributions of long-range interactions and electrostatics to the peptide's equilibrium persistence length can be investigated performing simulations, which do not calculate long-range interactions and electrostatics. The persistence length will be compared then to the persistence length obtained from standard simulations.

From recent stretching simulations, which did also not calculate long-range interactions and electrostatics, a value for the persistence length of our peptide has been determined [6]. We will finally compare this value to our equilibrium persistence length.

To investigate contributions of the peptides side chains to the effectively obtained equilibrium persistence length, simulations of a 20-Glycine chain have been performed. This chain has the same length like the investigated peptide, but no side chains.

The here investigated naturally unfolded peptide is a so called phenylalanine-glycine repeat (FG-repeat), which occurs in nuclear pore complexes (NPCs). NPCs are large protein complexes comprised of 30 distinct proteins called nucleoporins (Nups). 13 of the Nups contain FG-repeats. These Nups are IDPs, which are highly flexible and lack secondary structure [7]. The FG-repeat, which is investigated during this thesis, is an repeated motif of the cNup153 molecule with sequence

SDTSKPAFSFGAKPDEKKDS.

2 Theory

Several models describe the structure of polymers. From literature, there are at least two different definitions of the persistence length, which depend on the used chain model.

First, Paul Flory gives a definition of persistence length for an infinitely long discrete chain consisting of identical segments.

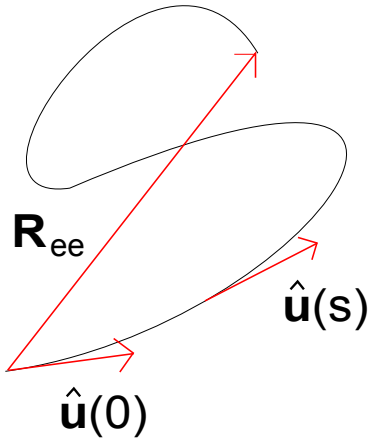

Figure 1 – The worm-like chain model

The chain segments are connected by bond vectors. Here the persistence length is defined as the average sum of scalar products of all bond vectors $j \geq i$ with bond vector i [8, p. 111].

Second, there is the worm-like chain (WLC) model [2, p. 316 – 317], which will be solely discussed in this thesis. The WLC model is also referred as the Kratky-Porod model. This model gives a definition of persistence length for a continuous chain having a finite length. In the limit of an infinitely long WLC and after its discretization, the two definitions of persistence length from WLC model and Paul Flory's model become equivalent [9].

A drawing of the WLC model is shown in figure 1. The continuity is due to an infinitely large number of chain segments each having an infinitesimally small segment length, while the chain's contour length $L = \int_0^L \mathrm{d}s$ is kept constant. The chain is described using the tangent vector

$$\mathbf{u}(s) = \frac{\partial \mathbf{R}}{\partial s}.$$

The WLC model only considers the chain's internal bending

energy

$$U_{\text{bend}} = \frac{1}{2} E \int_0^L \mathrm{d}s \left(\frac{\partial \mathbf{u}}{\partial s} \right)^2$$

with a persistence length $P = E/k_B T$.

The persistence length is interpreted as a measure of the chain's stiffness. The meaning of stiffness in this context can be emphasized discussing the normalized correlation function

$$C(s) = \langle \hat{\mathbf{u}}(s) \cdot \hat{\mathbf{u}}(0) \rangle.$$

This function is assumed to decay exponentially, which can be derived assuming the polymers motion in solvent being Brownian. The rate of decay is scaled by the persistence length, which yields the equation

$$\langle \hat{\mathbf{u}}(s) \cdot \hat{\mathbf{u}}(0) \rangle = \exp(-s/P). \tag{1}$$

Hence, a large persistence length means directional accordance of averaged tangent chain vectors over large distances on the chain's contour. A large persistence length is therefore an indication for stiffness. In contrast, a small persistence length results from low directional tangent vector correlation, which means high flexibility.

The relationship between persistence length P and the averaged squared end-to-end distance $\langle \mathbf{R}_{ee}^2 \rangle$ can be derived directly from the definition

$$\langle \mathbf{R}_{ee}^2 \rangle = \left\langle \left(\int_0^L \mathrm{d}s \; \hat{\mathbf{u}}(s) \right) \left(\int_0^L \mathrm{d}s' \; \hat{\mathbf{u}}(s') \right) \right\rangle.$$

Using equation 1, integration leads to

$$\langle \mathbf{R}_{ee}^2 \rangle = 2LP - 2P^2 \left[1 - \exp\left(-\frac{L}{P} \right) \right]. \tag{2}$$

