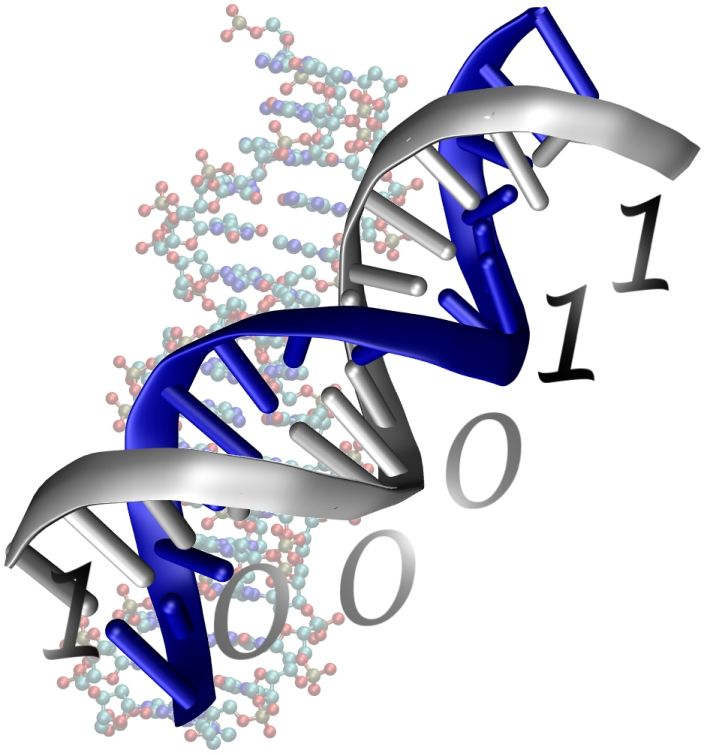
Technische Universität München
Physik Department
Theoretische Physik T37

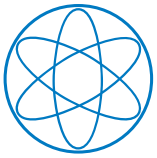
Dissertation

Folded Bio-Polymers

Structure, Pulling and Diffusion

Thomas Rudolf Einert





Technische Universität München

Physik Department

Theoretische Physik T37

Folded Bio-Polymers

Structure, Pulling and Diffusion

Thomas Rudolf Einert

Dipl.-Phys. (Univ.)

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Andreas Bausch
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Roland Netz
	2. Univ.-Prof. Dr. Holger Stark, Technische Universität Berlin

Die Dissertation wurde am 21. Dezember 2010 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 24. Januar 2011 angenommen.

Science is like sex: sometimes something useful comes out,
but that is not the reason we are doing it.

Richard P. Feynman (1918 – 1988)

ABSTRACT

In this work denaturation transitions of nucleic acids and bio-polymers in and out of equilibrium are studied theoretically. First, the over-stretching transition of DNA is modeled successfully by a three-state model, which allows to disentangle effects of the stacking energy of base pairs and the loop free energy. Second, the thermodynamics of RNA folding is studied: A marked dependence on the parameterization of loops, which are a common theme in folded RNA structures, and on salt concentration is found. Including these effects yields perfect agreement with experimental data. Third, coarse grained Brownian dynamics simulations of homopolymers reveal the influence of cohesive strength and chain length on internal friction. The existence of two distinct globular phases is shown.

Diese Arbeit behandelt Denaturierungsübergänge von Nukleinsäuren und Biopolymeren im Gleichgewicht und Nichtgleichgewicht. Zunächst wird der Überdehnungsübergang bei DNA durch ein Dreizustandsmodell erfolgreich beschrieben, welches es ermöglicht die Einflüsse der Stapelwechselwirkung von Basenpaaren und der Schleifenentropie zu trennen. Im Anschluss wird der Einfluss der Parametrisierung von Schleifen, ein häufig auftretendes Motiv in RNA-Strukturen, und der Salzkonzentration auf die Thermodynamik von RNA untersucht. Die Berücksichtigung dieser Effekte liefert hervorragende Übereinstimmung mit experimentellen Daten. Schließlich wird mit Hilfe von Brown'sche-Dynamik-Simulationen gezeigt, dass die interne Reibung bei Homopolymeren von der Kettenlänge und den Kohäsionskräften abhängt. Des Weiteren können im Polymerglobul zwei Phasen unterschieden werden.

CONTENTS

Chapter 1

Introduction and outline

The double helix is indeed a remarkable molecule. Modern man is perhaps 50 000 years old, civilization has existed for scarcely 10 000 years and the United States for only just over 200 years; but DNA and RNA have been around for at least several billion years. All that time the double helix has been there, and active, and yet we are the first creatures on Earth to become aware of its existence [9].

Francis Crick (1916 – 2004)

In 1869 Johann Friedrich Miescher discovered a substance in pus cells, which he termed nuclein and about which he wrote *"Wir haben vielmehr hier Körper sui generis, mit keiner jetzt bekannten Gruppe vergleichbar."* [10]. He analyzed the nuclein's elementary composition and already stated that it has to be a *"at least four basic acid"* [11]. It turned out that Miescher was the first to get a glimpse on the molecules that determine our lives: deoxyribonucleic acid – DNA – and ribonucleic acid – RNA.

Yet there was a gap of 75 years between the discovery of DNA and the realization of its role for inheritance. Avery, MacLeod, and McCarty [12] showed that DNA is responsible for the transformation of bacteria and hence carries genetic information. Later, Hershey and Chase [13] corroborated this thesis by demonstrating that DNA is the genetic material of viruses.

"We wish to suggest a structure for the salt of deoxyribose nucleic acid." – with these words James D. Watson and Francis H. C. Crick announced the discovery of the famous double helix in their seminal paper in 1953 [14]. The revelation of the structure of DNA finally solved the puzzle of how our hereditary material is encoded and replicated. Since then, scientists from all disciplines have striven to enhance our un-

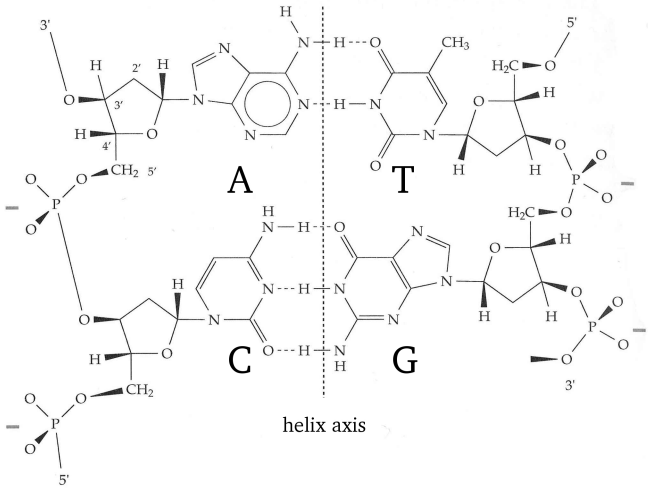


Figure 1.1.: The chemical structure of the two Watson-Crick base pairs is shown. The sugar phosphate backbone on either side connects the bases on each strand. The bases are recognized as the nitrogen containing ring compounds near the helix axis. Complementary bases establish hydrogen bonds – three hydrogen bonds in a CG pair, two for AT – which leads to the formation of the double helix.

derstanding of nucleic acids and to employ this knowledge in research and biotechnological applications.

1.1. Nucleic acids – versatile bio-polymers

Nucleic acids are linear chain-like molecules that are made up of four different types of monomers – the nucleotides. A nucleotide consists of three different parts: a phosphoric acid, a five-carbon sugar, and a nitrogen containing ring compound generally referred to as a base [15]. For DNA, there are four different bases, namely cytosine (C), guanine (G), adenine (A), and thymine (T), which are linked to the sugar, see fig. 1.1. DNA and RNA differ in only two respects: while the sugar in DNA is deoxyribose and thymine is the fourth base, RNA contains ribose and features uracil (U) instead of thymine. In the present work, these differences between DNA and RNA are of minor importance and most of the results obtained are valid for both. With the nucleotides as the basic building blocks, a nucleic acid can be polymerized by linking the sugar of one nucleotide to the phosphate of another. Nucleic acids occur in a huge variety of lengths starting from small oligonucleotides consisting of less than a dozen bases up to huge molecules containing several million bases as, for instance, the DNA in a human chromosome [16]. By arranging the four bases in different ways specific sequences can be designed. The purpose of the sequence is twofold. First, it serves as the storage of genetic information and, second, it determines the three-dimensional structure the

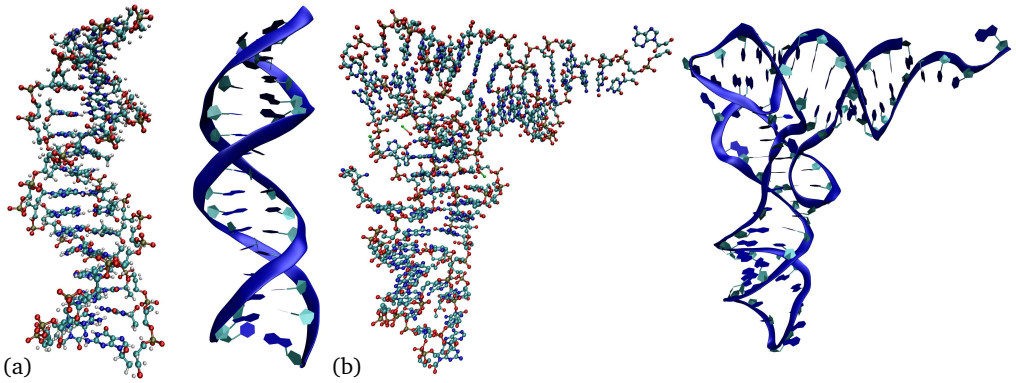


Figure 1.2.: The all atom structure and a schematic image of two nucleic acids are shown. In the schematic schemes, the sugar phosphate backbone is recognized as the thick blue ribbon, while bases are depicted as plates. Images were generated with VMD [20]. (a) NMR structure of a 16 base pair DNA oligomer [21, PDB: 1SS7]: the two DNA strands intertwine and form the famous double helix. Complementary bases on the opposite strands form Watson-Crick base pairs via hydrogen bonding. The helical structure is further stabilized by the stacking interaction between two neighboring base pairs lying on top of each other. (b) Crystal structure of yeast tRNA-phe obtained by X-ray diffraction [22, PDB: 1I9V]: this single stranded RNA consists of 76 bases and assumes an L-shaped structure typical for transfer RNA. The molecule interacts with itself featuring helical sections and loops with unbound bases.

molecule will fold into and hence its functionality. While DNA mainly acts as a genetic information storage, RNA accomplishes numerous tasks and may act as an information carrier, catalyst, or regulator for protein expression [15, 17]. For example, messenger RNA (mRNA) serves as a blueprint into which the genetic information in the DNA is transcribed. In the ribosome, mRNA is translated into proteins, where the amino acids – the individual building blocks of proteins – are delivered by a transfer RNA (tRNA). The ribosome itself contains yet another type of RNA, ribosomal RNA (rRNA), which catalyses protein synthesis. Further, RNA regulates protein expression in the form of riboswitches, which are able to sense the presence of certain metabolites [18], or as small interfering RNA (siRNA) molecules that bind to mRNA and initiate its degradation [19]. All these abilities of nucleic acids are brought about by the structure of the molecule, which is encoded in the very sequence written in a four letter alphabet.

The most important interaction, which guides the molecule towards its structure, is the specific pairing inducing the famous Watson-Crick base pairs: cytosine and guanine on the one hand as well as adenine and thymine/uracil on the other hand. The binding between the bases is brought about by hydrogen bonds – three for CG and two for AT/AU, see fig. 1.1. It is base pairing that causes two complementary nucleic acid strands to hybridize and to form the well known double helix, which is observed in siRNA or double stranded DNA (dsDNA), see fig. 1.2a. However, a single strand may

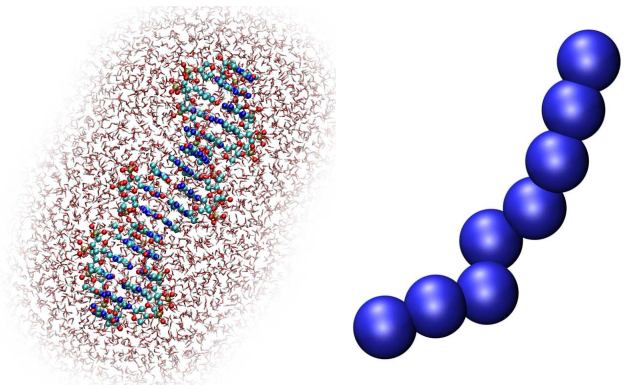


Figure 1.3.: Illustration of coarse graining DNA. On the left, a 16 base pair long DNA molecule solvated in water is shown. In a molecular dynamics simulation all atoms ($\sim 10\,000$) would be considered. After coarse graining this system, the water molecules are no longer modeled explicitly and two consecutive base pairs (~ 100 atoms) are represented by a single bead as depicted on the right. A Brownian dynamics simulation of this bead spring model would require to consider only eight monomers; the effect of the solvent would be included by introducing a fluctuating random force acting on the beads. Therefore, coarse graining allows to study larger length scales and longer time scales.

also interact with itself and form intricate structures, which are, for instance, found in riboswitches, rRNA, or tRNA, see fig. 1.2b.

The structure of nucleic acids may be decomposed into three hierarchy levels. The primary structure is the mere sequence of bases along the chain. Base pairing leads to the secondary structure, which – on an abstract level – can be viewed as a list of all base pairs. The tertiary structure is the complete three-dimensional configuration and contains the positions of all atoms.

1.2. A physicists view on bio-polymers

Physicists try to understand complex systems by focusing on their essentials and by comprehending them from first principles. Experimentally this can be achieved by investigating idealized systems in a very controlled setup. Theory complements experiments and aims to explain reality by models, which comprise equations and simulations that can be checked against experimental data.

To describe bio-polymer systems theoretically an abundance of degrees of freedom has to be dealt with: for instance, to study the movement of tRNA through the ribosome several million atoms were simulated [23]. But not always so much detail is necessary to understand nature, which leads to main idea of the powerful concept of coarse graining: the deliberate procedure of neglecting irrelevant degrees of freedom. A possible first step of coarse graining nucleic acids is illustrated in fig. 1.3: Instead of describing the positions of all atoms of the DNA and the surrounding water, one

represents several base pairs by a single bead and desists from modeling water explicitly. Various levels of coarse graining exist and successfully describe experiments. For example, in molecular dynamics simulations typically all atoms of the bio-polymer and the solvent are considered (left hand side of fig. 1.3), but quantum effects are neglected [24]. Further simplification leads to bead-spring models, where solvent molecules are omitted and successive monomers are represented by a single bead (right hand side of fig. 1.3), which allows to study larger systems on a longer time scale [25]. Apart from simulations, bio-polymers can be investigated analytically. To describe the elastic properties, DNA can be modeled as a semiflexible rod [26]. Pushing coarse graining to its extremes, phenomena as complex as the melting of double stranded DNA can be described by discrete models; within the Poland-Scheraga theory, a single variable, which is either 1 or 0, indicates whether or not two bases are paired without considering the three-dimensional conformation of the molecule [27, 28]. In this work, we employ various levels of coarse graining ranging from simulations of bead-spring models to discrete models that can be treated analytically.

Bio-polymers belong to the class of soft matter systems since they are easily deformed by thermal fluctuations as typical energies are of the order of the thermal energy. These materials are particularly interesting as they constitute model systems for many physical fields and phenomena – e. g. statistical mechanics, (non-) equilibrium thermodynamics, stochastic dynamics, critical phenomena, phase transitions, or glasses – where models and concepts may be tested against experiments. Quantum effects are generally unimportant and classical theories may be employed. Soft matter and bio-physics is a rich and active field of research being relevant for our understanding of life and for many biotechnological applications. This work tries to contribute to this scientific community and to enhance our knowledge in the area of bio-polymers.

1.3. Outline of this work

The aim of this thesis is to give an extensive picture of the statistical mechanics and the denaturation transitions associated with the secondary structure of nucleic acids. In chapters 2, 3, and 4 we develop a theory for equilibrium properties of nucleic acids. Only secondary structure interactions are considered, whereas the tertiary structure is neglected. This approximation relies on the idea of hierarchical folding proposed by Tinoco and Bustamante [29], who claim that given a sequence, the secondary structure forms independently of the tertiary structure. In chapter 5 we shift our focus to dynamical aspects and assume a more general point of view by considering homopolymers without specific base pairing interactions.

Double stranded DNA is known to exhibit a highly cooperative over-stretching transition when subject to a tensile force [30]. This gave rise to a still ongoing and controversial debate whether this over-stretching transition produces a distinct DNA state or not. In chapter 2 we introduce a three-state model for a single DNA chain under tension that distinguishes between native B-DNA, over-stretched S-DNA, and dena-

tured segments [3]. We show that the existence of S-DNA sensitively depends on the model parameters and suggest that the stability of the S-state is very susceptible to experimental conditions. Further, by fitting to experimental data our approach allows to extract the loop exponent, which characterizes the entropy of denatured loops, and the stacking interaction independent from each other. Finally, we establish a global phase diagram in the force-temperature plane.

Upon approaching the denaturation transition of double stranded nucleic acids, molten loops proliferate and inter-strand base pairing abates. This leads to an interesting and biotechnologically extremely relevant question: Do secondary structure elements form in the single strands inside loops or not? To answer this question we characterize the thermodynamic behavior of single stranded nucleic acids, which occur in molten DNA loops and particularly in the form of RNA, such as tRNA, rRNA, or mRNA. In chapter 3 we analytically calculate the partition function, determine the phase diagram, and find non-universal critical exponents, which depend on the loop exponent [1]. For the case of denaturation of double stranded nucleic acids, we demonstrate that inter-strand base pairing is always in competition with intra-strand base pairing if the sequence allows intra-strand base pairing [5]. These findings have important implications on theory and software development, which aim to predict the stability of nucleic acid compounds.

In chapters 2 and 3 sequence effects are neglected. However, the structure of nucleic acids and hence their functioning is mainly controlled by the specific Watson-Crick base pairing between C and G and between A and T/U. In addition, the stability of nucleic acids is strongly influenced by the salt concentration of the solution. While for DNA numerous corrections of the base pairing free energies due to varying salt concentrations exist, analogous results for the salt dependence of RNA energy parameters are very sparse. In chapter 4 we derive salt dependent free energies for base pairing and loops [4]. We include a logarithmic loop entropy characterized by the loop exponent and demonstrate that also the non-critical behavior sensitively depends on the loop exponent [1]. Subsequently we consider thermal and force induced denaturation of RNA and observe excellent agreement between our results and experimental data. We show that for a proper description of RNA melting and stretching, both salt and loop entropy effects are needed. We derive the phase diagram in the three-dimensional phase space spanned by temperature, external stretching force, and salt concentration.

In chapter 5 we assume a more general point of view by omitting the specific base pairing interactions and study dynamical and non-equilibrium properties of homopolymeric globules by means of Brownian dynamics simulations of bead-spring polymers [6]. We find a marked dependence of the internal dynamics on the chain length as well as on the cohesive force and observe a transition from liquid-like to solid-like globules [6, 7]. While monomers are quite mobile inside the globule in the liquid phase, the dynamics are very slow in the solid phase. The results in this chapter have implications for many systems as conformational dynamics of polymers play a crucial role in biological systems and biotechnological applications: transcription, translocation, and translation of RNA impose huge conformational changes on the

molecule and require to consider the dynamics of folding and unfolding. DNA packing in the chromosome or accessing genes for transcription involves rearrangement and reorganization of the molecule. Further, proteins experience many intermediate structures along their way towards the folded state and the adsorption behavior of polymers to surfaces depends on their configuration [8].

In chapter 6 we summarize the main results of this work and give an outlook for future studies.

CHAPTER 2

THREE-STATE MODEL FOR THE OVER-STRETCHING TRANSITION OF DNA

2.1. Introduction

2.1.1. Motivation

DNA continuously stays in focus of polymer scientists due to its unique mechanical and structural properties. In particular the possibility to trigger phase transformations in this one-dimensional system has intrigued theorists from different areas [27]. In fact, the thermal denaturation or melting transition of DNA was shown to correspond to a true phase transition, brought about by a logarithmic contribution to the configurational entropy of molten loops or bubbles, $S \sim -\hat{c}\ln n$, as a function of the loop size n [28]. The value of the exponent $\hat{c}$ is crucial since it determines the resulting transition characteristics. For $\hat{c} = 3/2$, the value for a phantom chain without self-avoidance, the transition is continuous, while self-avoidance increases $\hat{c}$ slightly beyond the threshold $\hat{c} = 2$ above which the transition becomes discontinuous [28, 31]. A distinct mechanism for transforming DNA involves the application of an extensional force. For forces around $F \approx 65$ pN DNA displays a highly cooperative transition and its contour length increases by a factor of roughly 1.7 to 2.1 over a narrow force range [30, 32, 33]. These experiments sparked a still ongoing debate on whether this over-stretching transition produces a distinct DNA state, named S-DNA, or merely the denatured state under external tension. According to the first view S-DNA is a highly stretched state with paired bases but disrupted base stacking [34–40]. In the other view the over-stretched state consists of two non-interacting strands [41–44]. Evidence for the existence of a distinct S-state comes from theoretical models [34, 36], molecular dynamics simulations [37–40] and from AFM experiments of Rief et al. [45–47] where in addition to the over-stretching transition a second weak transition at forces between 150 pN and 300 pN is discerned, which has been interpreted as a force induced melting of the S-state. The critical force of both tran-

sitions depends on the actual sequence [46] and the salt concentration [48], but the interpretation of the second transition is complicated by the occurrence of pronounced hysteresis effects that depend on various parameters such as pulling velocity, salt concentration, or presence of co-solutes such as cisplatin [46, 47]. On the other hand, support for the view according to which S-DNA is not a distinct state comes from theoretical models [41, 42], simulations [49] and recent experiments by Shokri et al. [43] and van Mameren et al. [44].