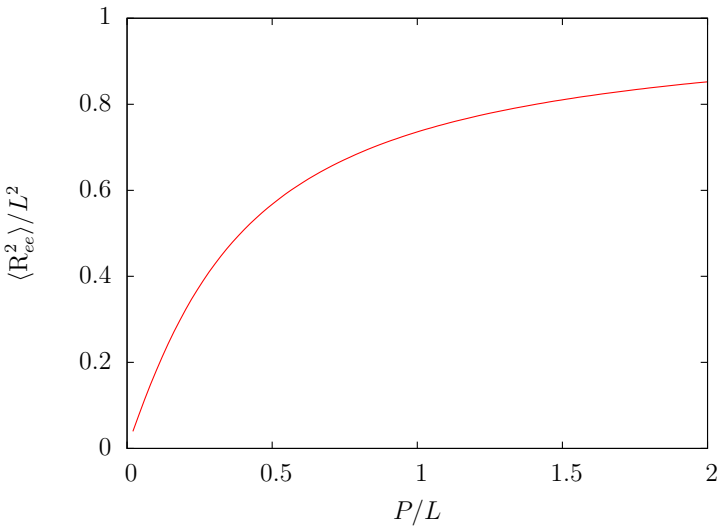

Figure 2 – Relation between persistence length and end-to-end distance

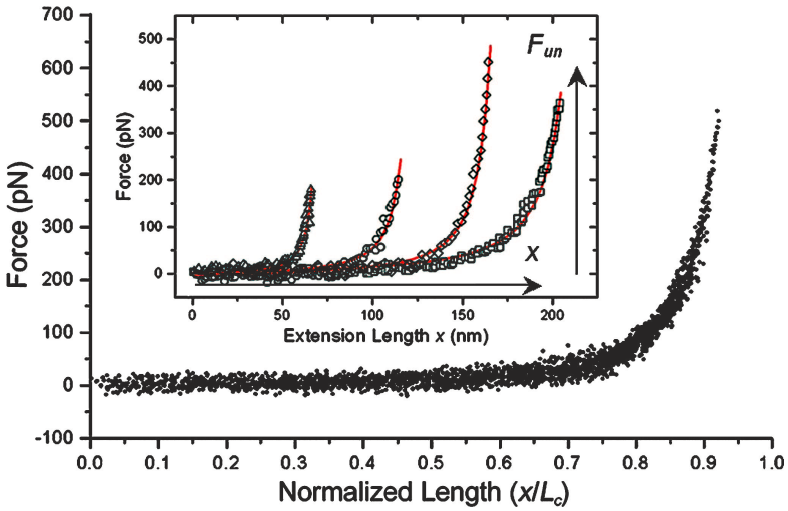

Figure 3 – An example of protein stretching measurement. This plot is taken from Lim et al. [5] and shows stretching measurements for cNup153 molecules. Fits were done from equation 3.

Figure 2 shows the discussion of equation 2. For values $P/L << 1$ the linear relation $\langle \mathbf{R}_{ee}^2 \rangle = 2LP$ is found, which is called 'random flight limit'. The 'rigid rod limit' for $P \to \infty$ is $\langle \mathbf{R}_{ee}^2 \rangle = L^2$.

An example of force stretching measurements is given from figure 3. Marko and Siggia [1] have derived a force-extension formula from the WLC model. They considered not only the internal bending energy, but also an affecting force, which yields the stretching potential

$$U_{\text{stretch}} = \frac{1}{2} E \int_0^L \mathrm{d}s \left(\frac{\partial \mathbf{u}}{\partial s} \right)^2 - Fz.$$

Here F is the stretching force and $z = \hat{\mathbf{z}} \mathbf{R}_{ee}$ the chain extension. From this potential, an approximate interpolation formula

$$F(z) = \frac{k_B T}{P} \left[\frac{z}{L} + \frac{1}{4(1 - z/L)^2} - \frac{1}{4} \right] \tag{3}$$

was derived. This equation is asymptotically exact for large- and small-force limits.

3 Methods

3.1 Analyzing Effects of Long-Range Interactions, Electrostatics and Side Chains on Persistence Length

To investigate long-range and hydrophobic contributions to the persistence length, two different simulation types were performed. The first type is a standard MD simulation, which is in the following called *full simulation*. The other type calculates particle trajectories ignoring long-range interactions and electrostatics. This simulation type is in the following called *modified simulation*.