Apart from simulations [37–40], existing theoretical works that grapple with experimental force traces or DNA melting fall into three categories with increasing computational complexity, for reviews see refs. [50–52]. In the first group are Ising-like models for DNA under tension which give excellent fitting of the over-stretching transition but by construction cannot yield the denaturing transition [42, 53–55]. The work of Marko [56] is similar but employs a continuous axial strain variable. In the second group are models that include a logarithmic entropy contribution of molten loops in the spirit of the classical model by Poland and Scheraga [28] [31, 36, 57–62]. This gives rise to effectively long-ranged interactions between base pairs (bp) and thus to a true phase transition. The third group consists of models which explicitly consider two strands [63–68]. Those models thereby account for the configurational entropy of loops – at the cost of considerable calculational efforts – and correspond to loop exponents $\hat{c} = d/2$ in the absence of self-avoidance effects, where the dimensionality of the model is $d = 3$ for ref. [64–66] and 1 for ref. [67, 68]. All these above mentioned works consider only two different base states (paired versus unpaired) and thus do not allow to distinguish between B-DNA, S-DNA, and denatured bases. Recently, three-state models were introduced that yield very good fits of experimental force traces at ambient temperatures. However, in previous analytic treatments of such three-state models [34, 69], the loop entropy was neglected and therefore the temperature-induced denaturation in the absence of force cannot be properly obtained, while the loop entropy was included in a simulation study where most attention was given to dynamic effects [36].

2.1.2. Outline

In this chapter we combine the Poland-Scheraga formalism with a three-state transfer matrix approach which enables us to include three distinct local base pairing states and at the same time to correctly account for long-ranged interactions due to the configurational entropy of molten DNA bubbles. Our approach thus allows for a consistent description of thermal denaturation and the force induced BS-transition within one framework and yields the global phase diagram in the force-temperature plane. We derive a closed form expression for the partition function of three-state DNA under tension. This allows to systematically investigate the full parameter range characterizing the three states and the DNA response to temperature and external force. In our model we allow for the existence of S-DNA but stress that the actual occurrence of S-DNA is governed by the model parameters. By assuming such a general point of view, our

work is able to shed new light on the question of the existence of S-DNA. The extensible worm-like chain model is employed for the stretching response of each state. The loop exponent is found to have quite drastic effects on the force extension curve. For realistic parameters for the stacking energy, the experimental force extension curves are fitted best for small loop exponents $0 \leq \hat{c} \leq 1$, hinting that the DNA in the experiments contained nicks. Loop exponents $\hat{c} > 1$, which give rise to a genuine phase transition, are not compatible with experimental force distance curves. Under external force, the effects of stacking energy and loop exponent are largely decoupled, since the stacking energy only determines the cooperativity of the BS-transition while the loop exponent influences the second SM-transition found at higher forces. This allows to disentangle these two parameters, in contrast to the denaturation transition at zero force where the effects of these two parameters are essentially convoluted. The precise value of $\hat{c}$ is important also from a practical point of view, as it impacts the kinetics of DNA melting [36, 70], which is omnipresent in biological and bio-technological processes.

2.2. Three-state model

Double-stranded DNA is modeled as a one-dimensional chain with bases or segments that can be in three different states, namely paired and in the native B-state, in the paired stretched S-state, or in the molten M-state. The free energy of a region of n segments in the same state reads

$$\mathcal{G}_i(n,F) = n\, g_i(F) - \delta_{i,\mathrm{M}}\, \mathrm{k_B}T \ln n^{-\hat{c}}\,, \tag{2.1}$$

with $i = \mathrm{B}, \mathrm{S}, \mathrm{M}$. The force F dependent contribution

$$g_i(F) = g_i^0 + g_i^{\mathrm{stretch}}(F) + g_i^{\mathrm{WLC}}(F) \tag{2.2}$$

is split into three parts. g_i^0 is a constant that accounts for the base pairing as well as the difference of reference states of the worm-like stretching energy, cf. appendix A.1 eq. (A.1). The stacking energy of neighboring bases in the same state is absorbed into g_i^0, too, so that the stacking energy will appear explicitly only as an interfacial energy V_{ij} between two regions which are in a different state. The second term $g_i^{\mathrm{stretch}}(F) = -F^2 l_i/(2\kappa_i)$ takes into account stretching along the contour with l_i and κ_i the segmental contour length and the elastic stretch modulus. Finally, $g_i^{\mathrm{WLC}}(F)$ is the free energy of a worm-like chain (WLC) in the Gibbs ensemble (constant force F), based on the heuristic relation between the force F and the projected extension x [26] $F_i^{\mathrm{WLC}}(x)\xi_i/\mathrm{k_B}T = \big(1 - x/(nl_i)\big)^{-2}/4 + x/(nl_i) - 1/4$ where ξ_i is the persistence length and n the number of segments. The Gibbs free energy $ng_i^{\mathrm{WLC}}(F)$ of a stretch of n segments is extensive in n and follows via integration, see appendix A.1. We note that this is only valid if the persistence length is smaller than the contour length of a region, $\xi_i < nl_i$, which is a plausible assumption because of the high domain wall

energies. Likewise, the decoupling of the free energy into contour stretching elasticity and worm-like chain elasticity is only approximate [71, 72] but quite accurate for our parameter values [73]: For small force, WLC bending fluctuations dominate and the contour extensibility is negligible, while contour stretching sets in only when the WLC is almost completely straightened out. The last term in eq. (2.1) is the logarithmic configurational entropy of a molten loop ($i = \mathrm{M}$), characterized by an exponent $\hat{c}$ [1, 31, 58], see appendix A.3. The exponent is $\hat{c} = 3/2$ for an ideal polymer [74] and 2.1 for a self avoiding loop with two attached helices [31, 75]. If the DNA loop contains a nick the exponent is reduced to $\hat{c} = 0$ for an ideal polymer and 0.092 for a self avoiding polymer [31]. We consider the simple case $\hat{c} = 0$, where transfer matrix methods can be used to yield results in the canonical ensemble with a fixed number of segments N [34], as well as the case of finite $\hat{c}$ where we introduce a modified Poland-Scheraga method to obtain results in the grand canonical ensemble. Intra-strand base pairing in molten loops, i. e. the interaction of a single strand with itself, is not considered here. We will come back to this point in chapter 3.

2.3. Partition functions

2.3.1. Modified Poland-Scheraga approach for $\hat{\mathbf{c}} \neq \mathbf{0}$

The molecule is viewed as an alternating sequence of different regions each characterized by grand canonical partition functions. Various techniques for going back to the canonical ensemble are discussed below. The canonical partition function of a stretch of n segments all in state $i = \mathrm{B}$, S, or M is

$$\mathcal{Q}_i(n) = \exp(-\beta \mathcal{G}_i(n)) \,, \tag{2.3}$$

where $\beta = (\mathrm{k_B} T)^{-1}$ is the inverse thermal energy. The grand canonical partition functions are defined as $\mathcal{Z}_i = \sum_{n=1}^{\infty} z^n \mathcal{Q}_i(n)$ with $z = \exp(\beta\mu)$ the fugacity and μ the chemical potential. The grand canonical partition function of the whole DNA chain, which contains an arbitrary number of consecutive B, S, and M stretches reads

$$\mathcal{Z} = \sum_{k=0}^{\infty} \boldsymbol{\nu}^{\mathrm{T}} \cdot (\mathbf{M}_{\mathrm{PS}} \mathbf{V}_{\mathrm{PS}})^k \mathbf{M}_{\mathrm{PS}} \cdot \boldsymbol{\nu} = \boldsymbol{\nu}^{\mathrm{T}} \cdot (\mathbf{1} - \mathbf{M}_{\mathrm{PS}} \mathbf{V}_{\mathrm{PS}})^{-1} \mathbf{M}_{\mathrm{PS}} \cdot \boldsymbol{\nu}, \tag{2.4}$$

see fig. 2.1, with the matrices in this Poland-Scheraga approach given by

$$\mathbf{M}_{\mathrm{PS}} = \begin{pmatrix} \mathcal{Z}_{\mathrm{B}} & 0 & 0 \\ 0 & \mathcal{Z}_{\mathrm{S}} & 0 \\ 0 & 0 & \mathcal{Z}_{\mathrm{M}} \end{pmatrix} , \quad \mathbf{V}_{\mathrm{PS}} = \begin{pmatrix} 0 & \mathrm{e}^{-\beta V_{\mathrm{BS}}} & \mathrm{e}^{-\beta V_{\mathrm{BM}}} \\ \mathrm{e}^{-\beta V_{\mathrm{SB}}} & 0 & \mathrm{e}^{-\beta V_{\mathrm{SM}}} \\ \mathrm{e}^{-\beta V_{\mathrm{MB}}} & \mathrm{e}^{-\beta V_{\mathrm{MS}}} & 0 \end{pmatrix} , \quad \boldsymbol{\nu} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{2.5}$$

and where $\mathbf{1}$ is the unity matrix. The energies V_{ij} are the interfacial energies to have neighboring segments in different states and are dominated by unfavorable base pair un-stacking. The diagonal elements of $\mathbf{V}_{\mathrm{PS}}$ are zero which ensures that two neighboring regions are not of the same type and thus prevents double counting. The explicit

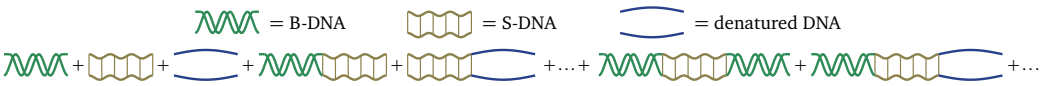


Figure 2.1.: Illustration of the calculation of the grand canonical partition function: The three different pictograms denote the grand canonical partition functions. The sum depicts the first terms of the sum in eq. (2.4).

form of $\mathcal{Z}$ is given in appendix A.2, see eq. (A.2). The partition function in eq. (2.4) is general and useful for testing arbitrary models for the three DNA states as given by the different $\mathcal{Z}_i$. This approach is also easily generalized to higher numbers of different states. Using the parameterization eq. (2.1) for vanishing loop exponent $\hat{c} = 0$ the partition functions of the different regions are given by

$$\mathcal{Z}_i = \sum_{n=1}^{\infty} z^n \mathcal{Q}_i(n) = \frac{z\mathrm{e}^{-\beta g_i}}{1 - z\mathrm{e}^{-\beta g_i}}\,, \tag{2.6}$$

for $z\mathrm{e}^{-\beta g_i} < 1$, $i = \mathrm{B}, \mathrm{S}, \mathrm{M}$. Insertion into eq. (2.4) yields

$$\mathcal{Z}_{\hat{c}=0} = \frac{a_1 z + a_2 z^2 + a_3 z^3}{a_4 + a_5 z + a_6 z^2 + a_7 z^3}\,. \tag{2.7}$$

which is a rational function of the fugacity z, whose coefficients a_i – determined by eqs. (2.4) and (2.6) – are smooth functions of the force F and the temperature T. For $\hat{c} \neq 0$ the partition function of a molten stretch is modified to

$$\mathcal{Z}_{\mathrm{M}} = \sum_{n=1}^{\infty} z^n \mathcal{Q}_{\mathrm{M}}(n) = \sum_{n=1}^{\infty} z^n \left(\mathrm{e}^{-\beta g_{\mathrm{M}}}\right)^n \frac{1}{n^{\hat{c}}} = \mathrm{Li}_{\hat{c}}(z\mathrm{e}^{-\beta g_{\mathrm{M}}})\,, \tag{2.8}$$

for $z\mathrm{e}^{-\beta g_{\mathrm{M}}} < 1$, where $\mathrm{Li}_{\hat{c}}(x) = \sum_{n=1}^{\infty} x^n n^{-\hat{c}}$ for $x < 1$ is the polylogarithm [76] and exhibits a branch point at $x = 1$, see appendix A.4. The functional form of the grand canonical partition function for $\hat{c} \neq 0$ reads

$$\mathcal{Z}_{\hat{c}\neq 0} = \frac{b_0 z + b_1 z^2 + b_2 \mathrm{Li}_{\hat{c}}(z/z_{\mathrm{b}}) + b_3 z \mathrm{Li}_{\hat{c}}(z/z_{\mathrm{b}}) + b_4 z^2 \mathrm{Li}_{\hat{c}}(z/z_{\mathrm{b}})}{b_5 + b_6 z + b_7 z^2 + b_8 z \mathrm{Li}_{\hat{c}}(z/z_{\mathrm{b}}) + b_9 z^2 \mathrm{Li}_{\hat{c}}(z/z_{\mathrm{b}})}\,, \tag{2.9}$$

where $z_{\mathrm{b}} = \mathrm{e}^{\beta g_{\mathrm{M}}}$ denotes the position of the branch point and the coefficients b_i, determined by eqs. (2.4), (2.6) and (2.8), are smooth functions of F and T.

The grand canonical ensemble where N, the total number of segments fluctuates, does not properly describe a DNA chain of fixed length. We therefore have to investigate the back-transformation into the canonical ensemble where the number of segments N is fixed. For the back-transformation there are three options:

Calculus of residues route

The grand-canonical partition function $\mathcal{Z}(z) = \sum_{N=1}^{\infty} z^N \mathcal{Q}(N)$ can be viewed as a Laurent series, the coefficients of which are the canonical partition functions $\mathcal{Q}(N)$ determined exactly by

$$\mathcal{Q}(N) = \frac{1}{2\pi \mathrm{i}} \oint_{\mathcal{C}} \frac{\mathcal{Z}(z)}{z^{N+1}} \mathrm{d}z \,. \tag{2.10}$$

The contour $\mathcal{C} = z_0 \mathrm{e}^{2\pi \mathrm{i} t}$, $0 \leq t \leq 1$, is a circle in the complex plane around the origin with all singularities of $\mathcal{Z}(z)$ lying outside. This complex contour integral can be evaluated using calculus of residues [77] which becomes technically involved for large N and thus limits the practical relevance of this route.

Legendre transformation route

The canonical Gibbs free energy

$$\mathcal{G}(N) = -\mathrm{k_B} T \ln \mathcal{Q}(N) \tag{2.11}$$

and the grand potential

$$\Phi(\mu) = -\mathrm{k_B} T \ln \mathcal{Z}(z) \,, \tag{2.12}$$

are related via a Legendre transformation

$$\mathcal{G}(N) = \Phi(\mu(N)) + N \cdot \mu(N) \,. \tag{2.13}$$

The chemical potential μ as a function of the segment number N is obtained by inverting the relation

$$N(\mu) = -\frac{\partial \Phi(\mu)}{\partial \mu} \,. \tag{2.14}$$

Let us briefly review the origin of eqs. (2.13) and (2.14) in the present context. Changing the integration variable in eq. (2.10) from z to $\mu = \mathrm{k_B} T \ln z$, the complex path integral can be transformed into

$$\mathcal{Q}(N) = \int_{\mathcal{C}'} \mathrm{e}^{-\beta \Phi(\mu) - \beta N \mu} \mathrm{d}\mu \approx \mathrm{e}^{-\beta \Phi(\mu_{\mathrm{sp}}(N)) - \beta N \mu_{\mathrm{sp}}(N)} \,, \tag{2.15}$$

with the contour $\mathcal{C}' = \mu_0 + 2\pi \mathrm{i} t/\beta$, $0 \leq t \leq 1$ and $\mu_0 = \mathrm{k_B} T \ln z_0$. The integral in eq. (2.15) has been approximated by the method of steepest descent, where the contour $\mathcal{C}'$ is deformed such that it passes through the saddle point μ_{sp} [77] determined by equation (2.14). If $\Phi(z)$ features singularities, deformation of the contour $\mathcal{C}'$ requires extra care. In the present case the presence of a pole $z_{\mathrm{p}} = \exp(\beta \mu_{\mathrm{p}})$ of $\mathcal{Z}(z)$ produces no problem as $\mu_{\mathrm{sp}} < \mu_{\mathrm{p}}$ holds, meaning that the deformed contour does not enclose the pole singularity. This is different for the branch point singularity μ_{b}, where we will encounter the case $\mu_{\mathrm{b}} < \mu_{\mathrm{sp}}$ for large $\hat{c} > 2$.

Dominating singularity route

For large systems, i. e. $N \gg 1$, one approximately has $-\ln \mathcal{Q}(N) \sim N \ln z_\mathrm{d}$, where the dominant singularity $z_\mathrm{d} = \exp(\beta\mu_\mathrm{d})$ is the singularity – in the general case a pole or a branch point – of $\mathcal{Z}(z)$ which has the smallest modulus. One thus finds

$$\mathcal{G}(N) = \mathrm{k_B} T N \ln z_\mathrm{d} \,. \tag{2.16}$$

This easily follows from eq. (2.10): In the limit of $N \gg 1$ the integral can be approximated by expanding $\mathcal{Z}(z)$ around z_d and deforming $\mathcal{C}$ to a Hankel contour, which encircles z_d, see appendix A.5 [78, 79]. For the case where $N(\mu) \propto (\mu_\mathrm{d} - \mu)^{-\alpha}$, $\alpha > 0$, this can be understood also in the context of a Legendre transform. As $N = -\partial\Phi/\partial\mu$ one has $\Phi \propto (\mu_\mathrm{d} - \mu)^{-\alpha+1}$ and therefore the first term of eq. (2.13) scales like $\Phi(\mu(N)) \propto N^{1-1/\alpha}$. Thus, the second term $N\mu(N) \propto N\mu_\mathrm{d} - N^{1-1/\alpha} \propto N\mu_\mathrm{d}$ is dominant. Since the saddle point behaves as $\mu_\mathrm{sp} = \mu(N) \to \mu_\mathrm{d}$ for $N \to \infty$, it follows that the dominating singularity expression eq. (2.16) equals the Legendre transform, eq. (2.13), in the thermodynamic limit $N \to \infty$.

2.3.2. Transfer matrix approach for $\hat{c} = 0$

For $\hat{c} = 0$ only interactions between nearest neighbors are present and straight transfer matrix techniques are applicable. We introduce N spin variables i_n, $n = 1, \ldots, N$, which can have the values $i_n = \mathrm{B}, \mathrm{S}, \mathrm{M}$. The energetics are given by the Hamiltonian

$$H(i_1, i_2, ..., i_N) = \sum_{n=1}^{N} g_{i_n} + \sum_{n=1}^{N-1} V_{i_n i_{n+1}} \tag{2.17}$$

where g_{i_n} and $V_{i_n i_{n+1}}$ are the previously introduced parameters for the segment and interfacial free energies. The canonical partition function of the molecule can be written as

$$\mathcal{Q}(N) = \sum_{i_1,\ldots,i_N} \mathrm{e}^{-\beta H(i_1,i_2,\ldots,i_N)} = \boldsymbol{\nu}^\mathrm{T} \cdot \mathbf{T}^{N-1} \mathbf{M}_\mathrm{TM} \cdot \boldsymbol{\nu} \,, \tag{2.18}$$

where we introduced the transfer matrix $\mathbf{T} = \mathbf{M}_\mathrm{TM}\mathbf{V}_\mathrm{TM}$ and

$$\mathbf{M}_\mathrm{TM} = \begin{pmatrix} \mathrm{e}^{-\beta g_\mathrm{B}} & 0 & 0 \\ 0 & \mathrm{e}^{-\beta g_\mathrm{S}} & 0 \\ 0 & 0 & \mathrm{e}^{-\beta g_\mathrm{M}} \end{pmatrix}, \mathbf{V}_\mathrm{TM} = \begin{pmatrix} 1 & \mathrm{e}^{-\beta V_\mathrm{BS}} & \mathrm{e}^{-\beta V_\mathrm{BM}} \\ \mathrm{e}^{-\beta V_\mathrm{SB}} & 1 & \mathrm{e}^{-\beta V_\mathrm{SM}} \\ \mathrm{e}^{-\beta V_\mathrm{MB}} & \mathrm{e}^{-\beta V_\mathrm{MS}} & 1 \end{pmatrix}, \boldsymbol{\nu} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} . \tag{2.19}$$

$\mathcal{Q}(N)$ is readily calculated by diagonalizing $\mathbf{T}$

$$\mathcal{Q}(N) = \boldsymbol{\nu}^\mathrm{T} \cdot \mathbf{U}\mathbf{D}^{N-1}\mathbf{U}^{-1}\mathbf{M}_\mathrm{TM} \cdot \boldsymbol{\nu} = \boldsymbol{\nu}_\mathrm{l}^\mathrm{T} \cdot \mathbf{D}^{N-1} \cdot \boldsymbol{\nu}_\mathrm{r} = \sum_{i=1}^{3} \nu_{\mathrm{l},i} \nu_{\mathrm{r},i} x_i^{N-1} \,. \tag{2.20}$$

where $\mathbf{D} = \mathbf{U}^{-1}\mathbf{T}\mathbf{U}$ is a diagonal matrix with eigenvalues x_i, the columns of $\mathbf{U}$ are the right eigenvectors of $\mathbf{T}$, $\boldsymbol{\nu}_\mathrm{l}^\mathrm{T} = \boldsymbol{\nu}^\mathrm{T} \cdot \mathbf{U}$ and $\boldsymbol{\nu}_\mathrm{r} = \mathbf{U}^{-1}\mathbf{M}_\mathrm{TM} \cdot \boldsymbol{\nu}$. By virtue of the Perron-Frobenius theorem one eigenvalue x_max is larger than the other eigenvalues and thus the free energy is in the thermodynamic limit dominated by x_max and reads

$$\mathcal{G} = -\mathrm{k_B}T \ln \mathcal{Q}(N) \approx -\mathrm{k_B}TN \ln x_\mathrm{max} \,. \tag{2.21}$$

The transfer matrix eigenvalues x_i and the poles $z_{\mathrm{p},i}$ of $\mathcal{Z}_{\hat{c}=0}$ eq. (2.7) are related via $x_i = 1/z_{\mathrm{p},i}$. Therefore, the free energies from the transfer matrix approach, eq. (2.21), and from the Poland-Scheraga approach for $\hat{c} = 0$, using the dominating singularity approximation eq. (2.16), are identical in the limit $N \to \infty$. Clearly, for $\hat{c} \neq 0$ the modified Poland-Scheraga approach yields new physics that deviates from the transfer matrix results.

Although not pursued here, the transfer matrix approach allows to calculate correlators. For example the probability $p_M(k,m)$ of a denatured region with k consecutive molten base pairs starting at base m is given by

$$p_M(k,m) = \mathcal{Q}(N)^{-1} \boldsymbol{\nu}^\mathrm{T} \cdot \mathbf{T}^{m-2}(\mathbf{P}_\mathrm{B}\mathbf{T} + \mathbf{P}_\mathrm{S}\mathbf{T})(\mathbf{P}_\mathrm{M}\mathbf{T})^k(\mathbf{P}_\mathrm{B}\mathbf{T} + \mathbf{P}_\mathrm{S}\mathbf{T})\mathbf{T}^{N-m-k-1}\mathbf{M}_\mathrm{TM} \cdot \boldsymbol{\nu} \,. \tag{2.22}$$

The $\mathbf{P}_i$-matrices, which project a segment onto a certain state, are defined as

$$\mathbf{P}_\mathrm{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \,, \quad \mathbf{P}_\mathrm{S} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \,, \text{ and } \quad \mathbf{P}_\mathrm{M} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \,. \tag{2.23}$$

2.4. Force extension curves

The central quantity is $\mathcal{G}(F,T,N)$, the Gibbs free energy of a DNA chain with N base pairs, subject to a force F and temperature T. From $\mathcal{G}(F,T,N)$, obtained via the Legendre transform, eq. (2.13), the dominating singularity, eq. (2.16), or the exact transfer matrix partition function, eq. (2.20), we can calculate observables by performing appropriate derivatives. The number of segments in state $i = \mathrm{B},\mathrm{S},\mathrm{M}$ is obtained by

$$N_i = \left.\frac{\partial \mathcal{G}}{\partial g_i}\right|_{T,N,F} \,. \tag{2.24}$$

The force extension curve is readily calculated via

$$x(F) = -\left.\frac{\partial \mathcal{G}}{\partial F}\right|_{T,N} = -\sum_{i=\mathrm{B,S,M}} \frac{\partial \mathcal{G}}{\partial g_i}\frac{\partial g_i}{\partial F} = \sum_{i=\mathrm{B,S,M}} N_i \left(x_i^\mathrm{WLC}(F) + F l_i/\kappa_i\right) \tag{2.25}$$

where $x_i^\mathrm{WLC}(F)$ is the stretching response of a worm-like chain and given explicitly in appendix A.1, eq. (A.1).

2.4.1. Vanishing loop entropy, $\hat{c} = 0$

In this section we compare the prediction for vanishing loop exponent $\hat{c} = 0$ to experimental data and obtain estimates of the various parameters. We also demonstrate the equivalence of the grand canonical and canonical ensembles even for small chain length N.

In order to reduce the number of free fitting parameters we extract as many reasonable values from literature as possible. For the helical rise, the stretch modulus, and the persistence length of B-DNA we use $l_B = 3.4$ Å, $\kappa_B = 1$ nN, and $\xi_B = 48$ nm [80]. For the M-state, which is essentially single stranded DNA (ssDNA), ab initio calculations yield $l_M = 7.1$ Å and $\kappa_M = 2 \cdot 9.4$ nN [81], where κ_M is valid for small forces $F < 400$ pN and the factor 2 accounts for the presence of two ssDNA strands. Our value for the stretch modulus is considerably larger than previous experimental fit estimates [30, 82], which might be related to the fact that experimental estimates depend crucially on the model used to account for conformational fluctuation effects; however, the actual value of κ_M is of minor importance for the stretching response as will be discussed later. The persistence length of ssDNA is given by $\xi_M \approx 3$ nm [83]. It turns out that the quality of the fit as well as the values of the other fit parameters are not very sensitive to the exact value of the persistence length ξ_S and the stretch modulus κ_S of the S-state as long as 10 nm $\lesssim \xi_S \lesssim$ 50 nm and κ_S is of the order of κ_M. Therefore we set $\xi_S = 25$ nm, which is an intermediate value between the persistence lengths of ssDNA and B-DNA, and $\kappa_S = \kappa_M = 2 \cdot 9.4$ nN [34]. The segment length of the S-state l_S will be a fit parameter.

The chemical potentials g_i^0, $i = B, S, M$, account for the free energy of base pairing and, since we set the interaction energies between neighboring segments of the same type to zero, $V_{ii} = 0$, also for the free energy gain due to base pair stacking [84]. They also correct for the fact that the reference state of the three different WLCs, which is $x = 0$ in the Helmholtz ensemble (constant extension x), cf. eq. (A.1), is not the same as contour and persistence lengths differ for B-, S-, and molten M-DNA. We choose $g_B^0 = 0$ and treat g_S^0 and g_M^0 as fitting parameters. Each of these parameters controls a distinct feature of the force-extension curve: The chemical potentials g_i^0 determine the critical forces, the segment lengths l_i affect the maximal extensions of each state and the off-diagonal V_{ij} control the cooperativity of the transitions. In appendix A.6 we summarize all parameter values that are used. We stress that our approach not necessarily implies the existence but merely allows for the possibility of observing three distinct states. The actual occurrence of the disputed S-state depends on the values of the parameters after fitting and is not presupposed in the model.

Force extension curve

In fig. 2.2 force extension curves based on three different levels of approximation are compared, using the same parameters that we extracted from DNA stretching data as will be detailed below. It turns out that the force extension curve obtained via the

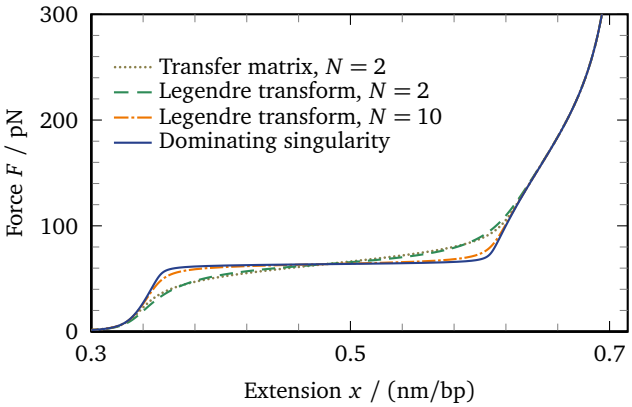


Figure 2.2.: Comparison of force extension curves obtained by different methods for $\hat{c} = 0$. The curve obtained via the exact transfer matrix calculation eq. (2.20) is already for $N = 2$ accurately reproduced by the approximate Legendre transformation eq. (2.13). The dominating singularity method, eqs. (2.16) or – equivalently – (2.21), is strictly valid in the thermodynamic limit but agrees with the Legendre transform already for a modest value of $N = 10$. The units of the abscissa is extension per base pair (bp). Parameters for λ-DNA in the absence of DDP are used, see appendix A.6.

Legendre transformation route eq. (2.13) (dashed line) is a very good approximation of the results obtained from the exact transfer matrix results eq. (2.20) (dotted line) already for $N = 2$. For $N = 10$ and larger virtually no differences between these two approaches are detectable. The deviations from the dominating singularity route, eq. (2.16) (solid line), which gives a result independent of N, are somewhat larger. But one sees that already the Legendre transform for $N = 10$ (dash-dotted line) matches the dominating singularity result very closely. Therefore the use of the dominating singularity, eq. (2.16), or the largest transfer matrix eigenvalue, eq. (2.21), is a very good approximation already for oligonucleotides and will be used in the rest of this work.

In fig. 2.3 experimental stretching curves of λ-DNA with and without DDP (cisplatin) are presented [47]. When B-DNA is converted into S-DNA or M-DNA the base stacking is interrupted which gives rise to an interfacial energy between B and S as well as between B and M of the order of the stacking energy [37, 38]. For untreated DNA, we use the value $V_{\mathrm{BS}} = V_{\mathrm{SB}} = V_{\mathrm{BM}} = V_{\mathrm{MB}} = 1.2 \cdot 10^{-20}$ J. Later we will show that variations down to $0.8 \cdot 10^{-20}$ J do not change the resulting curves much. V_{SM} is presumably small as the stabilizing stacking interactions are already disrupted [34], we thus set $V_{\mathrm{SM}} = V_{\mathrm{MS}} = 0$ for the fits in fig. 2.3 – but we will come back to the issue of a non-zero V_{SM} later on. Cisplatin is thought to disrupt the stacking interaction between successive base pairs and thereby to reduce the cooperativity of the BS-transition. We incorporate this fact by setting all interfacial energies to zero, $V_{ij} = 0$, for DDP treated DNA. The three remaining undetermined parameters (l_{S}, g_{S}^0, g_{M}^0) have distinct consequences

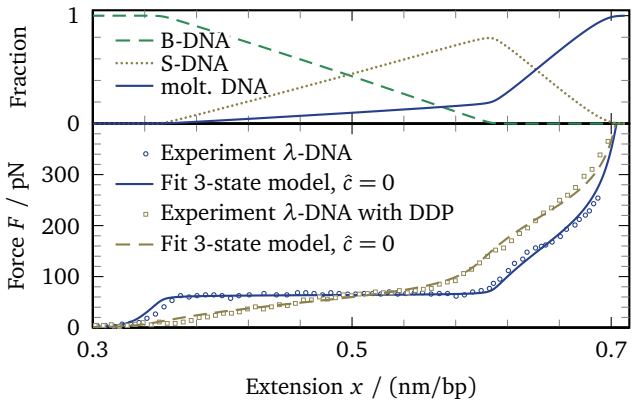


Figure 2.3.: Bottom panel: Force extension curve of double-stranded λ-DNA with and without DDP. Symbols denote experimental data [47], lines are fits with the three-state model for $\hat{c} = 0$. The main difference between the two curves is the lack of cooperativity in the BS-transition in the presence of DDP, which we take into account by choosing vanishing interaction energies $V_{ij} = 0$, $i, j = \mathrm{B}, \mathrm{S}, \mathrm{M}$. Top panel: Fraction N_i/N of segments in the different states as follows from eq. (2.24) in the absence of DDP.

on the force-extension curve. The segment length l_S and the chemical potential g_S^0 determine the position of the BS-plateau with respect to the polymer extension and applied force, respectively, while the chemical potential g_M^0 controls the force at which the second transition appears. Fitting to experimental data is thus straightforward and yields for untreated λ-DNA $l_\mathrm{S} = 6.1\,\text{Å}$, $g_\mathrm{S}^0 = 1.6 \cdot 10^{-20}\,\mathrm{J}$, $g_\mathrm{M}^0 = 2.4 \cdot 10^{-20}\,\mathrm{J}$ and for λ-DNA in the presence of DDP (cisplatin) $l_\mathrm{S} = 6.0\,\text{Å}$, $g_\mathrm{S}^0 = 1.2 \cdot 10^{-20}\,\mathrm{J}$, $g_\mathrm{M}^0 = 2.8 \cdot 10^{-20}\,\mathrm{J}$, see fig. 2.3. We also fit the number of monomers N and allow for an overall shift along the x-axis. The main difference between the two stretching curves is the cooperativity of the BS-transition, which is controlled by the interfacial energies V_BS and V_BM. Note that, although the over-stretching transition is quite sharp for DNA without DDP, it is not a phase transition in the strict statistical mechanics sense. A true phase transition arises only for $\hat{c} > 1$, as will be shown in the next section. The fitted value of g_M^0 is about two times larger than typical binding energies [84] for pure DNA. As a possible explanation, we note that the force extension curve of DNA without DDP exhibits pronounced hysteresis (especially at higher force), which will increase the apparent binding energy due to dissipation effects [69]. Any statements as to the stability of S-DNA based on our fitting procedures are thus tentative. However, such complications are apparently absent in the presence of DDP [47] which rules out kinetic effects as the reason for our relatively high fit values of g_M^0 and the stability of S-DNA. Cisplatin most likely stabilizes base pairs due to cross-linking and thus shifts the subtle balance between B-, S-, and M-DNA. Therefore, the relative stability of B-, S-, and M-DNA is sensitively influenced by co-solute effects. We note that even with $\hat{c} = 0$ a good fit of the data is possible. In the top panel of fig. 2.3 we show the

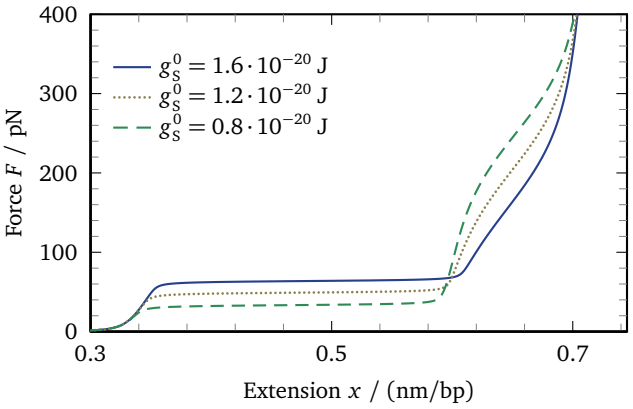


Figure 2.4.: The chemical potentials g_i^0 influence the transition forces. The solid line shows the force extension curve of λ-DNA without DDP. The dashed and the dotted line show stretching curves with different chemical potentials g_S^0 of the S-state while all other parameters are kept constant. Decreasing g_S^0 leads to a lower plateau force of the BS-transition and to an increased force of the SM-transition.

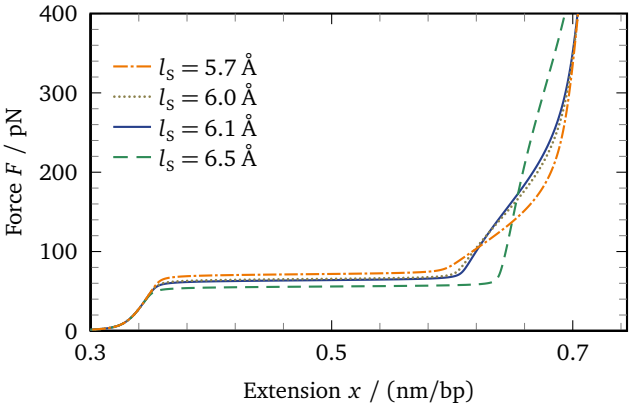


Figure 2.5.: The segment lengths l_i determine the extensions between which the transitions occur. The solid line shows the force extension curve of λ-DNA without DDP. The other curves are obtained by varying the segment length l_S of the S-state while keeping all other parameters constant. In addition, increasing l_S shifts the equilibrium towards the S-state.

fraction of segments in B-, S-, and M-states for untreated λ-DNA. There is a balanced distribution of bases in all three states across the full force range, in agreement with previous results [34].

Different parameters of the model are linked to certain features of the force extension curve. Fig. 2.4 illustrates how the chemical potentials g_i^0 influence the transition forces. Lowering g_S^0 leads to a decreasing plateau force of the BS-transition as the free energy difference between B- and S-states ($g_S(F) - g_B(F)$) decreases, see eq. (2.2). At the same time the force of the SM-transition increases as the free energy difference between S- and M-states ($g_M(F) - g_S(F)$) increases. The segment lengths l_i affect the maximal extensions between which the transitions occur. This can be seen in fig. 2.5, where l_S is varied. The effect of changing l_S is twofold. First, increasing l_S elongates the BS-plateau. As a side effect a larger l_S stabilizes the S-state as $g_S^{WLC}(F)$ is lowered. This can be concluded from eq. (A.1) in appendix A.1. Thus, increasing l_S shifts the equilibrium towards the S-state and has the same effects as decreasing g_S^0, cf. fig. 2.4. The off-diagonal entries V_{ij} of the interaction matrix determine the cooperativity of the transitions, see fig. 2.6. Lowering V_{BS} (dotted, dashed, and dash-dotted lines) smoothes the BS-transition and renders it less cooperative, which has been ob-

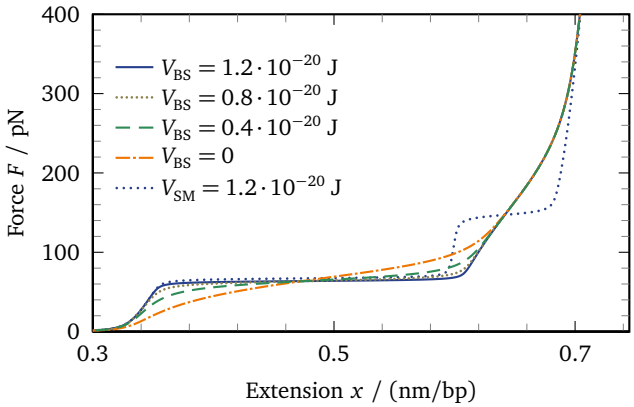


Figure 2.6.: The interfacial energies V_{ij} determine the cooperativity of the transitions. The solid line shows the force extension curve of λ-DNA without DDP. The other force extension curves are obtained by varying V_{BS} or V_{SM} while keeping all other parameters constant. Decreasing V_{BS} renders the BS-transition less cooperative. V_{SM} has similar effects on the melting transition: Increasing V_{SM} from its original value 0 to $1.2 \cdot 10^{-20}$ J (dotted blue curve) sharpens the denaturation transition.

served in the experiments by Krautbauer et al. [47] for λ-DNA with DDP, cf. fig. 2.3. The influence of the stretch modulus on force extension curves is quite moderate, see fig. 2.7. The solid line shows the force extension curve for λ-DNA without DDP. The dashed and the dotted line show the data with $\kappa_M = 40$ nN and $\kappa_M = 9.4$ nN, respectively, while keeping all other parameters unchanged. One observes that changing κ_M by a factor 4 results in deviations of the extension by just a few percent even at extreme forces $F = 400$ pN. The dash-dotted line shows a force extension trace where $\kappa_S = \kappa_M = 1.6$ nN, and $l_S = l_M = 0.59$ nm are used, which are the typical experimental values for two parallel ssDNA strands [30]. With this choice of parameters the force cannot discriminate between the S- and M-state anymore – the two states are degenerate. Therefore we have to set the chemical potentials equal, fit them again and obtain $g_S^0 = g_M^0 = 1.6 \cdot 10^{-20}$ J. Consequently, the shoulder indicating a denaturation transition does not appear in the force extension curve in contrast to the other curves shown.

2.4.2. Non-vanishing loop entropy, $\hat{c} \neq 0$

We now turn to non-zero loop exponents $\hat{c} \neq 0$ and in specific try to estimate $\hat{c}$ from the experimental stretching data. The partition function $\mathcal{Z}_{\hat{c}\neq 0}$ in eq. (2.9) exhibits two types of singularities. First, simple poles at $z = z_{p,i}$, which are the zeros of the denominator of eq. (2.9) and which are determined as the roots of the equation

$$-\frac{b_5 + b_6 z + b_7 z^2}{b_8 z + b_9 z^2} = \mathrm{Li}_{\hat{c}}(z/z_b)\,. \tag{2.26}$$

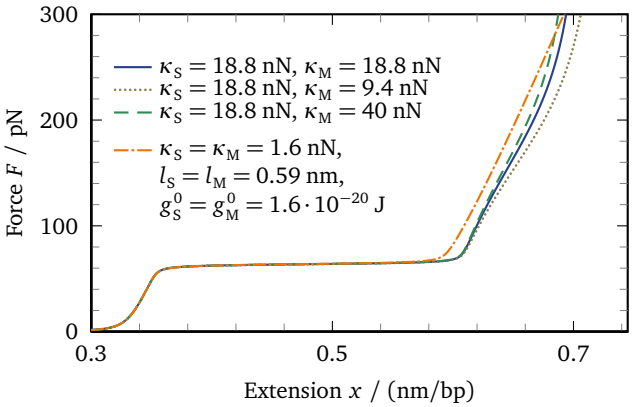
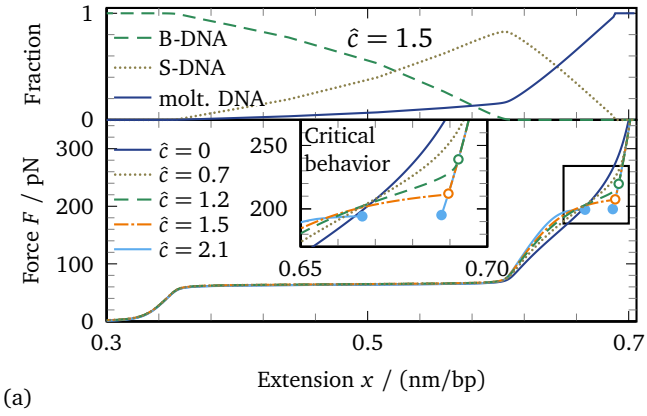


Figure 2.7.: The influence of the stretch modulus on force extension curves is quite moderate. The solid line shows the force extension curve of λ-DNA without DDP. The dashed and the dotted line show data with $\kappa_M = 40$ nN and $\kappa_M = 9.4$ nN, respectively, while keeping all other parameters unchanged. The dash-dotted line shows a force extension trace where $\kappa_S = \kappa_M = 1.6$ nN and $l_S = l_M = 0.59$ nm are used, which are the typical experimental values for two parallel ssDNA [30]. With this choice of parameters the force cannot discriminate between the S- and M-state anymore and the shoulder characterizing the SM-transition is not present.