Figure 4 shows the two types of potential curves, effectively calculated during MD simulations. During full simulations the

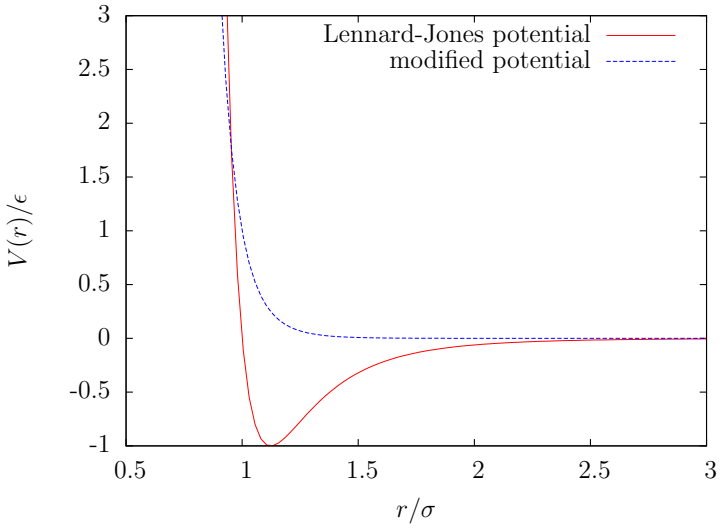

Figure 4 – Potentials of full simulation (red line) and modified simulation (dashed blue line)

usual Lennard-Jones potential and electrostatics from the Coulomb potential were calculated. At modified simulations, the potential $V(r) = \epsilon(\sigma/r)^{12}$ was used instead of the Lennard-Jones potential and electrostatics were not calculated.

The modified potential only includes the Lennard-Jones repulsion term. Therefore this simulation type "ignores nonlocal interactions and solvent-induced effects", but still "incorporates correct volume exclusion, backbone geometry and conformational freedom" [10].

To investigate side chain contributions to the persistence length, two different chains were simulated, which are shown in figure 5. The 20-Glycine chain should imitate the FG repeat without sidechains. It is in the following called *backbone chain*. The original FG-repeat chain including all side chains is called *complete chain*.

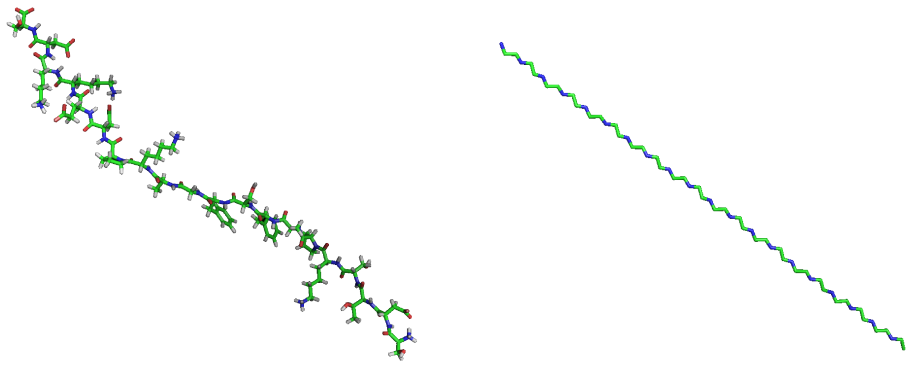

Figure 5 – Complete chain (left) and backbone chain (right) at $t = 0$. Both chains were constructed using the software SYBYL-X 1.0.

3.2 Calculation and Error Estimation of Persistence Length

To extract the persistence length from our simulations, both equations 1 and 2 were used. From the first, the persistence length was extracted via an exponential fit to calculated correlation values over chain contour. The persistence length from the second equation can be obtained calculating the averaged end-to-end distance from peptide trajectory.

For the calculation of persistence length from equation 1, the 19 bond vectors connecting the 20 amino acids were used. A bond vector is here defined as the connection between each two C_α backbone atoms. Thus, 19 correlation values

$$C_i = \langle \mathbf{u}_0 \cdot \mathbf{u}_i \rangle \qquad (i = 0,\ 1,\ ...,\ 18)$$

were computed, at which for all the 19 scalar products a time average with statistical error was calculated.

The bond vector length $L/19$ was used. The contour length $L = 6.631$ nm was taken from Gräter et al. [6]. This value has a deviation smaller than 1% and was therefore used without statistical error.

Finally $\exp(-ks)$ fits were done to the correlation values C_i. The exponential decay rate k is equal to P^{-1}. Because relatively small deviations of values C_1 and C_2 were found, which are due to local bond geometry, non-weighted fits were calculated. The statistical error of persistence length was calculated using the propagation error $\sigma_P = \sigma_k|1/k^2|$.

To calculate persistence length from equation 2, the averaged end-to-end distance $\langle \vec{R}_{ee}^2 \rangle$ and its statistical error were extracted from trajectory file. The equation was then solved numerically to obtain P. Here $\langle \vec{R}_{ee}^2 \rangle$ was subtracted from both sides of equation 2 and the root of the resulting function $f(P)$ was calculated.

To find an estimation of the statistical error for the here obtained persistence length, we used the general error propagation

$$\sigma_f = \left| \frac{\mathrm{d}f}{\mathrm{d}P} \right| \sigma_P.$$

The function $f(P)$ has an error σ_f, which is equal to the calculated mean squared error of $\langle \vec{R}_{ee}^2 \rangle$. The derivative $\mathrm{d}f/\mathrm{d}P$ was calculated analytically from equation 2.