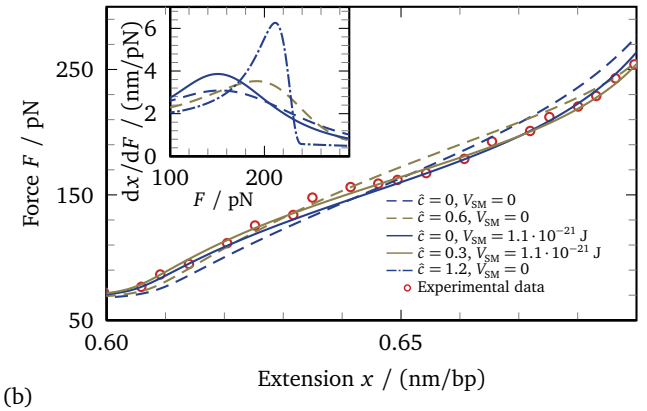
Second, a branch point that occurs at

$$z = z_b = e^{\beta g_M} . \tag{2.27}$$

The singularity with the smallest modulus is the dominant one [27, 78], and we define the critical force F_c as the force where both equations, eq. (2.26) and (2.27), hold simultaneously. For $\hat{c} \leq 1$ the dominant singularity is always given by the pole z_p and thus no phase transition is possible. For $1 < \hat{c} \leq 2$ a continuous phase transition occurs. By expanding eq. (2.26) around F_c one can show that all derivatives of the free energy up to order n are continuous, where $n \in \mathbb{N}$ is defined as the largest integer with $n < (\hat{c} - 1)^{-1}$ [31]. For instance, $\hat{c} = 3/2$ leads to a kink in the force extension curve. For $\hat{c} = 1.2$, this leads to a kink in $x'''(F)$. If $\hat{c} > 2$, the transition becomes first order and the force extension curve exhibits a discontinuity at $F = F_c$. In fig. 2.8a we plot force extension curves for different values of the loop exponent $\hat{c}$ with all other parameters fixed at the values fitted for untreated DNA. It is seen that finite $\hat{c}$ leads to changes of the force extension curves only at rather elevated forces. In order to see whether a finite $\hat{c}$ improves the comparison with the experiment and whether it is possible to extract the value of $\hat{c}$ from the data, we in fig. 2.8b compare the untreated DNA data with a few different model calculations for which we keep the parameters l_S, g_S^0, g_M^0, V_{BS}, V_{BM} fixed at the values used for the fit with $\hat{c} = 0$ in fig. 2.3. Allowing for finite $\hat{c}$ but fixing a zero domain wall energy between S- and M-regions, $V_{SM} = 0$, leads to an optimal exponent $\hat{c} = 0.6$ and slightly improves the fit to the data which show the onset of a plateau at a force of about 100 pN. The same



(a)



(b)

Figure 2.8.: Various force-extension curves of the three-state model with fit parameters for λ-DNA without DDP. (a) Lower panel: Force extension curves for different values of the loop exponent $\hat{c}$, showing no phase transition ($\hat{c} \leq 1$), a continuous ($1 < \hat{c} \leq 2$), or a first order phase transition ($\hat{c} > 2$). The critical forces are denoted by open circles. The inset is a magnification of the region around the transition. Upper panel: Fraction of bases in the three states for $\hat{c} = 3/2$. The critical transition, above which all bases are in the molten M-state, is discerned as a kink in the curves. (b) Comparison of experimental data (circles) and theory for $\hat{c} \neq 0$. The curve for $\hat{c} = 0$ and $V_{SM} = 0$, already shown in fig. 2.3, is obtained by fitting l_S, g_S^0, g_M^0 to the experimental data, the values of which are kept fixed for all curves shown. The curve for $V_{SM} = 0$ and $\hat{c} = 0.6$ results by fitting $\hat{c}$ and slightly improves the fit quality. The curve $\hat{c} = 0$ and $V_{SM} = 1.1 \cdot 10^{-21}$ J is obtained by fitting V_{SM}. The curve for $V_{SM} = 1.1 \cdot 10^{-21}$ J and $\hat{c} = 0.3$ is obtained by fitting $\hat{c}$ and keeping V_{SM} fixed. The inset shows the first derivative of $x(F)$ and illustrates that increasing $\hat{c}$ leads to a growing asymmetry around the transition region.

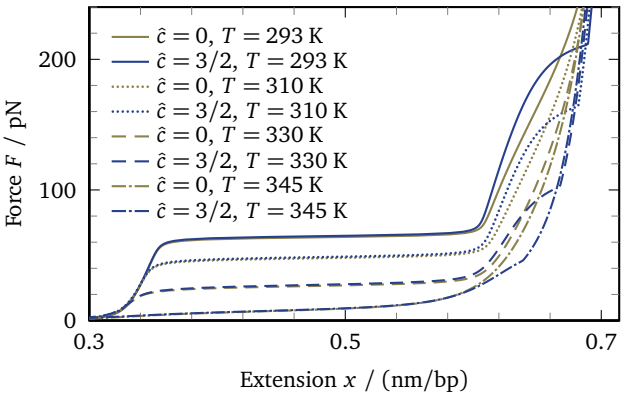


Figure 2.9.: Temperature dependence of the force extension curves with fit parameters for λ-DNA without DDP. Increasing temperature leads to a decrease of the BS-plateau force. In the presence of a true denaturing transition, i.e. for $\hat{c} > 1$, the critical force F_c decreases with increasing temperature and for $F > F_c$ the force extension curve follows a pure WLC behavior.

effect, however, can be produced by fixing $\hat{c} = 0$ and allowing for a finite V_{SM}, which yields the optimal value of $V_{SM} = 1.1 \cdot 10^{-21}$ J. Finally, fixing $V_{SM} = 1.1 \cdot 10^{-21}$ J and optimizing $\hat{c}$ yields in this case $\hat{c} = 0.3$ and perfect agreement with the experimental data. However, the significance of this improvement is not high, as the experimental data are quite noisy and possibly plagued by kinetic effects. What the various curves illustrate quite clearly, however, is that a non-zero exponent $\hat{c}$ leads to modifications of the stretching curves that are similar to the effects of a non-vanishing domain wall energy V_{SM}. Although V_{SM} should be considerably smaller than V_{BM} or V_{BS}, a finite value of $V_{SM} = 1.1 \cdot 10^{-21}$ J as found in the fit is reasonable and cannot be ruled out on general grounds. The maximal value of $\hat{c}$ is obtained for vanishing V_{SM} and amounts to about $\hat{c} = 0.6$. A value of $\hat{c} = 2.1$, which would be expected based on the entropy of internal DNA loops [31], on the other hand does not seem compatible with the experimental data, as follows from fig. 2.8a. This might have to do with the presence of nicks. Nicks in the DNA drastically change the topology of loops and result in a reduced loop exponent which is $\hat{c} = 0$ for an ideal polymer and 0.092 for a self avoiding polymer [31]. Therefore, the low value of $\hat{c}$ we extract from experimental data might be a signature of nicked DNA. Additional effects such as salt or co-solute binding to loops are also important. Therefore $\hat{c}$ can be viewed as a heuristic parameter accounting for such non-universal effects as well. We note in passing that $\hat{c}$ only slightly affects the BS-transition, as seen in fig. 2.8a.

2.5. Finite temperature effects

The temperature dependence of all parameters is chosen such that the force extension curves at $T = 20\,^{\circ}$C that were discussed up to now remain unchanged. The persistence lengths are modeled as $\xi_i(T) = \xi_i \cdot (T/293\,\text{K})^{-1}$. The S-state free energy is split into enthalpic and entropic parts as $g_S^0(T) = \tau_S(h_S - Ts_S)$ where we use $h_S = 7.14 \cdot 10^{-20}$ J and $s_S = 1.88 \cdot 10^{-22}$ J/K from Clausen-Schaumann et al. [46]. The correction factor $\tau_S = 0.98$ accounts for slight differences in the experimental setups and is determined

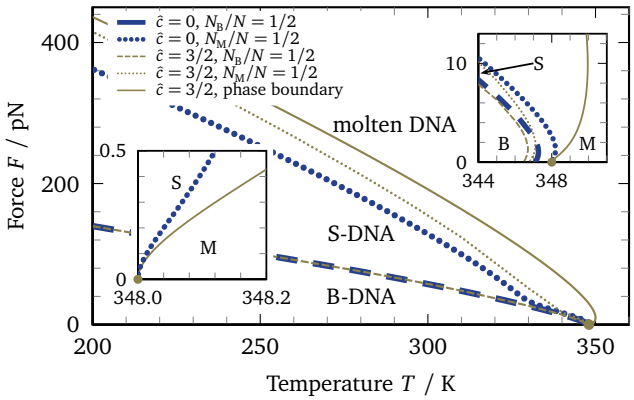


Figure 2.10.: The solid line shows the critical force F_c for $\hat{c} = 3/2$, at which a singularity occurs according to eq. (2.28). Phase boundaries for $\hat{c} = 0$ (thick lines) and $\hat{c} = 3/2$ (thin lines) are defined by $N_M/N = 0.5$ (dotted) and $N_B/N = 0.5$ (dashed). The melting temperature T_c is denoted by the dot. The insets show the behavior of the phase boundaries near the melting temperature, $F \propto \sqrt{T_c - T}$. Parameters for λ-DNA without DDP are used.

such that $g_S^0(T = 20\,^\circ\mathrm{C})$ equals the previously fitted value. The molten state energy $g_M^0(T) = (h_M - Ts_M)$ is also chosen such that $g_M^0(T = 20\,^\circ\mathrm{C})$ agrees with the previous fit value and that the resulting denaturing temperature in the absence of force agrees with experimental data. Assuming a melting temperature of $T_c = 348$ K for λ-DNA [85], we obtain $g_M^0(T) = 1.5 \cdot 10^{-19}\,\mathrm{J} - T \cdot 4.2 \cdot 10^{-22}\,\mathrm{J/K}$ for $\hat{c} = 0$ and $g_M^0(T) = 1.6 \cdot 10^{-19}\,\mathrm{J} - T \cdot 4.6 \cdot 10^{-22}\,\mathrm{J/K}$ for $\hat{c} = 3/2$. In fig. 2.9 we plot a few representative stretching curves for different temperatures. It is seen that increasing temperature lowers the BS-plateau and makes this transition less cooperative. Differences between $\hat{c} = 0$ and $\hat{c} = 3/2$ are only observed at elevated forces, where for $\hat{c} = 3/2$ one encounters a singularity characterized by a kink in the extension curves.

For $\hat{c} > 1$ the critical force F_c is defined as the force where the pole and the branch point coincide and eqs. (2.26) and (2.27) are simultaneously satisfied. The phase boundary in the force-temperature plane is thus defined by

$$-\frac{b_5 + b_6 z_b + b_7 z_b^2}{b_8 z_b + b_9 z_b^2} = \zeta_{\hat{c}}\,, \tag{2.28}$$

where b_i and z_b depend on T and F and $\zeta_{\hat{c}} = \mathrm{Li}_{\hat{c}}(1)$ is the Riemann zeta function. The phase boundary $F_c(T)$ for $\hat{c} = 3/2$ is shown in fig. 2.10 and agrees qualitatively with the experimental data of ref. [41]. For exponents $\hat{c} < 1$ no true phase transition exists. We therefore define crossover forces as the force at which half of the segments are in the molten state or in the B-state, i.e. $N_M/N = 1/2$ or $N_B/N = 1/2$. In fig. 2.10 we show these lines for both $\hat{c} = 0$ and $\hat{c} = 3/2$. Note that the parameters for the $\hat{c} = 0$ case are adjusted so that $N_M/N = 1/2$ at $T = 348$ K and $F = 0$. The broken lines on which $N_B/N = 1/2$ for $\hat{c} = 0$ and $\hat{c} = 3/2$ are virtually the same, showing again that loop entropy is irrelevant for the BS-transition. The S-state is populated in the area between the $N_B/N = 1/2$ and $N_M/N = 1/2$ lines, which for $T > 330$ K almost coalesce, meaning that at elevated temperatures the S-state is largely irrelevant. Force-induced re-entrance at constant temperature is found in agreement with previous two-state

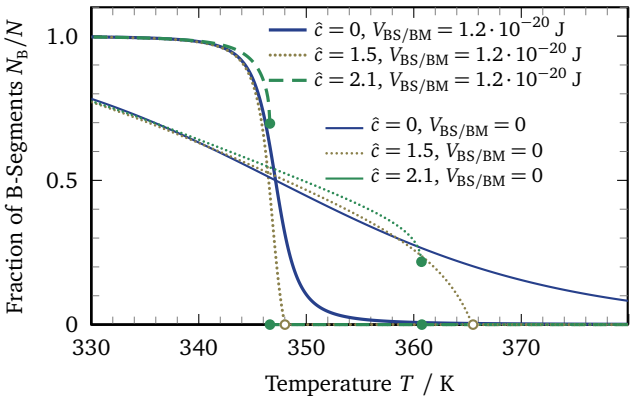


Figure 2.11.: Fraction of segments in the B-state, N_B/N, as a function of temperature for different loop exponents $\hat{c} = 0,\ 1.5,\ 2.1$ and for finite BS interfacial energy $V_{BS} = V_{BM} = 1.2 \cdot 10^{-20}$ J (bold lines) and for $V_{BS} = V_{BM} = 0$ (thin lines). Circles indicate the positions of the phase transition. For all curves parameters for λ-DNA without DDP are used, $g_M^0(T) = 1.5 \cdot 10^{-19}\,\text{J} - T \cdot 4.2 \cdot 10^{-22}\,\text{J/K}$ for $\hat{c} = 0$ and $g_M^0(T) = 1.6 \cdot 10^{-19}\,\text{J} - T \cdot 4.6 \cdot 10^{-22}\,\text{J/K}$ for $\hat{c} > 0$.

models [58, 64]. Re-entrance at constant force as found for a Gaussian model [58] is not reproduced, in agreement with results for a non-extensible chain [64].

As we have shown so far, a non-zero loop exponent $\hat{c}$ only slightly improves the fit of the experimental stretching data and the optimal value found is less than unity. This at first sight seems at conflict with recent theoretical work that argued that a loop exponent larger than $\hat{c} = 2$ is needed in order to produce denaturation curves (at zero force) that resemble experimental curves in terms of the steepness or cooperativity of the transition [31]. To look into this issue, we plot in fig. 2.11 the fraction of native base pairs, N_B/N as a function of temperature for zero force and different parameters. As soon as the domain-wall energies V_{BS} and V_{BM} are finite, the transition is quite abrupt, even for vanishing exponent $\hat{c}$. Therefore even loop exponents $\hat{c} < 1$ yield melting curves which are consistent with experimental data, where melting occurs over a range of the order of 10 K [85, 86].

2.6. Conclusions

The fact that the domain-wall energy due to the disruption of base pair stacking and the loop entropy embodied in the exponent $\hat{c}$, give rise to similar trends and sharpen the melting transition has been realized and discussed before [36, 86]. The present three-state model and the simultaneous description of experimental data, where the denaturation is induced by application of force and by temperature allows to disentangle the influence of these two important effects. By the application of a force, the de-stacking and the loop formation possibly occur subsequently, allowing to fit both

parameters separately. As our main finding, we see that for a finite domain-wall energy $V_{\mathrm{BS}} = V_{\mathrm{BM}} > 0$, the additional influence of the loop exponent on the force stretching curves and the denaturation curves is small. In fact, the optimal value for $\hat{c}$ turns out to be of the order of $\hat{c} \approx 0.3$-0.6, even if we choose a vanishing value $V_{\mathrm{SM}} = 0$. This estimate for $\hat{c}$ is smaller than previous estimates. One reason for this might be nicks in the DNA. So it would be highly desirable to redo stretching experiments with un-nicked DNA with which the value of $\hat{c}$ under tension could be determined. In the next chapter we will show that – if the sequence allows – secondary structure formation in molten loops significantly alter the denaturation transition and a pronounced sequence dependence should be observed. In our approach we allow for an intermediate S-state, but do not enforce its existence – it rather emerges as a result of the values of the fitting parameters. The second transition at high forces of about $F \approx 200$ pN which is seen in the experimental data used in this work, inevitably leads via the fitting within our three-state model to this intermediate S-DNA state. But we stress that the occurrence of such an intermediate S-state depends on the fine-tuning of all model parameters involved, which suggests that in experiments the S-state stability also sensitively depends on the precise conditions. One drawback of the current model is that sequence effects are not taken into account. This means that our fitted parameters have to be interpreted as coarse-grained quantities, which average over sequence disorder. Calculations with explicit sequences have been done for short DNA strands but should in the future be doable for longer DNA as well.

CHAPTER 3

SECONDARY STRUCTURES OF HOMOPOLYMERIC SINGLE-STRANDED NUCLEIC ACIDS

3.1. Introduction

3.1.1. Motivation

In the previous chapter only double stranded nucleic acids have been considered and base pairing within the same strand was not allowed. However, base pairing within single stranded nucleic acids is known to be vital for their structure and functioning [17]. Therefore, we will now turn to single stranded nucleic acids and in particular to RNA. Theory on folding of single stranded nucleic acids vastly relies on the idea of hierarchical folding proposed by Tinoco et al. [29, 87], who claim that given a sequence (the primary structure), the secondary structure (the list of all base pairs) forms independently of the tertiary structure (the overall three-dimensional arrangement of all atoms). This is in contrast to the protein folding problem, which in general does not feature these well separated energy scales between the different structural levels and hence is more involved [88]. The idea of hierarchical folding therefore suggests to solely focus on the secondary structure, and to neglect the tertiary structure. This constitutes a major simplification and enables to calculate partition functions exactly and to predict the secondary structure formed by a given RNA sequence. In 1968 de Gennes [89] was the first to calculate the partition function of an ideal homopolymeric RNA chain by using a propagator formalism and solving the partition function by means of a singularity analysis of the generating functions. Due to his real space approach for an ideal polymer, the loop exponent was $c = 3/2$. Like in the previous chapter, the loop exponent c characterizes the logarithmic entropy contribution $\propto \ln m^{-c}$ of a loop of length m. Ten years later, Waterman and Smith [90] devised a recursion relation appropriate for the partition function of folded RNA, which now lies at the heart of most RNA secondary structure and free energy prediction algorithms currently used. Subsequently, several theoretical models were developed to study RNA and DNA: those

were focused on melting [1, 91, 92], stretching [3, 50, 58, 93–95], unzipping [96], translocation [97], salt influence [4, 57, 98, 99], pseudoknots [100, 101], and the influence of the loop exponent [1, 3, 86, 102]. In this context, an interesting question in connection with the melting of double stranded DNA (dsDNA) arises: Do secondary structure elements form in the single strands inside denatured dsDNA loops or not? Formation of such secondary structures in dsDNA loops would mean that inter-strand base pairing between the two strands – being responsible for the assembly of the double helix – is in competition with intra-strand pairing, where bases of the same strand interact. This question is not only important for the thermal melting of dsDNA but also for DNA transcription, DNA replication and the force-induced overstretching transition of DNA [15, 45]. Theory for single strand nucleic acids is accompanied and inspired by the advance of single molecule techniques [30, 45, 103–106], which allow to study single chains of nucleic acids under tension and varying solution conditions and thereby yield unprecedented insights into the behavior and folding properties of these essential molecules.

3.1.2. Outline

In this chapter, the influence of the loop exponent c on the behavior of RNA subject to varying temperature and external force is studied. Since all results obtained for RNA are also valid for DNA on our relative primitive level of modeling, we will mostly explicitly refer to RNA in this chapter but note that in principle all our results carry over to single stranded DNA molecules, as well. We neglect sequence effects and consider a long homopolymeric, single stranded RNA molecule. A closed form expression for the partition function is derived, which allows to study the thermodynamic behavior in detail. The phase diagram in the force-temperature plane is obtained. We find that the existence of a temperature induced phase transition crucially depends on the value of the loop exponent c: at vanishing force a melting transition is possible only for the limited range of loop exponents $2 < c < 2.479$. $c \approx 2.1$ is a typical exponent that characterizes the entropy of loops usually encountered in RNA structures – hairpin loops, internal loops, multi-loops with three or more emerging helices. That means that RNA molecules can experience a transition between a folded and an unfolded state. This is relevant for structure formation in DNA or RNA single strands and can in principle be tested in double laser trap force clamp experiments [107]. Our findings also have implications for the denaturation of double helical nucleotides (e. g. dsDNA). Since intra-strand and inter-strand base pairing competes, secondary structure formation of the single strands inside denatured regions of the duplex has to be taken into account in a complete theory of dsDNA melting. In the case where intra- and inter-strand base pairing occurs with the same weight, the classical Poland-Scheraga mechanism for the melting of a DNA duplex is augmented by our transition, as the Poland-Scheraga theory is only valid in the case where no intra-strand base pairing is possible; in chapter 2 the case where no intra-strand pairing is possible has been considered. Some simple arguments in the case where intra-strand pairing is possible suggest that the loop ex-

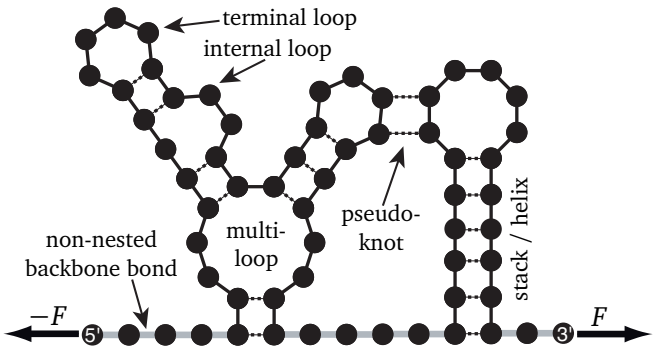


Figure 3.1.: Schematic representation of the secondary structure of an RNA molecule. Dots represent one base, i. e. cytosine, guanine, adenine, or uracil. Solid lines denote the sugar-phosphate backbone bonds, broken lines base pairs, and thick gray lines the non-nested backbone bonds, which are counted by the variable M, here $M = 11$. The thick arrows to either side illustrate the force F applied to the 5'- and 3'-end.

ponent governing the entropy of inter-strand loops is renormalized due to intra-strand secondary structure formation and takes on an effective universal value that only depends on whether the inter-strand loops are symmetric (consisting of two strands of the same length) or asymmetric. The resulting duplex melting transition is universal and turns out to be strongly discontinuous in the first, symmetric case, and on the border between continuous and discontinuous in the second, asymmetric case. In the case when both intra- and inter-strand base pairing is possible, but inter-strand base pairs are slightly stronger than intra-strand base pairs, the situation is more complex. All these effects can be studied experimentally. We make explicit suggestions for ds-DNA sequences, with which the formation of intra-strand secondary structures inside inter-strand loops can be selectively inhibited or favored.

3.2. Model for homopolymeric RNA

Single stranded RNA is modeled as a one-dimensional chain. RNA folding can be separated into three steps, which occur subsequently and do not influence each other to a fairly good approximation. This idea of the RNA folding process is known as hierarchical folding [29, 87]. The primary structure of RNA is the mere sequence of the four bases cytosine (C), guanine (G), adenine (A), and uracil (U). Due to base pairing, i. e. either the specific interaction of C with G or the interaction of A with U, the secondary structure is formed leading to helices and loops as the structural building blocks of an RNA secondary structure, see fig. 3.1. Only after the secondary structure has formed tertiary contacts arise. Pseudoknots [108, 109], helix stacking – the interaction of two helices emerging from the same loop – and base triples [110] as well as the overall three-dimensional arrangement of the molecule are considered as parts of the tertiary structure. The main assumption of hierarchical folding is that

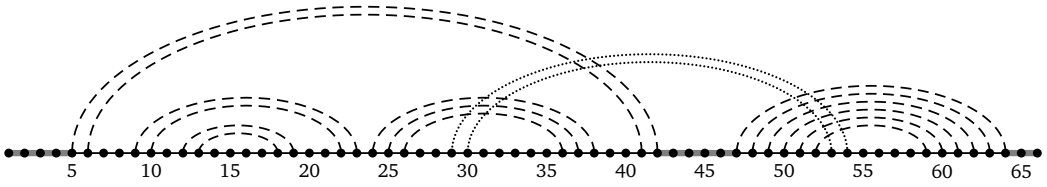


Figure 3.2.: The arc diagram is a representation of the secondary structure depicted in fig. 3.1. A dot represents one base. Solid lines denote the backbone bonds and thick gray lines the non-nested backbone bonds. Dashed arcs denote hydrogen bonds between two bases. A pseudoknot (dotted arc) is recognized here as crossing arcs. If no pseudoknots are present the structure is hierarchical, meaning that substructures are either nested or juxtaposed.

the tertiary structure formation operates only on already existing secondary structure elements [29, 87]. Although cases are known where this approximation breaks down, it generally constitutes a valid starting point [111].

On an abstract level, the secondary structure is given by the list of all base pairs present in the molecule with the constraint that a base can be part of at most one pair. Since pseudoknots are not allowed, we have for any two base pairs (i, j) and (k, l) with $i < j$, $k < l$, and $i < k$ either $i < k < l < j$ or $i < j < k < l$ [101]. This imposes a hierarchical order on the base pairs meaning that two base pairs are either nested and part of the same substructure or are independent and part of different substructures, fig. 3.2. Base pairs are stabilized by two different interactions. First, by hydrogen bonds between complementary bases and, second, by the stacking interaction between neighboring base pairs, which are accounted for by the sequence independent – we are considering homopolymeric RNA – parameters $g_{\mathrm{h}}^{\mathrm{hb}}$ and $g_{\mathrm{h}}^{\mathrm{stack}}$, respectively. A helix with h base pairs consequently has the free energy $h(g_{\mathrm{h}}^{\mathrm{hb}} + g_{\mathrm{h}}^{\mathrm{stack}}) - g_{\mathrm{h}}^{\mathrm{stack}} + g_{\mathrm{h}}^{\mathrm{i}}$, where $g_{\mathrm{h}}^{\mathrm{i}}$ is a helix initiation free energy. Therefore, the hydrogen bonding and the stacking interaction can be combined to yield the binding energy per base pair $\varepsilon = -(g_{\mathrm{h}}^{\mathrm{hb}} + g_{\mathrm{h}}^{\mathrm{stack}})$, which we define to be positive. ε can be measured experimentally by duplex hybridization [112] and contains the binding free energies as well as the extensive part of configurational polymer entropy. Here, we describe the binding free energy by a single, sequence independent parameter [1]

$$w = \exp(\varepsilon/(\mathrm{k_B}T)) \,, \tag{3.1}$$

where w is the statistical weight of a bound base pair, T is the absolute temperature, and $\mathrm{k_B}$ the Boltzmann constant. Further, the stacking interaction appears as an additional contribution to the helix initiation free energy and we define $\mathcal{G}_{\mathrm{h}}^{\mathrm{init}} = g_{\mathrm{h}}^{\mathrm{i}} - g_{\mathrm{h}}^{\mathrm{stack}}$. This is a model for homopolymeric RNA, which can be realized experimentally with synthetic alternating sequences $[\mathrm{AU}]_{N/2}$ or $[\mathrm{GC}]_{N/2}$. It has also been argued this homopolymeric model describes random RNA above the glass transition [113].

The non-extensive contribution of the free energy of a loop is given by

$$\mathcal{G}_{\mathrm{l}}^{\mathrm{conf}} = -\mathrm{k_B}T \ln m^{-c} \tag{3.2}$$

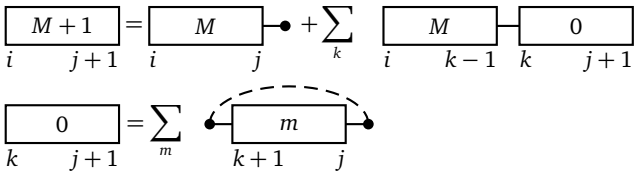


Figure 3.3.: Illustration of the recursion scheme for the canonical partition function in eqs. (3.3) and (4.11). Boxes denote partition functions of substrands (the range is given by the subindices). The numbers inside a box give the number of non-nested backbone bonds. To calculate the partition function of a strand ranging from i through $j+1$ consider the partition function of a strand ranging from i through j and add base number $j+1$, which may (right term in first row) or may not (left term in first row) establish a base pair with base number k. In the second row $Q^0_{k,j+1}$ is calculated by closing structures with m non-nested bonds with a hydrogen bond (dashed line). For homopolymeric RNA, the sequence dependence drops out and only the lengths of the substrands, $N = j-i$, $n = k-1-i$, $N-n = j+1-k$, need to be considered, see eq. (3.3).

and describes the entropy difference between an unconstrained polymer and a looped polymer. The loop exponent c is $c_{\text{ideal}} = 3/2$ for an ideal polymer and $c_{\text{SAW}} = d\nu \simeq 1.76$ for an isolated self avoiding loop with $\nu \simeq 0.588$ in $d = 3$ dimensions [74], see also appendix A.3. However, helices, which emerge from the loop, increase c even further. In the asymptotic limit of long helical sections renormalization group predicts $c_l = d\nu + \sigma_l - l\sigma_3$ for a loop with l emerging helices [75, 102], where $\sigma_l = \epsilon l(2-l)/16 + \epsilon^2 l(l-2)(8l-21)/512 + \mathcal{O}(\epsilon^3)$ in an $\epsilon = 4-d$ expansion. One obtains $c_1 = 2.06$ for terminal, $c_2 = 2.14$ for internal loops and $c_4 = 2.16$ for a loop with four emerging helices. For larger l the ϵ expansion prediction for c_l becomes unreliable. One sees that the variation of c over different loop topologies that appear in the native structures of RNA is quite small, which justifies our usage of the same exponent c for loops of all topologies that occur in a given RNA secondary structure in the following.

3.3. Partition functions

3.3.1. Canonical partition function

As we neglect pseudoknots, only hierarchical structures are present, which allows to write down a recursion relation for the partition function. Further, as we consider homopolymers and omit sequence effects by using a constant base pairing weight w, the system is translationally invariant. Hence, the canonical partition function $Q^M_{i,j}$ of a strand ranging from base i at the 5'-end through j at the 3'-end depends only on the total number of segments $N = j-i$ and on the number of non-nested backbone bonds M. A non-nested bond is defined as a backbone bond, which is neither part of a helix nor part of a loop. It is outside all secondary structure elements and therefore contributes to the extension, which couples to an external stretching force and which

can be observed for example in force spectroscopy experiments [1, 95, 97], see figs. 3.1 and 3.2. The recursion relations for Q_N^M can be written as

$$Q_{N+1}^{M+1} = \frac{\nu_\mathrm{f}(M+1)}{\nu_\mathrm{f}(M)} \left[Q_N^M + w \sum_{n=M}^{N-1} Q_n^M Q_{N-n}^0 \right] \tag{3.3a}$$

and

$$Q_{N-n}^0 = \sum_{m=-1}^{N-n-2} \frac{Q_{N-n-2}^m}{\nu_\mathrm{f}(m)} \frac{\exp\left(-\mathcal{G}_\mathrm{h}^\mathrm{init}\Theta(m-2)/(\mathrm{k_B}T)\right)}{(m+2)^c} \tag{3.3b}$$

and is illustrated in fig. 3.3. The Heaviside step function is $\Theta(m) = 0$ if $m \leq 0$ and $\Theta(m) = 1$ if $m > 0$. Eq. (3.3a) describes the elongation of an RNA structure by either adding an unpaired base (first term) or by adding an arbitrary substrand Q_{N-n}^0 that is terminated by a helix. Eq. (3.3b) constructs Q_{N-n}^0 by closing structures with m non-nested bonds, summed up in Q_{N-n-2}^m, by a base pair. $\nu_\mathrm{f}(M)$ denotes the number of configurations of a free chain with M links and can be completely eliminated from the recursion relation by introducing the rescaled partition function $\tilde{Q}_N^M = Q_N^M/\nu_\mathrm{f}(M)$. We set $\mathcal{G}_\mathrm{h}^\mathrm{init} = 0$ for computational simplicity and combine eqs. (3.3a) and (3.3b), which leads to the final recursion relation

$$\tilde{Q}_{N+1}^{M+1} = \tilde{Q}_N^M + w \sum_{n=M}^{N-1} \sum_{m=-1}^{N-n-2} \frac{\tilde{Q}_n^M \tilde{Q}_{N-n-2}^m}{(m+2)^c} \,, \tag{3.4}$$

with the boundary conditions $\tilde{Q}_{-1}^{-1} = 1$, $\tilde{Q}_M^N = 0$ for $M > N$, $N < 0$, or $M < 0$. The thermodynamic limit of an infinitely long RNA chain is described by the canonical Gibbs ensemble, which is characterized by a fixed number of segments N, but a fluctuating number of non-nested backbone bonds M. Therefore, we introduce the unrestricted partition function

$$Z_N(s) = \sum_{M=0}^{\infty} s^M \tilde{Q}_N^M \,, \tag{3.5}$$

which contains the influence of an external force F via the statistical weight s of a non-nested backbone bond. For RNA with no force applied to the ends, one has $s = 1$. We model the RNA backbone elasticity by the freely jointed chain (FJC) model, where the weight of a non-nested backbone bond subject to an external force is given by [114]

$$s = \frac{1}{4\pi} \int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta \, \mathrm{e}^{-\beta F b_\mathrm{ss} \cos\theta} = \frac{\sinh(\beta F b_\mathrm{ss})}{\beta F b_\mathrm{ss}} \,; \tag{3.6}$$

here, we introduced the inverse thermal energy $\beta = (\mathrm{k_B}T)^{-1}$ and the Kuhn length b_ss.

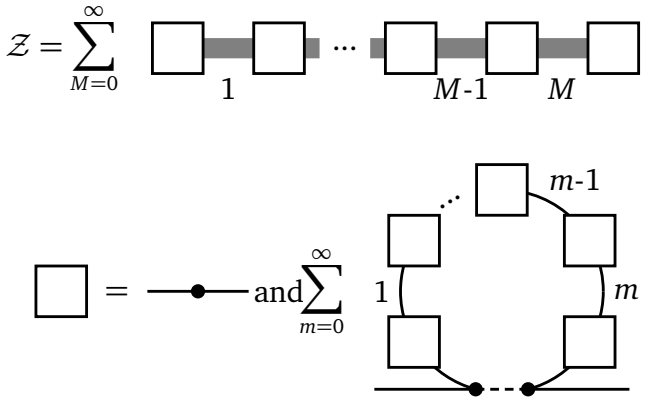


Figure 3.4.: Structure of the grand canonical partition function $\mathcal{Z}(z,s)$ according to eqs. (3.7) and (3.11). The grand canonical partition function of the Gibbs ensemble is a sum over all numbers of non-nested backbone bonds (thick gray lines) with statistical weight sz. Between two adjacent backbone bonds can be either a single nucleotide (dot), with statistical weight 1, or a structure with weight $\mathcal{Z}_\mathrm{b}$, whose first and last base are paired. The white squares serve as wildcards for either possibility and have the statistical weight $\kappa(w,z)$. Thin black lines depict backbone bonds that are part of a helix or loop and have statistical weight z.

3.3.2. Grand canonical partition function

For studying the phase transition and the critical behavior, it is useful to introduce the generating function or grand canonical partition function

$$\mathcal{Z}(z,s) = \sum_{N=0}^{\infty} z^N Z_N(s) = \sum_{N=0}^{\infty}\sum_{M=0}^{\infty} z^N s^M \tilde{Q}_N^M \,, \tag{3.7}$$

where $z = \exp(\beta\mu)$ is the fugacity of a segment. Performing the weighted double sum $\sum_{N=-1}^{\infty}\sum_{M=-1}^{\infty} z^N s^M$ on both sides of eq. (3.4) yields

$$(sz)^{-1}\mathcal{Z} = (sz)^{-1} + \mathcal{Z} + \big((sz)^{-1} + \mathcal{Z}\big)(\kappa - 1)\,, \tag{3.8}$$

which can be solved and one obtains the generating function

$$\mathcal{Z}(z,s) = \frac{\kappa(w,z)}{1 - sz\kappa(w,z)}\,. \tag{3.9}$$

Here we have defined $\kappa(w,z) = 1 + \mathcal{Z}_\mathrm{b}(w,z)$ as the grand canonical partition function of RNA structures with zero non-nested backbone bonds, i. e. structures which consist of just one nucleotide or structures where the terminal bases are paired,

$$\kappa(w,z) = 1 + \mathcal{Z}_\mathrm{b}(w,z) = 1 + wz^2 \sum_{N=-1}^{\infty}\sum_{M=-1}^{N} z^N \frac{\tilde{Q}_N^M}{(M+2)^{-c}}\,. \tag{3.10}$$

Eq. (3.9) has an instructive interpretation, which becomes clear by expanding the fraction in a geometric series

$$\mathcal{Z}(z,s) = \sum_{M=0}^{\infty} s^M z^M \kappa^{M+1} = \sum_{M=0}^{\infty} (1+\mathcal{Z}_\mathrm{b}) \cdot \big(sz(1+\mathcal{Z}_\mathrm{b})\big)^M \,, \tag{3.11}$$

where sz is the statistical weight for a backbone segment which is not nested. Between two adjacent segments we have the possibility to put either a single nucleotide (with statistical weight 1) or a structure whose first and last bases are paired (with statistical weight $\mathcal{Z}_\mathrm{b}$). See fig. 3.4 for an illustration.

In order to determine the function $\kappa(w,z)$, we compare the coefficients of the power series in s in eqs. (3.7) and (3.11) and obtain $z^M \kappa^{M+1} = \sum_{N=M}^{\infty} z^N \tilde{Q}_N^M$. The lower summation index is due to exchanging the summations in eq. (3.7) bearing in mind that $\tilde{Q}_N^M = 0$ for $M > N$. This identity can be inserted into eq. (3.10) and yields the equation

$$\kappa(w,z) - 1 = \frac{w}{\kappa(w,z)} \mathrm{Li}_c(z\kappa(w,z)) \,, \tag{3.12}$$

which determines $\kappa(w,z)$. $\mathrm{Li}_c(z\kappa) = \sum_{m=1}^{\infty} z^m \kappa^m / m^c$, for $z\kappa \leq 1$, is the polylogarithm [76], see appendix A.4. We introduce

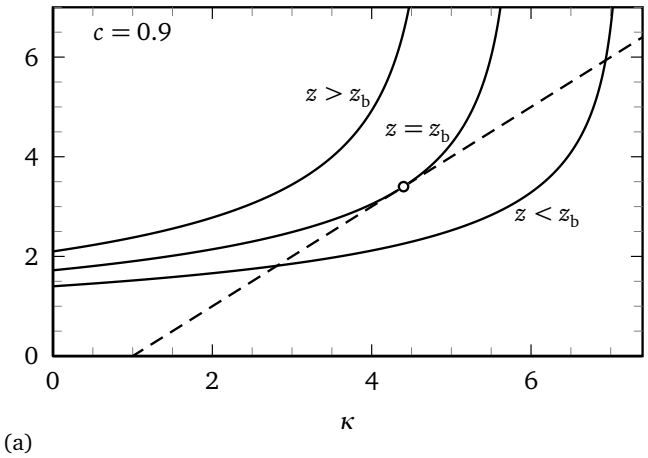
$$h(\kappa,z) = \frac{w}{\kappa} \mathrm{Li}_c(z\kappa) \tag{3.13}$$

and rewrite eq. (3.12) as $\kappa - 1 = h(\kappa,z)$. Eq. (3.12) has at most two positive and real solutions as can be seen from fig. 3.5, where we plot the two sides of eq. (3.12). $\kappa(w,z)$ is a continuous, monotonically increasing function of z with $\kappa(w,0) = 1$ as follows from eq. (3.10). Therefore, only the smallest positive root of eq. (3.12) yields the correct $\kappa(w,z)$. For $z \to 0$ there is always a positive and real solution for $\kappa(w,z)$. Increasing z increases $\kappa(w,z)$ until eventually, at $z = z_\mathrm{b}$, the real solution for $\kappa(w,z)$ vanishes. Thus, $\kappa(w,z)$ has a branch point at $z = z_\mathrm{b}$. Depending on the value of the loop exponent c, the polylogarithm $\mathrm{Li}_c(z\kappa)$ and $h(\kappa,z)$ (a) are divergent for $c \leq 1$, (b) are finite, but feature a diverging slope for $1 < c \leq 2$, or (c) have a finite value and derivative for $2 < c$ at $z\kappa = 1$, see fig. 3.5. This will become important later, when the existence of phase transitions is studied.

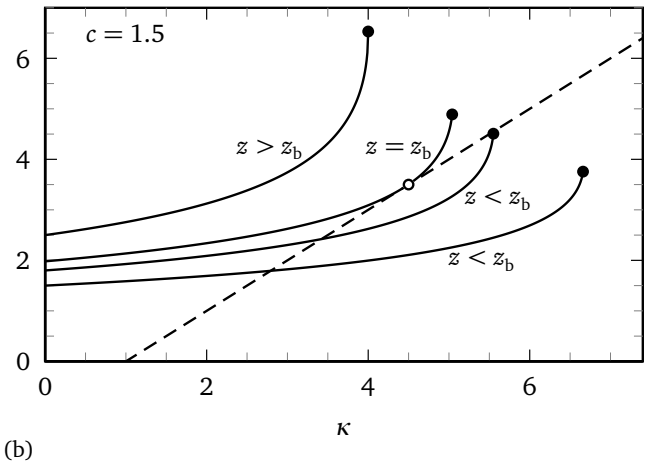
3.3.3. Back-transform to canonical ensemble

Since the thermodynamic limit $N \to \infty$ is defined in the canonical Gibbs ensemble, we now demonstrate how to obtain $Z_N(s)$, eq. (3.5), from $\mathcal{Z}(z,s)$, eq. (3.9). For large systems, $N \gg 1$, the canonical partition function is given by the dominant singularity $z_\mathrm{d}(s)$ of $\mathcal{Z}(z,s)$, which is defined as the singularity which is nearest to the origin in the complex z-plane, see eq. (2.16) and appendix A.5 [3, 78, 79]. In particular if $\mathcal{Z}(z,s) \sim K(s)\big(z_\mathrm{d}(s) - z\big)^{-\alpha}$ with $K(s)$ independent of z, we obtain

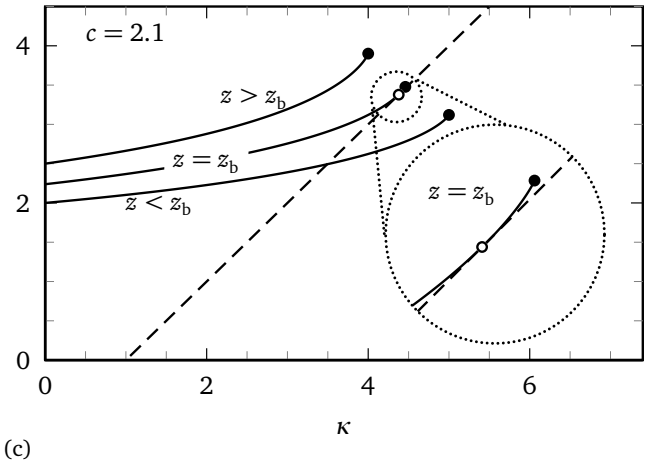
$$Z_N(s) \sim z_\mathrm{d}^{-N}(s) N^{\alpha-1} K(s)\, z_\mathrm{d}^{-\alpha}(s)/\Gamma(\alpha) \,, \tag{3.14}$$



(a)



(b)



(c)

Figure 3.5.: Graphical solution of the equations, which determine $\kappa(w,z)$, z_b, z_p. The functions $\kappa - 1$ (dashed line) and $h(\kappa,z)$ (solid lines), eq. (3.13), are plotted for $w = 10$ and different values of the fugacity z and the loop exponent c, (a) $c = 0.9$ (b) $c = 3/2$ (c) $c = 2.1$. Points at which both curves intersect are solutions of eq. (3.12) and determine $\kappa(w,z)$. In case of two positive solutions the smaller yields the correct solution as $\kappa(w,z)$ is to be a continuous, monotonically increasing function of z with $\kappa(w,0) = 1$, eq. (3.10). Points at which both curves are adjacent to each other (open circles) determine the branch point z_b. Points at which $z\kappa = 1$ (dots) determine the position of the pole z_p in the absence of force, $s = 1$.

where $\Gamma(\alpha)$ is the gamma function [115]. Therefore, the Gibbs free energy reads to leading order

$$\mathcal{G}/(\mathrm{k_B}T) = -\ln Z_N \sim N \ln z_\mathrm{d}(s) + (1-\alpha)\ln N \,. \tag{3.15}$$

In fact, $\mathcal{Z}(z,s)$ features two relevant singularities. First, the branch point $z_\mathrm{b}(w)$ of $\kappa(w,z)$, which is independent of s, and second a simple pole $z_\mathrm{p}(w,s)$, where the denominator of $\mathcal{Z}(z,s)$ vanishes, see eq. (3.9). Depending on which singularity has the smallest modulus, the molecule can be in different phases. In the following sections it will turn out that the low temperature, compact or folded phase is associated with z_b, whereas the high temperature, extended or unfolded phase is characterized by z_p.