3.3 Estimation of Equilibration Time

As it is shown in figure 5, our peptides were adjusted to an elongated conformation. This was done to estimate the minimum required simulation box size.

Using the software SYBYL-X 1.0 an energy minimization was performed. This minimization translated the system to a very local minimum in energy landscape. The peptides in figure 5 are shown *after* this minimization.

Their still elongated chain conformations are energetically not reasonable. Therefore, we expected the system to translate to a minimum, which is lower in energy landscape, during our simulations.

This translation is due to the artificially adjusted elongated initial conformation, which does not represent natural conditions. We therefore tried to estimate the required translation time. After estimation of this time, we excluded the trajectory during this time from evaluation.

We call this translation time here equilibration time. For a disordered peptide, however, it is not sure, whether a unique equilibrium conformation exists. 'Equilibration time' is therefore a mistakable term. Rather the time is meant here, after which the unnatural initial conformation has no more effects on the peptide dynamics.

To estimate this equilibration time, we performed a number of same simulations having different initial velocity distributions. Four different (Maxwell-Boltzmann) distributions were used, which are referred as A, B, C and D in the following.

The required equilibration time was then estimated qualitatively, analyzing root-mean-square deviation (RMSD) plots. The RMSD was calculated from the formula

$$RMSD(t) = \sqrt{\frac{1}{N^2} \sum_{i,j} |\mathbf{r}_{ij}(t) - \mathbf{r}_{ij}(0)|^2}.$$

The summation up to N was calculated for 295 and 143 atoms of complete chain and backbone chain, respectively. A low RMSD fluctuation over time is an indication, that the peptide fluctuates around a local minimum in energy landscape.

Not only the RMSD plots, but also plots of end-to-end distance were helpful at the estimation of equilibration time. A low fluctuating end-to-end distance is also a hint for a temporary stable conformation.

3.4 Performed Simulations

Gräter et al. [6] have also performed MD simulations of the investigated FG-repeat. To ensure a valid comparison of results

force field	OPLS/AA
water model	TIP4P
solvent	water salt concentration = 0.1 M
ensemble	NpT
temperature	Nosé-Hoover thermostat $T = 300$ K, coupling time $\tau_T = 0.1$ ps
pressure	Berendsen barostat $p = 1$ bar, coupling time $\tau_p = 0.1$ ps compressibility $= 4 \cdot 10^{-5}$ bar^{-1}
Lennard-Jones potential	cut-off radius = 1.4 nm
electrostatics (full simulation)	$r < 1.0$ nm: explicit Coulomb potential $r > 1.0$ nm: particle-mesh Ewald summation
electrostatics (modified simulation)	disabled
integration step	dt = 0.002 ps

Table 1 – Simulation setup

from their simulations to our results, same simulation parameters were chosen. These parameters are shown in table 1.

All MD simulations were carried out using the software package GROMACS 4.5.4 [11]. The OPLS/AA force field [12] and the TIP4P [13] water model were used. Table 2 gives an overview of the different performed simulations.

4 Results and Discussion

4.1 The Peptide Persistence Length can not be Extracted From Equilibrium Using the WLC Model

Full simulations did not yield consistent values of persistence length from equations 1 and 2. Values of persistence length extracted from correlation fits are systematically larger than those from averaged end-to-end distance. To have an overview of different obtained values for the persistence length, it is advisable

chain	simulation	velocity distribution	duration
complete	full	A	100 ns
		B	100 ns
		C	100 ns
		D	260 ns
complete	modified	A	100 ns
		B	100 ns
		C	100 ns
backbone	full	A	100 ns
backbone	modified	A	100 ns
		B	100 ns
		C	100 ns

Table 2 – Overview of performed simulations

to consult table 3 on page 21.

Figure 6 shows a representative plot of the calculated correlation values at full simulations. These values do not show the theoretically predicted exponential decay. Instead, a random cluster of values between -1 and 1 occurs. Because of values, which are significantly smaller than zero, an exponential behavior can be excluded.

Fitting an exponential to the obtained data, is obviously not reasonable. The persistence length can therefore not be extracted from our simulations using equation 1. The obtained results in table 3 must therefore be refused.

The relation between persistence length and end-to-end distance is derived using the exponential decay of correlation value. This exponential decay does effectively not exist, which explains the deviation between results from the two equations.

Moreover, the result extracted from end-to-end distance calculation can be excluded theoretically. A persistence length of less than one angstrom would mean, that neighboring amino acids were totally uncorrelated. This is, in fact, not the case

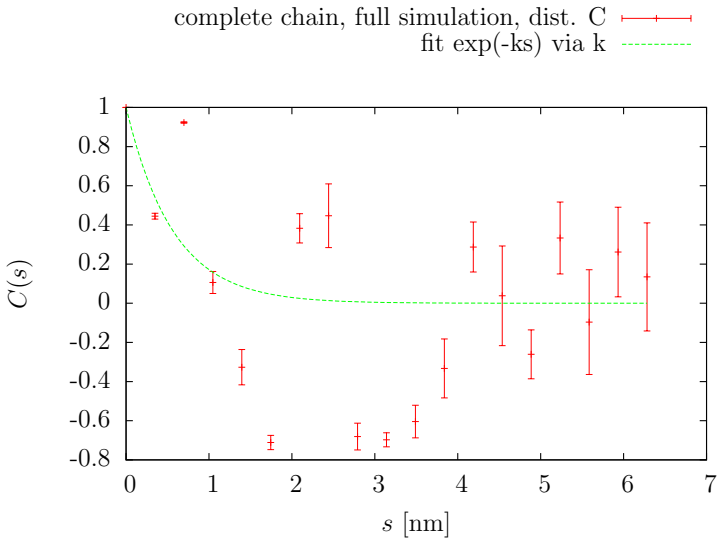

Figure 6 – Correlation values from full simulation, velocity distribution C

because of bond constraints, which were explicitly calculated by our MD simulations.

4.2 Mismatch of the WLC Model is due to Electrostatics or Long-range Interactions

To perform a first qualitative analysis, the peptide's trajectory was visualized. At full simulations the peptide was observed to collapse for all four velocity distributions. The peptide then adopted different globule conformations, which occasionally changed. Salt bridges often occurred, which influenced the peptides dynamics.

In contrast, at modified simulations no enduring globule conformations were adopted. Instead, the peptide collapses and elongates alternately.

These observations can be confirmed, comparing the end-to-end distance plots from figure 7. At full simulations (left) the peptide does not elongate to conformations with end-to-end dis-

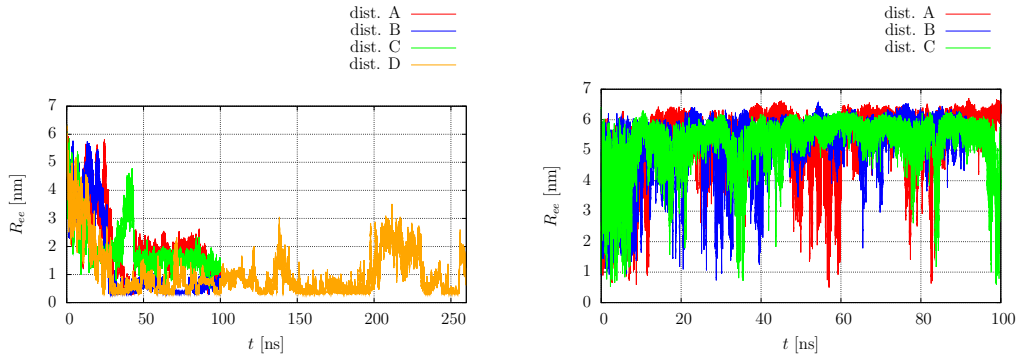

Figure 7 – Plot of R_{ee} from full simulations (left) and modified simulations (right)

tances longer than half its contour length after initial collapse. The very short end-to-end distances are due to the observed globule conformations.

At modified simulations (right), very short end-to-end distances do not occur enduring. Instead, the plot confirms the observed alternation of collapsed and elongated states over time.

The occurring globule conformations obviously cause the deviations from the WLC model. From the snapshot of a globule state from full simulations in figure 8, one can explain the significantly negative correlation values from the plot in figure 6.

To calculate the correlation, we defined vectors connecting two neighboring C_α atoms. The snapshot shows, that some of these vectors show in opposite directions of the first one $\mathbf{u}_0$, which yield negative scalar products. If similar globule conformations are adopted most over time, average scalar products can become negative.

Figure 9 shows, in contrast, correlation values calculated from a modified simulation. All 6 performed modified simulations, at which no enduring globule conformations occur, did not yield significantly negative values. Exponential fits were much more reasonable. Thus, values from equation 1 and 2 agree well with

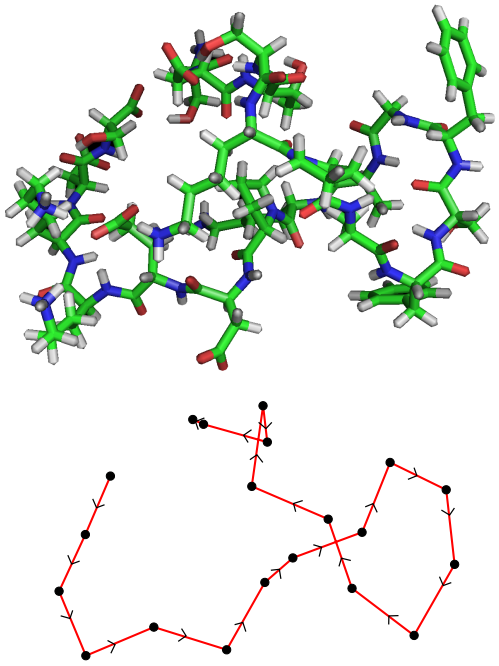

Figure 8 – Snapshot of a dense globule conformation (top) and its C_α atoms (black points) connected by the vectors $\mathbf{u}_i$ (bottom).