Let us consider the branch point first. It can be seen from fig. 3.5, that for $z < z_\mathrm{b}$ at least one real solution of eq. (3.12) exists, where the smaller solution determines $\kappa(w,z)$. Right at $z = z_\mathrm{b}$ the two solutions merge and the slope of h at the tangent point is $h'(\kappa, z_\mathrm{b}) = \partial h(\kappa, z_\mathrm{b})/\partial\kappa = 1$. This yields the equation for the position of the branch point singularity $z_\mathrm{b}(w)$, which is a function of w only,

$$\kappa(w,z_\mathrm{b})^2 = w\mathrm{Li}_{c-1}(z_\mathrm{b}\kappa(w,z_\mathrm{b})) - w\mathrm{Li}_c(z_\mathrm{b}\kappa(w,z_\mathrm{b}))\,. \tag{3.16}$$

The behavior of $\kappa(w,z)$ in the vicinity of the branch point can be obtained by expanding eq. (3.12) for $z \to z_\mathrm{b}$ and $z_\mathrm{b}\kappa(w,z_\mathrm{b}) < 1$

$$\kappa(w,z) \sim \kappa_\mathrm{b} - \Big(\frac{z_\mathrm{b}-z}{z_\mathrm{b}}\Big)^{1/2} K_\mathrm{b}(1-sz_\mathrm{b}\kappa_\mathrm{b})^2\,, \tag{3.17}$$

where we used the short notation $\kappa_\mathrm{b} = \kappa(w,z_\mathrm{b})$ and defined

$$K_\mathrm{b} = \Big(\frac{2w\mathrm{Li}_{c-1}(z_\mathrm{b}\kappa_\mathrm{b})}{w\mathrm{Li}_{c-2}(z_\mathrm{b}\kappa_\mathrm{b}) - w\mathrm{Li}_{c-1}(z_\mathrm{b}\kappa_\mathrm{b}) - 2\kappa_\mathrm{b}^2}\Big)^{1/2} \frac{\kappa_\mathrm{b}}{(1-sz_\mathrm{b}\kappa_\mathrm{b})^2}\,. \tag{3.18}$$

Due to the exponent $1/2$ in the above equation, the function $\kappa(w,z)$ exhibits a first order branch point at $z = z_\mathrm{b}$ and the grand canonical partition function, eq. (3.9), scales as

$$\mathcal{Z}(z,s) \sim \frac{\kappa_\mathrm{b}}{1-sz_\mathrm{b}\kappa_\mathrm{b}} - \Big(\frac{z_\mathrm{b}-z}{z_\mathrm{b}}\Big)^{1/2} K_\mathrm{b}\,. \tag{3.19}$$

Together with eq. (3.14), we obtain the following scaling for the canonical partition function

$$Z_N(s) \sim z_\mathrm{b}^{-N} N^{-3/2} K_\mathrm{b}/\sqrt{4\pi}\,, \tag{3.20}$$

which leads to a logarithmic N-contribution with universal prefactor $3/2$ to the free energy $\mathcal{G} = -\mathrm{k_B}T\ln Z_N$, in accord with the findings of de Gennes [89]. It will turn out that eq. (3.20) describes the low temperature or folded phase of the system.

Now let us consider the pole singularity z_p of the grand canonical partition function. $z_\mathrm{p}(w,s)$ is a function of w and s and is given by the zero of the denominator of $\mathcal{Z}(z,s)$ in eq. (3.9),

$$sz_\mathrm{p}\kappa(w,z_\mathrm{p}) = 1\,. \tag{3.21}$$

The position of the pole can be evaluated in a closed form expression by plugging eq. (3.21) into eq. (3.12) and solving the resulting quadratic equation for z. One obtains

$$z_\mathrm{p}(w,s) = \frac{2}{s}\left(1+\sqrt{1+4w\mathrm{Li}_c(1/s)}\right)^{-1}\,, \tag{3.22a}$$

$$\kappa(w,z_\mathrm{p}) = \frac{1}{2}\left(1+\sqrt{1+4w\mathrm{Li}_c(1/s)}\right)\,. \tag{3.22b}$$

The behavior of $\kappa(w,z)$ in the vicinity of the pole can be obtained by expanding eq. (3.12) for $z \to z_\mathrm{p}$, $z_\mathrm{p}\kappa(w,z_\mathrm{p}) = 1/s < 1$, and $c > 2$

$$\kappa(w,z) \sim \kappa_\mathrm{p} - \kappa_\mathrm{p}\frac{z_\mathrm{p}-z}{z_\mathrm{p}}\,\frac{w\mathrm{Li}_{c-1}(1/s)}{K_\mathrm{p}(2\kappa_\mathrm{p}-1)}\,, \tag{3.23}$$

where we used the short notation $\kappa_\mathrm{p} = \kappa(w,z_\mathrm{p})$ and introduced

$$K_\mathrm{p} = \frac{2\kappa_\mathrm{p}^2 - \kappa_\mathrm{p} - w\mathrm{Li}_{c-1}(1/s)}{2\kappa_\mathrm{p}-1}\,. \tag{3.24}$$

Therefore, the grand canonical partition function scales as

$$\mathcal{Z}(z,s) \sim \left(\frac{z_\mathrm{p}-z}{z_\mathrm{p}}\right)^{-1} K_\mathrm{p} \tag{3.25}$$

and together with eq. (3.14) we obtain the scaling of the canonical partition function

$$Z_N(s) \sim z_\mathrm{p}^{-N} K_\mathrm{p}\,. \tag{3.26}$$

Later we will see, that eq. (3.26) describes the denatured high temperature phase of the system. In contrast to the branch point phase, eq. (3.20), no logarithmic contribution to the free energy is present.

3.4. Thermodynamics of homopolymeric RNA

The two singularities $z_\mathrm{b}(w)$ and $z_\mathrm{p}(w,s)$ are smooth functions of external variables such as temperature T or force F, which enter via the weight of a base pair w, eq. (3.1), and the weight of a non-nested backbone bond s, eq. (3.6). As the system is described by the singularity, which is closest to the origin, a phase transition associated with a true

singularity in the free energy, eq. (3.15), is possible if these two singularities cross. For that purpose let us shortly review the three constitutive equations eqs. (3.12),(3.16), and (3.21). As observed earlier, the smallest positive root of eq. (3.12) yields the function $\kappa(w,z)$. The simultaneous solution of eqs. (3.12) and (3.16) yields the branch point $z_\mathrm{b}(w)$, whereas the simultaneous solution of eqs. (3.12) and (3.21) yields the pole $z_\mathrm{p}(w,s)$, which can be expressed in a closed form, see eq. (3.22).

3.4.1. Critical point and existence of a phase transition

The critical fugacity z_c and thus the phase transition is defined as the point where the branch point and the pole coincide

$$z_\mathrm{c} = z_\mathrm{b}(w_\mathrm{c}) = z_\mathrm{p}(w_\mathrm{c}, s_\mathrm{c})\,, \tag{3.27}$$

which means that all three eqs. (3.12),(3.16), and (3.21) have to hold simultaneously, see fig. 3.6. Assuming that a pair $(w_\mathrm{c}, s_\mathrm{c})$ exists so that eq. (3.27) is true, this can be evaluated further by plugging eqs. (3.22) into eq. (3.16) and we obtain

$$w_\mathrm{c} = \frac{\mathrm{Li}_{c-1}(1/s_\mathrm{c}) - \mathrm{Li}_c(1/s_\mathrm{c})}{(\mathrm{Li}_{c-1}(1/s_\mathrm{c}) - 2\mathrm{Li}_c(1/s_\mathrm{c}))^2}\,. \tag{3.28}$$

This constitutes a closed form relation between w_c and s_c or, by employing eqs. (3.1) and (3.6), the critical temperature T_c and force F_c. The melting temperature T_m is defined as the critical temperature at zero force.

The order of the branch point exactly at $T = T_\mathrm{m}$ and zero force, $s = 1$, is calculated by expanding eq. (3.12) in powers of $z/z_\mathrm{c} - 1$ and $\kappa(w,z)/\kappa_\mathrm{c} - 1$ while keeping $w = w_\mathrm{c}$ fixed. For vanishing force $\kappa_\mathrm{c} = 1/z_\mathrm{c}$ and we obtain

$$\kappa(w,z) \sim \kappa_\mathrm{c} - \frac{\kappa_\mathrm{c}}{K_{\mathrm{c,T}}}\Big(\frac{z_\mathrm{c} - z}{z_\mathrm{c}}\Big)^{1/(c-1)}\,, \tag{3.29}$$

where we used

$$K_{\mathrm{c,T}} = \Big(\frac{\zeta_{c-1}}{\Gamma(1-c)}\Big)^{-1/(c-1)}\,. \tag{3.30}$$

Thus, the asymptotic behavior of the generating function at $T = T_\mathrm{m}$ and zero force is

$$\mathcal{Z}(z,s) \sim K_{\mathrm{c,T}}\Big(\frac{z_\mathrm{c} - z}{z_\mathrm{c}}\Big)^{-1/(c-1)} \tag{3.31}$$

and we obtain the modified scaling of the canonical partition function right at the melting point temperature

$$Z_N(s=1) \sim z_\mathrm{c}^{-N} N^{(2-c)/(c-1)} K_{\mathrm{c,T}}/\Gamma(1/(c-1))\,. \tag{3.32}$$

We see that the loop statistics are crucial and enter via the loop exponent c, which gives rise to non-universal critical behavior.

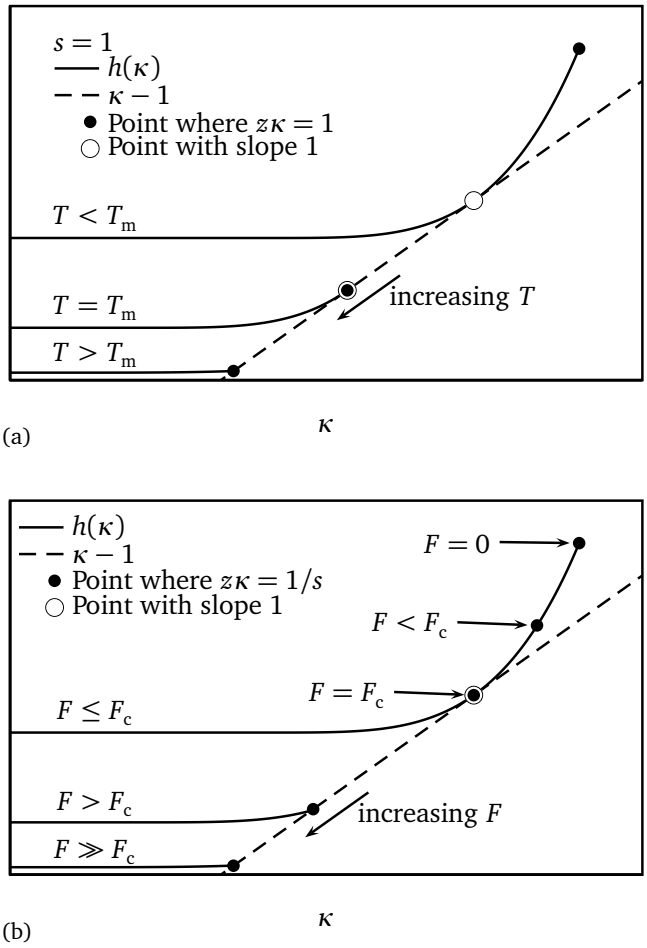


Figure 3.6.: Illustration of the graphical solution of eqs. (3.12),(3.16), and (3.21) at the phase transition. The solid line sketches $h(\kappa)$, eq. (3.13), the dashed line the function $\kappa - 1$. The open dots denote points where the conditions for the branch point are met: Curves have common points, eq. (3.12), and curves are tangent to each in these points, eq. (3.16). The black dots denote points where $sz\kappa = 1$, eq. (3.21). If the black dot lies on the dashed curve the conditions for a pole are met, eqs. (3.12) and (3.21). For a given temperature T and force F the fugacity z is increased from $z = 0$ to the value where either eqs. (3.12) and (3.16) hold (open dot on dashed line, folded phase) or eqs. (3.12) and (3.21) hold (black dot on dashed line, unfolded phase). (a) Illustration of the thermal phase transition at zero force, $s = 1$. For low temperatures the branch point is dominant. Upon increasing the temperature, i. e. decreasing w, the branch point and the pole approach each other until they eventually coincide at the melting temperature $T = T_m$ and cause a phase transition. For $T > T_m$ there is no real branch point anymore and the pole is dominant. (b) Illustration of the force induced phase transition at a temperature $T < T_m$. For small forces the branch point is dominant. Upon increasing the force, i. e. increasing s, the point where $z\kappa = 1/s$ (black dot) moves towards the branch point. As the branch point is independent of the force, see eqs. (3.12) and (3.16), no observable depends on F as long as $F < F_c$. At $F = F_c$ the branch point and the pole coincide and a phase transition occurs. For $F > F_c$ the pole is dominant and the force dependence of the observables sets in.

For finite force, $s > 1$, the branch point is first order and the scaling of the generating function at the critical point reads

$$\mathcal{Z}(z,s) \sim \Big(\frac{z_\mathrm{c} - z}{z_\mathrm{c}}\Big)^{-1/2} \frac{K_\mathrm{c,F}}{s}\,, \tag{3.33}$$

with

$$K_\mathrm{c,F} = \Big(\frac{w_\mathrm{c}\mathrm{Li}_{c-2}(1/s) - w_\mathrm{c}\mathrm{Li}_{c-1}(1/s) - 2\kappa_\mathrm{c}^2}{2w_\mathrm{c}\mathrm{Li}_{c-1}(1/s)}\Big)^{1/2}\,, \tag{3.34}$$

leading to the canonical partition function

$$Z_N(s) \sim z_\mathrm{c}^{-N} N^{-1/2} K_\mathrm{c,F}/(s\sqrt{\pi})\,, \tag{3.35}$$

with a scaling independent of the loop exponent c. Note that eq. (3.33) scales as $(z_\mathrm{c} - z)^{-1/2}$ in contrast to eq. (3.19), which leads to the different scaling of $Z_N(s)$ in eq. (3.35) when compared to eq. (3.20). In the rest of this section we compare thermal and force induced phase transitions and in particular determine the parameter range in which a phase transition is possible.

Thermal phase transition

First, we consider the thermal phase transition without external force, i. e. for $s = 1$. In this case the polylogarithm reduces to the Riemann zeta function, $\mathrm{Li}_c(1) = \zeta_c$. Since $s = 1$, we find that $z_\mathrm{b} \leq z_\mathrm{p}$ as long as $\kappa(w,z)$ has a real, positive branch point. This is due to the fact that $z_\mathrm{b}\kappa(w,z_\mathrm{b}) \leq 1$ (see fig. 3.5), $z_\mathrm{p}\kappa(w,z_\mathrm{p}) = 1$, and that $z\kappa(w,z)$ is a monotonically increasing function of z, see eq. (3.10).

No thermal phase transition for $\mathbf{c \leq 2}$ For $c \leq 2$ the function $\kappa(w,z)$ always features a branch point since $h'(\kappa) \to \infty$ for $\kappa \to 1/z$. This ensures that for every w a z_b is found, where $h(\kappa, z_\mathrm{b})$ is tangent to $\kappa - 1$, see figs. 3.5a and 3.5b. As the branch point is always dominant, we find the universal scaling

$$\mathcal{G}/(\mathrm{k_B}T) = N\ln z_\mathrm{b} + 3/2\ln N \tag{3.36}$$

for all temperatures and no phase transition is possible. The RNA chain is always in the folded phase.

No thermal phase transition for $\mathbf{c \geq c^* \approx 2.479}$ For $c > 2$ the function $h(\kappa,z)$ and its derivative are finite for $z\kappa = 1$. A sufficient condition for a branch point to exist is that the slope of h is greater than 1 for $z\kappa = 1$, see filled circles in fig. 3.5c, and hence

$$h'(\kappa(w,z_\mathrm{p}),z_\mathrm{p}) = \frac{2w(\zeta_{c-1} - 2\zeta_c)}{1 + \sqrt{1 + 4w\zeta_c}} \overset{!}{>} 1\,. \tag{3.37}$$

This can be achieved always for large enough w as long as the numerator is positive. On the other hand, for $c \geq c^* \approx 2.479$, where c^* is the root of

$$\zeta_{c^*-1} - 2\zeta_{c^*} = 0\,, \tag{3.38}$$

no branch point exists since the numerator in eq. (3.37) is negative. That means that for $c > c^*$ the pole $z_\mathrm{p}(w,s)$ is always the dominant singularity of $\mathcal{Z}(z,s)$ and the molecule is always in the unfolded state.

Thermal phase transition for $2 < \mathbf{c} < \mathbf{c}^*$ at $\mathbf{w} = \mathbf{w}_\mathrm{c}$ Only for $2 < c < c^*$ a thermal phase transition is possible. For $w > w_\mathrm{c}$, see eq. (3.28), the molecule is in the folded phase governed by the branch point singularity z_b, which is determined by eqs. (3.12) and (3.16). Decreasing w, i. e. increasing the temperature, causes the branch point and the pole to approach each other. At the critical point w_c, eq. (3.28), both singularities coincide and a phase transition occurs. For higher temperatures, $w < w_\mathrm{c}$, the RNA is unfolded and described by the pole z_p, eq. (3.22). See fig. 3.6a for an illustration. It will turn out that the temperature induced phase transition at zero force is very weak and that, in fact, the order of the phase transition is n, where n is the integer with $(c-2)^{-1} - 1 < n < (c-2)^{-1}$.

Force induced phase transition

For the force induced phase transition the situation is slightly different as the position of the pole z_p depends on the force, which enters via the weight of a non-nested backbone bond s, eq. (3.22). In contrast, the branch point z_b does not depend on s and hence is constant, eqs. (3.12) and (3.16). Therefore, the branch point z_b and the critical point z_c coincide and z_b can be determined exactly by the relation $z_\mathrm{b} = z_\mathrm{c} = z_\mathrm{p}(w_\mathrm{c}, s_\mathrm{c}) = \text{const}$.

No force induced phase transition if $\mathbf{w} < \mathbf{w}_\mathrm{c}(\mathbf{s} = \mathbf{1})$ or $\mathbf{c} \geq \mathbf{c}^*$ If the molecule is already in the unfolded phase, which can be due to high temperature, $w < w_\mathrm{c}$, or due to the non-existence of a branch point, $c \geq c^*$, a force induced phase transition is not possible. In these cases the pole always dominates the system, regardless of the value of the applied force.

Force induced phase transition if $\mathbf{w} > \mathbf{w}_\mathrm{c}(\mathbf{s} = \mathbf{1})$ and $\mathbf{c} < \mathbf{c}^*$ A system below the melting temperature, $w > w_\mathrm{c}(s = 1)$, is in the folded phase at zero force, $s = 1$. For small forces, i. e. $s < s_\mathrm{c}$, the system is described by the branch point singularity z_b, which is independent of s and hence does not depend on the force, eqs. (3.12) and (3.16) and fig. 3.6b. However, as the pole $z_\mathrm{p}(w,s)$ is a monotonically decreasing function of s, the branch point and the pole will eventually coincide at $s = s_\mathrm{c}$ and a phase transition occurs. The critical force fugacity s_c is defined as the root of eq. (3.28) for fixed w. For $s > s_\mathrm{c}$ the pole is the dominant singularity and governs the system.

Note that in contrast to the thermal phase transition, a force induced phase transition is possible even for $c \leq 2$.

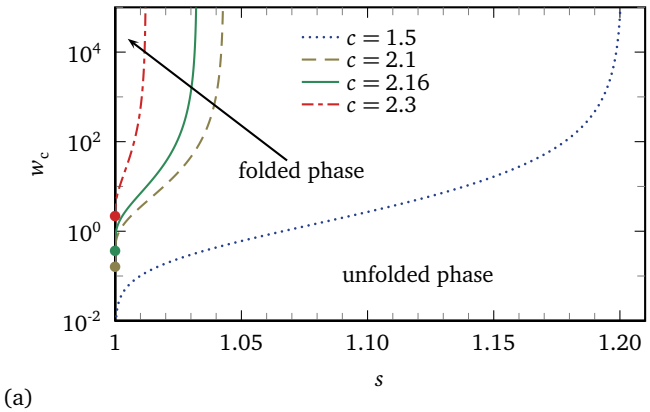
3.4.2. Global phase diagrams

Eq. (3.28) determines the phase diagram. In fig. 3.7a we show the phase transition between the folded and unfolded states of a homopolymeric RNA in the w-s plane for a few different values of the loop exponent c that correspond to an ideal polymer, $c = 3/2$, and values between $c = 2.1$ and $c = 2.3$ as they are argued to be relevant for terminal and internal loops of varying topology including the effects of self-avoidance. Below the transition lines in fig. 3.7a, the chain is in the unfolded (extended) state, above the line in the folded (compact) state. With growing loop exponent, the extent of the folded phase shrinks. In fact, $w_c(s, c)$ diverges for $c \to c^*$, where $c^* \approx 2.479$, cf. eq. (3.38). Thus, for $c \geq c^*$ only the unfolded phase is present, as observed earlier in the restricted case of zero force, $s = 1$ [1]. On the other hand, the critical line for $c \to 2$ and $s = 1$ goes down to zero, $w_c(s = 1) \to 0$ for $c \to 2$, which indicates that if no external force is applied to the molecule, there is only the folded phase for $c \leq 2$ and hence no thermal phase transition is possible. However, applying a sufficient force can drive the molecule into the unfolded state, as derived earlier. Concluding, only for $2 < c < c^*$ a thermal phase transition, denoted by the filled circles in fig. 3.7a, is possible. A force induced phase transition is possible whenever $c < c^*$ and $w > w_c$. For small force and $c > 2$, eq. (3.28) can be expanded around $s = 1$ and yields the universal asymptotic locus of the phase transition

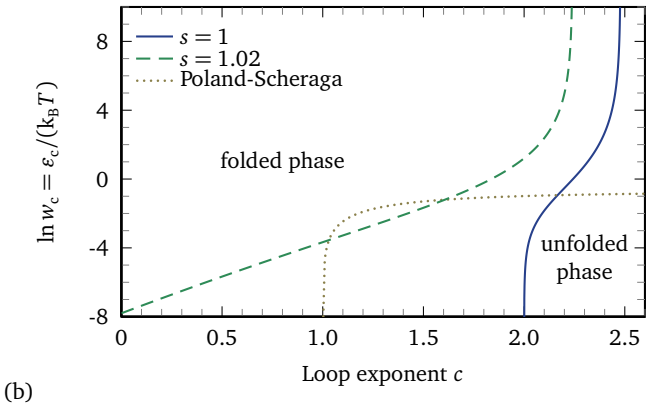
$$w_c(s) \sim w_c(s = 1) + (1 - s)^{c-2} \frac{\Gamma(2 - c)}{(\zeta_{c-1} - 2\zeta_c)^2} \,. \tag{3.39}$$

In fig. 3.7b we show the critical line $w_c(s, c)$ for two different values of s as a function of the loop exponent c. In the absence of an external pulling force, i. e. for $s = 1$ (solid line), the transition line only occurs in the limited range $2 < c < c^* \approx 2.479$. It is seen that for loop exponents around the relevant value of $c \approx 2.1$, the critical base pairing weight is quite small and of the order of $w_c \approx 0.1$. A base pairing weight smaller than unity corresponds to a repulsive base pairing free energy that is unfavorable. This at first sight paradoxical result, which means that the folded phase forms even when the extensive part of the base pairing free energy is repulsive, reflects the fact that the folded state contains a lot of topological entropy because of the different secondary structures. The dotted line shows the melting transition for a duplex DNA chain according to the classical two-state Poland-Scheraga theory, which only considers inter-strand but no intra-strand base pairing, given by the equation [3, 28], see appendix B.1,

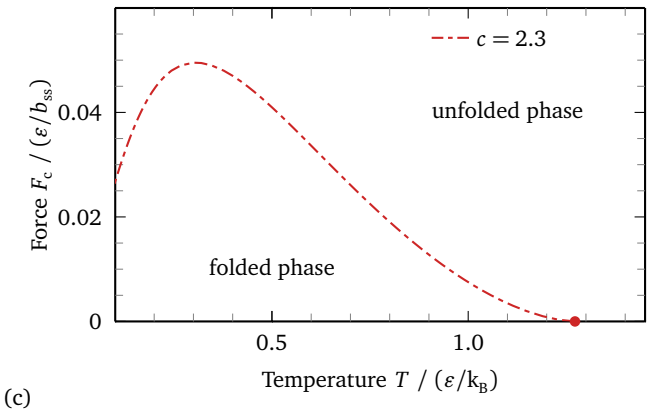
$$w_c^{\mathrm{PS}} = 1/(1 + \zeta_{\hat{c}}). \tag{3.40}$$



(a)



(b)



(c)

Figure 3.7.: (a) Phase diagram of homopolymeric RNA in the w-s plane for different values of the loop exponent $c = 1.5, 2.1, 2.16, 2.3$ featuring an unfolded phase (bottom right) and a folded phase (top left). For $c = c^* \approx 2.479$, the phase boundary approaches $s = 1$ and the melting point $w_c(s = 1)$ diverges; therefore only the unfolded phase exists for $c \geq c^*$. For $c \leq 2$, there is no melting transition at zero force as $w_c(s = 1) = 0$. Thus if no force is applied, the system is always in the folded phase regardless of the temperature. The molecule can be denatured, though, by applying an external force even for $c \leq 2$, as can be seen from the phase boundary for $c = 1.5$. The dots denote the thermal denaturation transition point $w_c(s = 1)$ in the absence of an external force for $c = 2.1, 2.16, 2.3$. (b) Phase diagram in the w-c plane. With zero force, $s = 1$ (solid line), the weight w_c drops to zero for $c \to 2$ and diverges as $c \to c^* \approx 2.479$. For finite force, $s > 1$ (dotted line), a phase transition is possible, even for $c \leq 2$. The dotted line denotes the two-state Poland-Scheraga prediction for the melting of a nucleotide duplex in the absence of intra-strand base pairing, eq. (3.40). (c) Phase diagram in the F-T plane for $c = 2.3$. Below the phase boundary the folded state is present, above the unfolded phase. Re-entrance at constant force is observed, as reported by Müller [116].

The crossing of the Poland-Scheraga duplex melting line eq. (3.40) and the folding transition line eq. (3.28) occurs at the threshold loop exponent

$$c^{**} = 2.17\,. \tag{3.41}$$

This means that for typical loop exponents for RNA loops around $c \approx 2.1$, the melting of a nucleotide duplex that can form inter-strand as well as intra-strand base pairs occurs at a substantially higher temperature since secondary structures arise in the two single strands of a denatured region of the duplex. The consequences for the theoretical description of systems, where single stranded nucleic acids occur, including DNA transcription, denaturation bubbles in dsDNA, untwisting of nucleic acids [35], and translocation [97] will be briefly discussed in section 3.5. For $c > c^{**}$ the Poland-Scheraga melting occurs at a higher temperature than our unfolding transition, which at first sight seems surprising. This finding can be rationalized by the fact that in our present model, even the unfolded state is characterized by pronounced base pairing, whereas the unfolded state of the Poland-Scheraga model exhibits no base pairing at all [3].

The phase diagram, eq. (3.28), can also be displayed in the F-T plane by virtue of eqs. (3.1) and (3.6) and is shown in fig. 3.7c. Here, re-entrance at constant force becomes visible, in line with previous predictions by Müller [116]. Expanding eq. (3.6) we obtain $s(F) \sim 1 + (b_{\mathrm{ss}}F/(\mathrm{k_B}T))^2/6$, for $F \to 0$. Eq. (3.39) yields the scaling of the critical force close to the melting temperature as

$$F_{\mathrm{c}} \propto (T_{\mathrm{m}} - T)^{1/(2c-4)}\,, \tag{3.42}$$

which depends on the loop exponent c and deviates from the predictions by Müller [116], who found a universal exponent $1/2$. We attribute this deviation to the use of slightly different models as Müller used a non-zero cooperativity parameter.

3.4.3. Thermodynamic quantities and critical exponents

We now consider the thermodynamic and critical behavior of various quantities. An arbitrary extensive quantity Y with the conjugated field f is obtained from the grand potential $\Phi = -\mathrm{k_B}T \ln \mathcal{Z}$ via differentiation with respect to f and the chemical potential μ held constant

$$Y = \left.\frac{\partial \Phi}{\partial f}\right|_{\mu}\,. \tag{3.43}$$

To evaluate the behavior of Y in the thermodynamic limit, $N \to \infty$, one sets $\mu \to \mu_{\mathrm{d}}$, where μ_{d} is defined as the chemical potential, at which $N(\mu) = -\partial\Phi/\partial\mu$ diverges, i. e. $N(\mu) \to \infty$ for $\mu \to \mu_{\mathrm{d}}$. Another route to obtain Y is to conduct the calculation in the canonical ensemble, i. e. $N = \mathrm{const}$, and to use the dominating singularity [3], where

$$Y = \frac{\partial \mathcal{G}}{\partial f} = \mathrm{k_B}TN\frac{\partial \ln z_{\mathrm{d}}}{\partial f}\,, \tag{3.44}$$

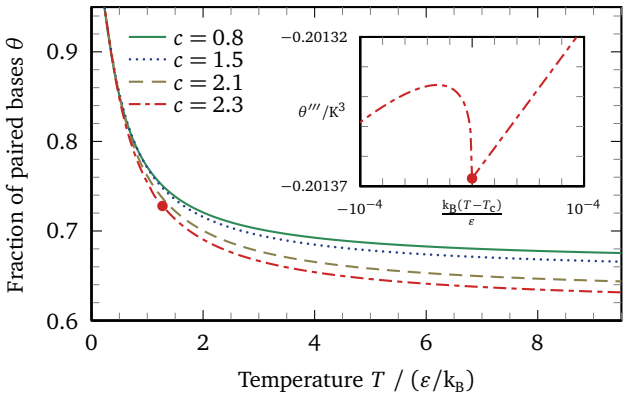


Figure 3.8.: Fraction of bound bases as a function of temperature for $w = \exp(\varepsilon/(\mathrm{k_B}T))$ and various $c = 0.8, 1.5, 2.1, 2.3$ at zero force, $s = 1$. A genuine phase transition is observed only for $c = 2.3$ (indicated by the dot) since for $c \lesssim 2.195$ the critical weight of a hydrogen bond is $w_\mathrm{c} < 1$, which can only be obtained for $\varepsilon < 0$ amounting to a repulsive interaction. For $c \leq 2$ no thermal phase transition can be observed at all. The inset shows the third derivative $\theta''' = \mathrm{d}^3\theta/\mathrm{d}T^3$ for $c = 2.3$, which reveals the phase transition.

see eq. (3.15). In fact, for $N \to \infty$ eqs. (3.43) and (3.44) are equivalent and μ_d is associated with the dominating singularity of $\mathcal{Z}(z,s)$, namely $z_\mathrm{d} = \exp(\beta\mu_\mathrm{d})$, which will be shown now. The Gibbs free energy $\mathcal{G}$ and the grand potential Φ are interrelated via a Legendre transform

$$\mathcal{G}(N) = \Phi + \mu(N)N \,. \tag{3.45}$$

Therefore,

$$\begin{aligned} Y &= \mathrm{k_B}TN\frac{\partial \ln z_\mathrm{d}}{\partial f} = \left.\frac{\partial \mathcal{G}}{\partial f}\right|_N = \left.\frac{\partial(\mathcal{G}-\Phi)}{\partial f}\right|_{N,\Phi} = N\left.\frac{\partial \mu}{\partial f}\right|_\Phi \\ &= -\left.\frac{\partial \Phi}{\partial \mu}\right|_f \left.\frac{\partial \mu}{\partial f}\right|_\Phi = \left.\frac{\partial \Phi}{\partial f}\right|_\mu , \end{aligned} \tag{3.46}$$

where the two final expressions are evaluated at $\mu = \mu_\mathrm{d}$. While performing derivatives of the dominant singularity and of the function $\kappa(w,z)$ is straightforward for z_p and κ_p, see eq. (3.22), one has to employ implicit differentiation of eq. (3.16) to obtain the derivative of z_b and κ_b, see appendix B.2. For the latter case it turns out that eq. (3.43) is more convenient to work with.

Fraction of bound bases

The fraction of paired bases is

$$\theta = \frac{1}{N}\frac{\partial \ln Z_N}{\partial \ln w} \,. \tag{3.47}$$

We obtain

$$\theta = \frac{2\mathrm{Li}_c(z_\mathrm{b}\kappa(w,z_\mathrm{b}))}{\mathrm{Li}_{c-1}(z_\mathrm{b}\kappa(w,z_\mathrm{b}))} \tag{3.48}$$

in the folded phase ($T < T_\mathrm{c}$, $F < F_\mathrm{c}$) and

$$\theta = 1 - \frac{1}{\sqrt{1+4w\mathrm{Li}_c(1/s)}} \tag{3.49}$$

in the unfolded phase ($T > T_\mathrm{c}$, $F > F_\mathrm{c}$). In fig. 3.8 the temperature dependence and in fig. 3.9 the force dependence of θ is shown. The singularity at the critical point of the thermal phase transition for zero force, $s = 1$, is very weak and becomes visible in the n^{th} derivative, with n being the integer with $(c-2)^{-1} - 1 < n < (c-2)^{-1}$, see appendix B.3. The n^{th} derivative exhibits a cusp, see inset of fig. 3.9,

$$\frac{\mathrm{d}^n\theta(T)}{\mathrm{d}T^n} \propto |T - T_\mathrm{m}|^{\lambda_T} + \mathrm{const}\,, \tag{3.50}$$

which is characterized by the critical exponent $\lambda_T = (c-2)^{-1} - n$ for $T < T_\mathrm{m}$ and $\lambda_T = 1$ for $T > T_\mathrm{m}$, see appendix B.3. The force induced phase transition is continuous, too, yet it exhibits a kink in $\theta(F)$

$$\theta(F) \propto |F - F_\mathrm{c}|^{\lambda_F} + \mathrm{const}\,, \tag{3.51}$$

which is characterized by the exponents $\lambda_F = 0$ for $F < F_\mathrm{c}$ and $\lambda_F = 1$ for $F > F_\mathrm{c}$. We note that for $T \to \infty$, i. e. $w \to 1$, a finite fraction of bases are still paired. This indicates that models as ours, which solely focus on secondary structures, tend to underestimate the entropy of the completely unfolded state, where no base pairs remain. The situation is different for the force induced phase transition where $\theta \to 0$ for $F \to \infty$.

Specific heat

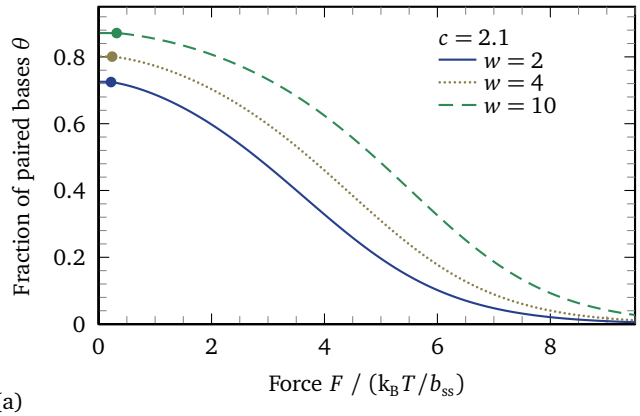
The specific heat is defined as

$$C = \frac{\mathrm{k_B}T}{N}\frac{\partial^2 T \ln Z_N}{\partial T^2}\,. \tag{3.52}$$

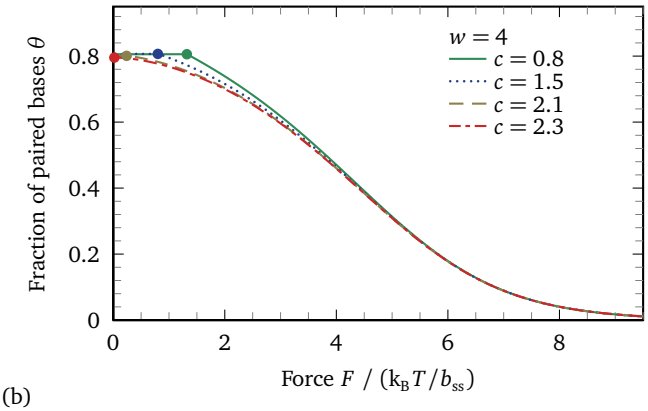
One observes that the specific heat in fig. 3.10 exhibits only a very weak dependence on the loop exponent, which stands in marked contrast to the findings for the short explicit sequence of tRNA-phe, see the following chapter 4 [1, 4], where a pronounced dependence of the heat capacity on c is observed. The non-analyticity of θ, eq. (3.50), translates into a divergence of the n^{th} derivative of the specific heat at the melting temperature

$$\frac{\mathrm{d}^n C(T)}{\mathrm{d}T^n} \propto |T - T_\mathrm{m}|^{-\chi} + \mathrm{const}\,, \tag{3.53}$$

with the critical exponent $\chi = n - (3-c)/(c-2)$ for $T < T_\mathrm{m}$ and $\chi = 1$ for $T > T_\mathrm{m}$. This singularity is illustrated in the inset of fig. 3.10 for $c = 2.3$.



(a)



(b)

Figure 3.9.: Fraction of bound bases as a function of force for various w and c. The phase transition is visible as a kink in the curves and indicated by a dot. (a) $c = 2.1$ and varying $w = 2, 4, 10$. (b) $w = 4$ and varying $c = 0.8, 1.5, 2.1, 2.3$.

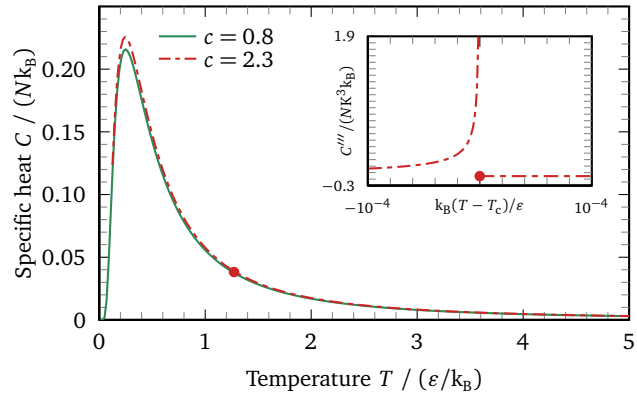


Figure 3.10.: Specific heat as a function of temperature for different loop exponents $c = 0.8, 2.3$. The non-critical behavior of the heat capacity curve depends on the loop exponent only marginally. However, the existence and position of the critical point (indicated by the dot) and the critical behavior depend on c. The inset depicts the third derivative of the specific heat, $C''' = d^3C/dT^3$ for $c = 2.3$ revealing the phase transition.

Fraction of non-nested backbone bonds

The fraction of non-nested backbone bonds is obtained by

$$\tau = \frac{1}{N}\frac{\partial \ln Z_N}{\partial \ln s} \tag{3.54}$$

and is

$$\tau = 0 \tag{3.55}$$

in the folded phase, as z_b does not depend on s, and reads

$$\tau = 1 - \frac{2w\mathrm{Li}_{c-1}(1/s)}{1+4w\mathrm{Li}_c(1/s)+\sqrt{1+4w\mathrm{Li}_c(1/s)}} \tag{3.56}$$

in the unfolded phase. For $T \to \infty$ the fraction of non-nested backbone bonds assumes a finite value smaller than one, which again indicates that the denatured phase in our model features pronounced base pairing. However, for large force $F \to \infty$ leads to $\tau \to 1$. As can be seen nicely in fig. 3.11 both $\tau(T)$ and $\tau(F)$ feature a kink at the critical point.