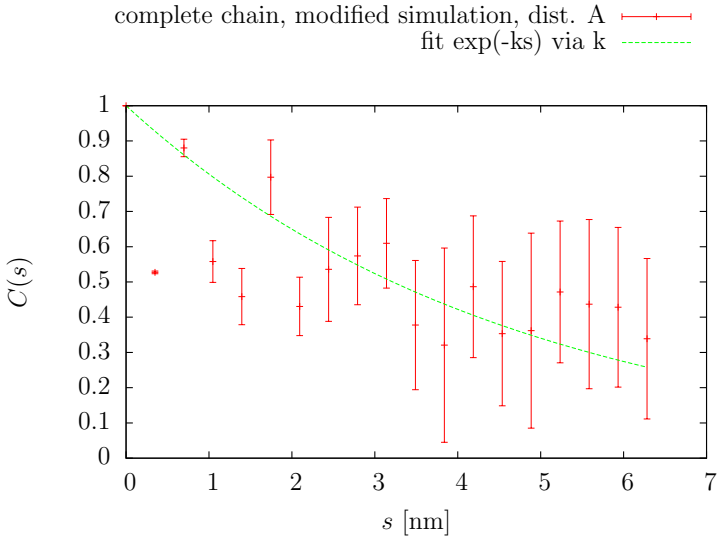

Figure 9 – Correlation values from modified simulation, velocity distribution A

each other at modified simulations.

This agreement was a result both at complete chain and backbone chain. At backbone chain simulations the agreement holds for 5 of 6 values. There is only one conspicuously irregular value from velocity distribution B, which causes a deviation between the two weighted averages.

We can conclude that the electrostatics and long-range interactions make the peptide collapse to enduring globule conformations. These conformations obviously lead to correlation values, which do not match an exponential decay. This mismatch means a deviation from the WLC model.

4.3 Side Chains do not Cause a Mismatch to the WLC Model, but Increase Persistence Length

We have seen, that the peptide at modified simulations can be well described as a WLC, because of the exponential decay in correlation values. This result is independent of existent side chains, although side chains are not considered by the WLC model.

For the complete chain we extracted a persistence length around roughly 5 nm, whereas, without side chains, five of six values are distributed around roughly 3,5 nm. Therefore we can conclude, that side chains — even in absence of electrostatics — have an effect on the persistence length.

This effect is obviously increasing, which is maybe due to a kind of mechanical friction of side chains with the solvent. The side chains obtain a certain volume within the solvent bulk and therefore confine the chain freedom.

Besides, we found, that the collapsing behavior does not require the presence of side chains. The plot in figure 10 shows the end-to-end distance of the backbone chain. Including electrostatics and long-range interactions, the backbone chain col-

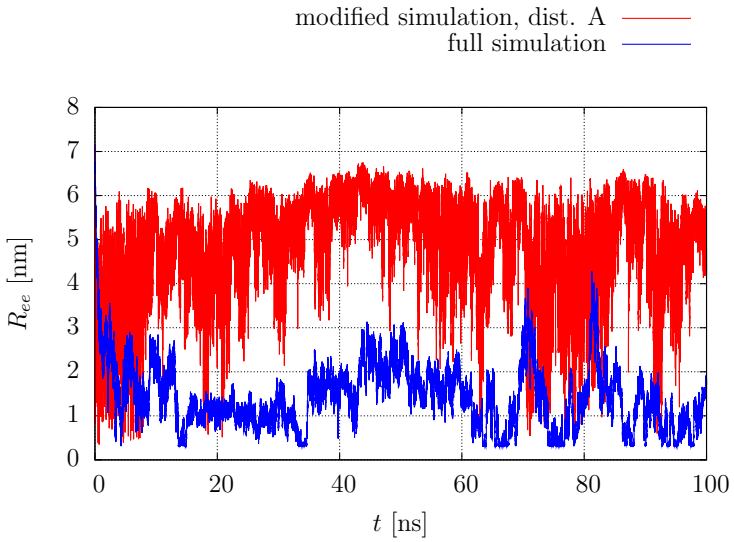

Figure 10 – End-to-end distance plot from backbone chain

lapses similar to the complete chain. Without electrostatics and long-range interactions also a behavior was found similar to the complete chain.

4.4 Improvements

First, the estimation of equilibration time is improvable. To demonstrate this, figure 11 shows the RMSD plots for the complete chain. For full simulations, the required equilibration time was estimated from this plot to be 50 ns, because up from here relatively low fluctuations occured. This was found similarly at R_{ee} plots. Data was therefore evaluated up from 50 ns.

This method is obviously a very rough estimation, which is done from only three different velocity distributions. It is for no reason clear, if three distributions can be a sufficient sample.

Especially at the modified simulations the estimation of equilibration time is disputable, because large RMSD fluctuations occur during the whole simulation time (figure 11, right). Thus, we could not even apply the criterion of low RMSD (or R_{ee})

	chain	simulation	velocity distribution	evaluated data [ns]	p. l. from correlation [nm]	p. l. from $\langle \mathbf{R}_{ee}^2 \rangle$ [nm]
	complete	full	A	50	0.80 ± 0.39	0.23 ± 0.10
			B	50	0.81 ± 0.52	0.03 ± 0.02
			C	50	0.57 ± 0.39	0.17 ± 0.06
			D	210	0.52 ± 0.23	0.09 ± 0.13
weighted mean				360	0.61 ± 0.17	0.05 ± 0.02
	complete	modified	A	90	4.64 ± 0.57	5.87 ± 4.95
			B	90	4.69 ± 0.61	5.18 ± 3.53
			C	90	7.34 ± 1.02	4.78 ± 2.95
weighted mean				270	5.05 ± 0.39	5.06 ± 2.06
	backbone	full	A	90	0.50 ± 0.10	0.17 ± 0.17
	backbone	modified	A	90	3.48 ± 0.29	3.45 ± 2.67
			B	90	0.60 ± 0.10	3.19 ± 2.23
			C	90	3.39 ± 0.37	4.34 ± 3.51
weighted mean				270	1.06 ± 0.10	3.49 ± 1.54

Table 3 – Overview of the results for persistence length (p. l.)

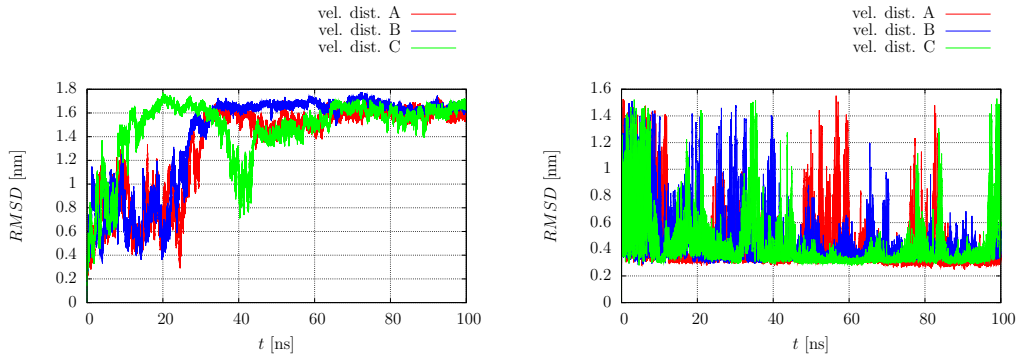

Figure 11 – RMSD plot from full simulations (left) and from modified simulations (right)

fluctuation. The fluctuation of first 10 ns is conspicuously high-frequent for all three velocity distributions and therefore this data was excluded.

More quantitative criteria and larger samplings would therefore be an improvement for the estimation of equilibration time.

Second, longer simulations would improve our simulations. This can be demonstrated, analyzing the 260 ns R_{ee} plot in figure 7 on page 17. This plot shows, that even after 200 ns relatively high fluctuations can occur.

5 Conclusion

In this thesis the question was, whether the persistence length of an intrinsically disordered peptide can be extracted from equilibrium simulations using the WLC model. We expected deviations from the WLC model and therefore asked, which effects cause these deviations.

The persistence length of the investigated peptide is not extractable from our equilibrium simulations. Electrostatics and long-range interactions make the peptide perform collapses to dense globule conformations. These conformations are adopted

most over time. This effect is not included within the WLC model and, therefore, causes an effective deviation from this model.

As we could not extract the equilibrium persistence length of our peptide, we can not compare it to recent stretching measurements. These stretching measurements have found a persistence length of 0.39 ± 0.14 nm [5].

However, we can compare our found persistence length, which we found from simulations without calculating long-range interactions and electrostatics to a value from Gräter et al. [6]. Surprisingly, the results do not match. We have found an equilibrium persistence length of 5.05 ± 0.38 nm, whereas stretching simulations yield a persistence length of 1.79 ± 0.77 nm.

This large deviation can be due to the fact, that Gräter et al. applied fits to a low-force regime. The force stretching formula by Marko and Siggia is known to be inaccurate at forces far from high- and low-force limits [14]. The fit by Gräter et al. had a poor match to their force-stretching values, which explains the large relative statistical error of their value.

We tried to figure out more in detail, which kind of electrostatics and long-range interactions lead to deviations from the WLC model. Interacting side chains (salt bridges) and backbone hydrophobicity can come into consideration.

Because of long-range interactions and electrostatics, we could also not apply the WLC model to simulations of a 20-Glycine peptide, which has no side chains. We can therefore conclude, that long-range interactions and electrostatics only from backbone are sufficient to cause deviations to the WLC model. For the 20-Glycine chain, long-range interactions and electrostatics can be summarized as hydrophobicity, as Glycine is a hydrophobic amino acid.

Our investigated intrinsically disordered peptide consists of hydrophobic residues distributed over two-thirds of the chain.

Its collapsing behavior confirms the observation, that the hydrophobicity of the backbone causes deviations from WLC model. This result is supported by recent studies, which also found intrinsically disordered peptides collapsing especially because of hydrophobicity [15].

For experimentalists, our results mean, that the persistence length extracted from equilibrium measurements using the WLC model not necessarily represents the persistence length in its model independent meaning.

Recent studies gave strong hints, that this problem can also occur extracting protein persistence length from stretching measurements. The force-extension formula by Marko and Siggia is derived by the WLC model and does therefore not consider electrostatics and long-range interactions. During stretching measurements, however, these effectively influence the protein.

Contributions of electrostatics and long-range interactions on persistence length of ubiquitin were investigated during stretching simulations. These simulations revealed, that electrostatics and long-range interactions cause a decrease of the effectively measured value [10]. Our equilibrium simulations confirm the tendency, that effectively obtained values of persistence length are smaller than those ignoring electrostatics and long-range interactions.

Thus, both at stretching and at equilibrium measurements the meaning of the obtained 'persistence length' is not clear. For future works, it is therefore a task to develop a protein model, which includes hydrophobicity and gives an applicable definition of persistence length.

References

[1] Marko, J., E. Siggia. 1995. Stretching DNA. *Macromolecules*. 28:8759–8770.

[2] Doi, M., S. F. Edwards. The Theory of Polymer Dynamics; Oxford University Press: New York, 1986.

[3] Rief, M., M. Gautel, F. Oesterhelt, J. Fernandez, H. Gaub. 1997. Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science*. 276:1109–1112.

[4] Schlierf, M., F. Berkemeier, M. Rief. 2007. Direct Observation of Active Protein Folding Using Lock-in Force Spectroscopy. *Biophys. J.* 93:3989–3998.

[5] Lim, R., N. Huang, ..., U. Aebi. 2006. Flexible Phenylalanine-Glycine Nucleoporins as Entropic Barriers to Nucleocytoplasmic Transport. *Proc. Natl. Acad. Sci. USA*. 103:9512–9517.

[6] Cheng, S., M. Cetinkaya, F. Gräter. 2010. How Sequence Determines Elasticity of Disordered Proteins. *Biophys. J.* 99:3863–3869.

[7] Denning, D., S. Patel, ..., M. Rexach. 2003. Disorder in the Nuclear Pore Complex: The FG Repeat Regions of Nucleoporins are Natively Unfolded. *Proc. Natl. Acad. Sci. USA*. 100:2450–2455.

[8] Flory, P. Statistical Mechanics of Chain Molecules; Interscience: New York, 1969.

[9] Ullner, M., B. Jönsson, ..., B. Söderberg. 1997. The Electrostatic Persistence Length Calculated from Monte Carlo, Variational and Perturbation Methods. *J. Chem. Phys.* 107:1279–1287.

[10] Gräter, F., P. Heider, R. Zangi, B. Berne. 2008. Dissecting Entropic Coiling and Poor Solvent Effects in Protein Collapse. *J. Am. Chem. Soc.* 130:11578–11579.

[11] Hess, B., D. van der Spoel, ..., E. Lindahl. 2010. Gromacs User Manual, version 4.5.4.

[12] Jorgensen, W., J. Tirado-Rives. 1988. The OPLS Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* 110:1657-1666.

[13] Jorgensen, W., J. Chandrasekhar, J. Madura. 1983. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79:926-935.

[14] Bouchiat, C., M. Wang, ..., V. Croquette. 1999. Estimating Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements. *Biophys. J.* 76:409–413.

[15] Vitalis, A., X. Wang, R. Pappu. 2007. Quantitative Characterization of Intrinsic Disorder in Polyglutamine: Insights from Analysis Based on Polymer Theories. *Biophys. J.* 93:1923–1937.

6 Acknowledgments

I would like to thank my supervisor Helmut Grubmüller for his very nice and patient assistance. I also thank my second supervisor Jörg Enderlein for giving me the opportunity to make my presentation. Furthermore, I would like to thank Rudolf Schemm, Carsten Kutzner and the whole staff of the department of theoretical and computational biophysics. Special thanks go to Marc André Heller for inspiring discussions and for reading this thesis. At last, I would like to thank my mother, who always supported me, wherever she could.

Erklärung

Erklärung nach §13 Abs. 8 der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder singemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe. Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 30. Juli 2012

(Cornelius Müthel)