Force extension curve

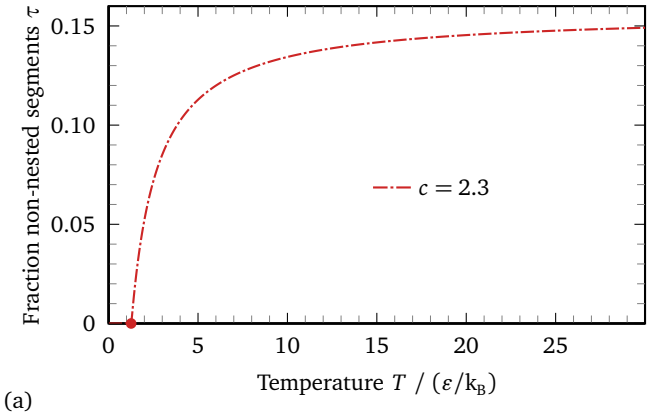
The force extension curve is closely related to the fraction of non-nested backbone bonds τ. The extension per monomer is given by

$$\begin{aligned} x(F) &= \frac{\mathrm{k_B}T}{N}\frac{\partial \ln Z_N}{\partial F} = \frac{\mathrm{k_B}T}{N}\frac{\partial \ln Z_N}{\partial \ln s}\frac{\partial \ln s}{\partial F} \\ &= b_{\mathrm{ss}}\tau\,(\coth(\beta F b_{\mathrm{ss}}) - 1/(\beta F b_{\mathrm{ss}})) = b_{\mathrm{ss}}\tau\,\mathcal{L}(\beta F b_{\mathrm{ss}})\,, \end{aligned} \tag{3.57}$$

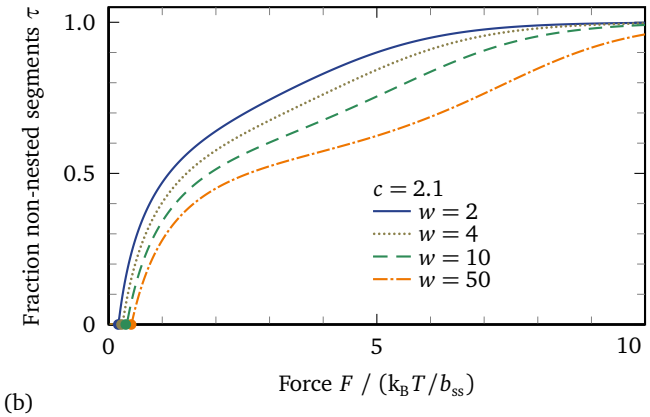
Since the Langevin function $\mathcal{L}$ is smooth, the critical behavior of $x(F)$ is governed by the behavior of the fraction of non-nested bonds τ. As can be seen in fig. 3.12a, the stretching behavior of a freely jointed chain, which is described by the Langevin function, is approached as the base pairing weight decreases, otherwise pronounced deviations are seen in the stretching curves. Also, a finite stretching force to unravel the folded state is needed.

3.5. Implications for DNA melting

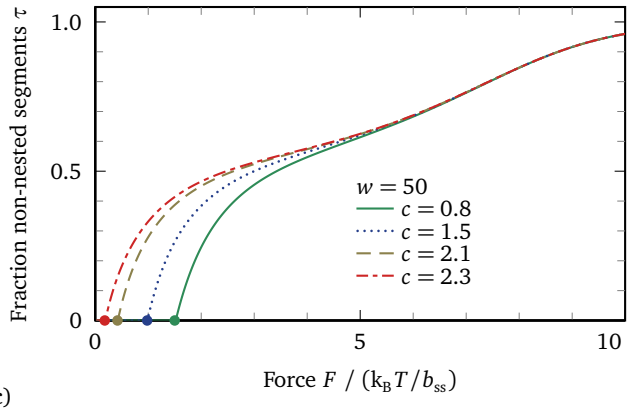
How do the previous results impact on the theoretical description of the denaturation of double stranded nucleic acid systems, particularly DNA melting? When double stranded DNA approaches the denaturation transition, more and more inter-strand base pairs break up and inter-strand loops proliferate. In the traditional theories based on the Poland-Scheraga model [27, 28] as our three-state model in chapter 2, the possibility of intra-strand base pairing is not considered. These models are thus accurate



(a)

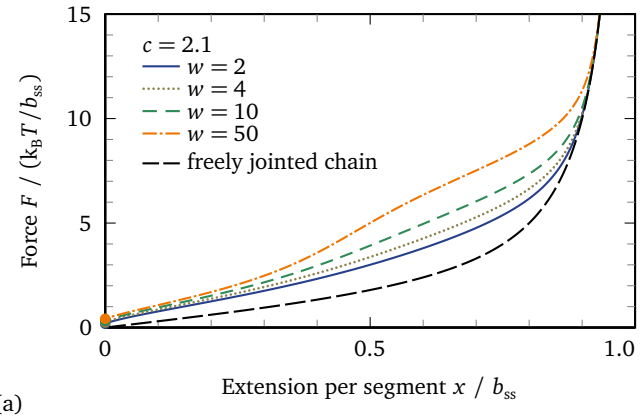


(b)

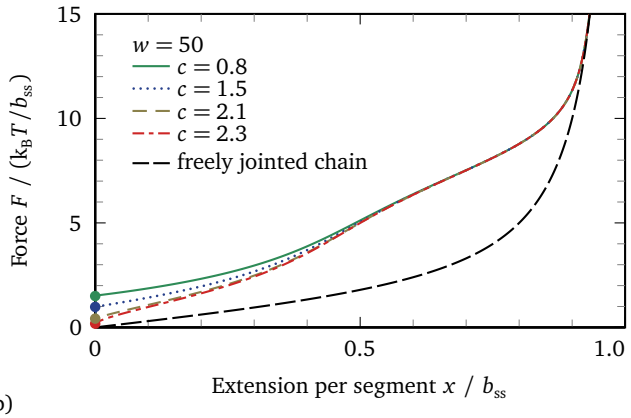


(c)

Figure 3.11.: (a) Fraction of non-nested backbone bonds as a function of temperature for $c = 2.3$. For $c \lesssim 2.195$ the critical weight of a hydrogen bond is $w_c < 1$, which can only be obtained for $\varepsilon < 0$ amounting to a repulsive interaction. Thus, for $c \lesssim 2.195$ or $T < T_m$ all segments are parts of loops or helices and hence $\tau = 0$. Dots indicate the melting temperature T_m. (b) and (c) fraction of non-nested backbone bonds as a function of force for various w and c. Again, the phase transition is visible as a kink in the curves and is indicated by a dot. (b) $c = 2.1$ and varying $w = 2, 4, 10, 50$. (c) $w = 50$ and varying $c = 0.8, 1.5, 2.1, 2.3$.



(a)



(b)

Figure 3.12.: Force extension curve as a function of force for various w and c. The phase transition is indicated by dots and occurs at zero extension and a finite threshold force needed to unfold the compact folded structure. (a) $c = 2.1$ and varying $w = \exp(\varepsilon/k_B T) = 2, 4, 10, 50$. (b) $w = 50$ and varying $c = 0.8, 1.5, 2.1, 2.3$. Additionally the force extension curve of a freely jointed chain is plotted, which is the limiting form for $w = 0$ or $s = \infty$.

for duplexes formed between strands with sequences $[AG]_{N/2}$ and $[TC]_{N/2}$, where indeed base pairs (between A and T and between G and C) can only form between the two strands, not within one strand, cf. fig. 3.13b. For the case of duplexes formed by two strands with the sequence $[AT]_{N/2}$ or $[GC]_{N/2}$, both intra- and inter-strand base pairs can form and have identical statistical weights, , cf. fig. 3.13a. In this case, the model presented in this chapter describes duplex melting. In fig. 3.7b we already saw that intra-strand base pairing leads to increased transition temperatures compared to the two-state Poland-Scheraga model.

The above sequence examples are prototypes for the two extreme cases for zero intra-strand base pairing weight and for equal intra- and inter-strand base pairing weight. Intermediate values for the relative interaction strengths between bases on different strands and on the same strand can be achieved experimentally, for example, by $[ATT]_{N/3}$ and its complementary sequence $[AAT]_{N/3}$, in which case 2/3 of the intra-strand base pairs can be of the Watson-Crick type and which would effectively lead to a lower weight of intra-strand base pairs. In naturally occurring DNA a similar situation might be present above the glass transition where, for certain sequences, self-hybridization in a single strand is possible to a certain extent [113]. The weights

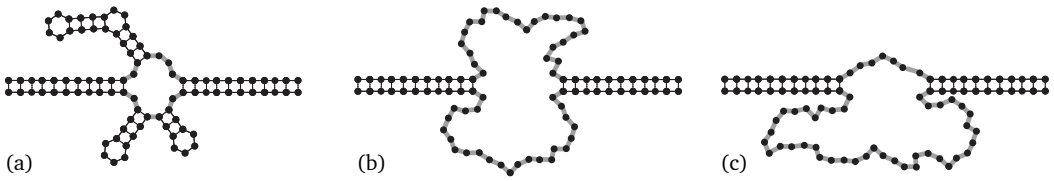


Figure 3.13.: Illustration of a double stranded DNA molecule near the melting transition. (a) If intra-strand base pairing is possible, in inter-strand loops the vast majority of bases will form intra-strand secondary structure elements. (b) If the sequence does not favor intra-strand base pairs inter-strand loops without secondary structure will form. (c) Illustration of an asymmetric loop. In all graphs, the thick gray lines denote the non-nested backbone bonds counted by the variables M, M' in eq. (3.58).

of inter-strand and intra-strand base pairs can also be changed by applying an external force or torque on the duplex, for instance in the setup by Léger et al. [35]. We expect the weight of inter-strand base pairs to decrease when the duplex is untwisted. This might lead to denatured regions in the duplex, where secondary structure can form if the sequence allows. We speculate that subsequent pulling first leads to the denaturation of the intra-strand secondary structures, whose signature would be a threshold force around 1 pN - 10 pN [4, 93], see fig. 3.12, followed by the over-stretching transition of DNA [3]. According to our previous arguments a marked dependence on the sequence should be observed. The full calculation of this system is cumbersome and not shown here.

To gain a more direct understanding of the effects of intra-strand base pairing on duplex melting, let us now consider the statistical weight of a single inter-strand loop, where secondary structure formation in the two single strands is possible. Explicitly, we assume the intra-strand base pairing weight w to be above the critical value w_c so that we have essentially folded single strands. For instance, $w > w_c$ regardless of the temperature for $c = 2.1$ and attractive base pairing interaction $\varepsilon = k_B T \ln w > 0$, see eq. (3.28). The statistical weight of a denatured region of the double strand, which is N base pairs long, where no inter-strand pairs are present, and where intra-strand secondary structures may form, as depicted in fig. 3.13a, is given by

$$Z_N^{\mathrm{DNA}} = \sum_{M,M'=0}^{N} s^{M+M'+4} \frac{\tilde{Q}_N^M \tilde{Q}_N^{M'}}{(M+M'+4)^{\hat{c}}} \,. \tag{3.58}$$

$\hat{c} \approx 2.1$ is the loop exponent describing inter-strand loops in DNA, $M + M' + 4$ counts the number of non-nested back-bones that contribute to the loop entropy, and $\tilde{Q}_N^M$ is given by eq. (3.4). The denominator in eq. (3.58) amounts to an effective interaction between the two strands as the expectation value of M depends on M' and vice versa.

The asymptotic behavior of Z_N^{DNA} can be estimated by establishing two inequalities. The first is

$$Z_N^{\mathrm{DNA}} \leq s^4 \sum_{M=0}^{N} s^M \tilde{Q}_N^M \sum_{M'=0}^{N} s^{M'} \tilde{Q}_N^{M'} \propto z_{\mathrm{b}}^{-2N} N^{-3} \tag{3.59}$$

where the scaling is due to eqs. (3.5) and (3.20) as the branch point is dominant if $w > w_{\mathrm{c}}$. The upper scaling boundary is obtained by factorizing eq. (3.58) and therefore removing the effective interaction, but retaining the loop entropy for each individual strand

$$Z_N^{\mathrm{DNA}} \geq \sum_{M=0}^{N} \frac{s^{M+2} \tilde{Q}_N^M}{(M+2)^{\hat{c}}} \sum_{M'=0}^{N} \frac{s^{M'+2} \tilde{Q}_N^{M'}}{(M'+2)^{\hat{c}}} \propto z_{\mathrm{b}}^{-2N} N^{-3} \,. \tag{3.60}$$

We thus conclude that

$$Z_N^{\mathrm{DNA}} \propto z_{\mathrm{b}}^{-2N} N^{-3} \tag{3.61}$$

to leading order and the bare inter-strand loop exponent $\hat{c}$ is replaced by an effective renormalized exponent with the value $\hat{c}' = 3$. The last equality in eq. (3.60) follows from the dominant singularity analysis of the generating function

$$\mathcal{Z}^{\sqrt{\mathrm{DNA}}}(z,s) = s^2 z^2 \sum_{N=0}^{\infty} \sum_{M=0}^{N} s^M z^N \frac{\tilde{Q}_N^M}{(M+2)^{\hat{c}}} = \frac{1}{\kappa(w,z)} \mathrm{Li}_{\hat{c}}(sz\kappa(w,z)) - sz \,. \tag{3.62}$$

$\mathcal{Z}^{\sqrt{\mathrm{DNA}}}(z,s)$ features the same singularities as $\mathcal{Z}(z,s)$, see eq. (3.9), where the branch point z_{b} of $\kappa(w,z)$ is dominant.

To estimate the size of the DNA loop, which is the number of non-nested bonds of the two secondary structures, at zero force, $s = 1$, we calculate

$$\begin{aligned} \frac{\langle M + M' \rangle}{N} &= \frac{1}{N Z_N^{\mathrm{DNA}}} \sum_{M,M'=0}^{N} (M+M') \frac{\tilde{Q}_N^M \tilde{Q}_N^{M'}}{(M+M'+4)^{\hat{c}}} \\ &\leq \frac{1}{N Z_N^{\mathrm{DNA}}} \sum_{M,M'=0}^{N} (M+M') \tilde{Q}_N^M \tilde{Q}_N^{M'} = 2\tau \frac{\left(Z_N(s=1)\right)^2}{Z_N^{\mathrm{DNA}}} = 0 \,, \end{aligned} \tag{3.63}$$

as each of the secondary structures is assumed to be in the folded phase and $\tau = 0$ for $N \to \infty$ and $w > w_{\mathrm{c}}$, see eq. (3.55). Eq. (3.63) shows that in a duplex loop, the vast majority of bases form intra-strand secondary structure elements. As $Z_N^{\mathrm{DNA}} \propto N^{-3}$, see eq. (3.61), the effective loop exponent of a duplex loop is fixed at the universal value $\hat{c} = 3$, and the thermal melting transition of DNA would be rendered first order according to the classic result of Poland and Scheraga [27, 28]. There are $2N$ ways of constructing an asymmetric inter-strand loop [60], where the number of bases in the lower and the upper part of a loop is not required to be identical, see fig. 3.13c for an

illustration. This additional factor changes the scaling to $Z_N^{\mathrm{DNA}} \propto N^{-2}$ and thus changes the effective loop exponent of a duplex loop to $\hat{c} = 2$, right at the threshold between continuous and discontinuous transitions. We add that these results are obtained with the assumption that the inter- and intra-strand base pairing weights are such, that the perturbative treatment of the duplex within the Poland-Scheraga model with loops that are decorated with secondary structures on the single strands, is valid.

As a main result, the effective inter-strand loop exponent is renormalized and takes on universal values, depending on whether one allows for asymmetric loops or not. The formation of intra-strand secondary structure influences the melting temperature and has implications for the determination base pairing free energy parameters [84, 112] and other biotechnological applications where DNA melting and hybridization is involved. Thus, intra-strand interaction might be important to include in software packages predicting the stability of nucleic acids based on a Poland-Scheraga scheme. Algorithms for cofolding of multiple nucleic acids already account for this [117–119].

3.6. Conclusions

The partition function of RNA secondary structures has been evaluated including arbitrary pairing topologies in the absence of pseudoknots, including the configurational entropy of loops in the form of the loop length dependent term $\mathcal{G}_{\mathrm{l}}^{\mathrm{conf}} = -\mathrm{k_B}T \ln m^{-c}$. Exact expressions for the fraction of paired bases, the heat capacity, and the force extension curves are derived in the presence of an external pulling force. The observed thermal phase transition is very weak and of higher order, the force induced transition is found to be second order. The critical behavior and the critical exponents are found to depend on the loop exponent c. A temperature induced melting transition is only possible for $2 < c < c^* \approx 2.479$. Our results have consequences on the denaturation of double stranded DNA molecules, in particular when intra-strand base pairs as well as inter-strand base pairs can form. In this case, the native double strand is in competition with intra-strand base pairing effecting the secondary structures discussed in this chapter. Future directions will include loop exponents, that depend on the number of helices emerging from a given loop, treatment of pseudoknots, and cofolding nucleic acids to study the influence of intra-strand interactions during double strand denaturation.

CHAPTER 4

SEQUENCE AND SALT EFFECTS ON RNA FOLDING AND STRETCHING

4.1. Introduction

4.1.1. Motivation

So far, only sequence independent properties of nucleic acids have been studied. Yet, it is the very sequence encoded in the four different bases, which induces the intriguing structural and functional properties of nucleic acids. RNA has multiple functions: beyond being an information carrier it has regulatory and catalytic abilities [17]. Comprehending how RNA folds and what influences the folding process are key questions [87]. Thus, the reliable prediction of RNA structure and stability under various conditions is crucial for our understanding of the functioning of RNA and nucleic acid constructs in general [120, 121].

The influence of temperature and solution conditions on RNA folding stays in the interest of experimental groups. Traditionally the thermal melting of RNA was monitored via differential scanning calometry or UV spectroscopy for the bulk ensemble [112, 122–124]. More recently, single molecule pulling and unzipping experiments have been used to unveil the influence of different solution conditions and even determine energy parameters [125–127].

On the theoretical side RNA denaturation is modeled on various levels of coarse graining. By focusing on the secondary structure, namely the base pairs (bp), and omitting tertiary interactions equilibrium folding and unfolding can be modeled very successfully [1, 5, 92–94, 97, 116, 118, 128–130]. In chapters 2 and 3 we have shown that – if a logarithmic contribution to the loop entropy is included – homopolymeric nucleic acids, where sequence effects are neglected, feature a genuine phase transition, which can be induced by temperature or force [1, 5, 116]. However, the specific sequence influences the stretching response of a molecule, which has been shown by Gerland et al. [94, 131]. More detailed insights can be obtained by simulations,

which are numerically quite costly, though, when compared to models focusing only on secondary structure. Coarse grained, Go-like simulations of short RNA hairpins allow to analyze the dynamics of the folding and unfolding process [132, 133]. Ion specific effects are studied by performing molecular dynamics [134] or coarse grained simulations [98, 135, 136]. To our knowledge, the salt dependence of denaturation transitions of RNA has not been studied yet.

While for DNA numerous corrections of the base pairing free energies due to varying salt concentrations exist, see refs. [57, 137] and references therein, analogous results for the salt dependence of RNA energy parameters are sparse [98]. However, molecular biology and biotechnological applications depend on the reliable prediction of RNA stability for different solution conditions.

4.1.2. Outline

In this chapter we extend the formalism of chapter 3 and develop a theory that allows to include all these effects – sequence, salt dependence, logarithmic loop entropy, stretching force – and demonstrate that all are necessary to obtain a complete picture of the thermodynamics of the secondary structure of RNA. Neglecting tertiary interactions, we again use a recursion relation, which allows to correctly account for logarithmic free energy contributions due to the configurational entropy of loops [1]. To include the influence of monovalent salt on RNA stability we model the RNA backbone as charged polymers interacting via a Debye-Hückel potential and give heuristic formulas for the modification of the loop free energy as well as of the base pairing and stacking free energy parameters. Debye-Hückel is a linear theory, yet we allow for the non-linear effects brought about by counterion condensation using Manning's concept [138]. The backbone elasticity of single stranded RNA (ssRNA) is described by the freely jointed chain (FJC) model. Our approach allows for a complete description of the behavior of RNA in the three-dimensional phase space spanned by temperature, salt concentration, and external stretching force. We find that for an improved description of RNA melting curves one needs to include both salt effects and loop entropy. Only the combined usage of these two contributions enables to predict the shift of the melting temperature (due to salt) and the cooperativity (due to logarithmic loop entropy), which is illustrated in the case of tRNA-phe. As an independent check we consider the force induced unfolding of the P5ab RNA hairpin and observe excellent agreement with experimental values with no fitting parameters. The influence of salt is illustrated by melting curves and force extension traces for various salt concentrations. For the P5ab hairpin the phase diagram is determined and slices through the three-dimensional phase space are shown.

4.2. Free energy parameterization of secondary structures

Like in chapter 3 we employ the idea of hierarchical folding. Since the main point is the influence of the sequence, the loop entropy, and the salt concentration on the secondary structure, we again neglect tertiary interactions, see section 3.2.

Given a set of base pairs the secondary structure consists of helices and loops as the basic structural units, cf. fig. 3.1 on page 31. Since pseudoknots are neglected a nucleotide can be attributed to exactly one subunit. The free energy of a certain secondary structure is then given by the sum of the free energy contributions of the individual structural subunits as we will detail now.

4.2.1. Free energy of a loop

We model the free energy of a loop consisting of m backbone bonds, see fig. 3.1, with

$$\mathcal{G}_{\mathrm{l}}(m) = \mathcal{G}_{\mathrm{l}}^{\mathrm{conf}}(m) + \mathcal{G}_{\mathrm{l}}^{\mathrm{salt}}(m) + \mathcal{G}_{\mathrm{l}}^{\mathrm{init}} \,. \tag{4.1}$$

The first term is the loop entropy difference between an unconstrained polymer and a ring-like polymer, which characterized by the loop exponent c [1, 75, 102]

$$\mathcal{G}_{\mathrm{l}}^{\mathrm{conf}}(m) = -\mathrm{k_B}T \ln m^{-c} \,, \tag{4.2}$$

with $\mathrm{k_B}$ the Boltzmann constant and T the absolute temperature. c denotes the loop exponent, which has been discussed in section 3.2 and appendix A.3. We assume a constant loop exponent $c = 2.1$ in this work and only compare with the case of vanishing loop entropy characterized by $c = 0$.

The second term in eq. (4.1) describes the free energy difference between a charged ring of length ml_{ss} and a straight rod of the same length due to electrostatic interactions, with $l_{\mathrm{ss}} = 6.4$ Å the length of one ssRNA backbone bond [136]. The electrostatics are modeled on the Debye-Hückel level [140]

$$\begin{aligned}
\mathcal{G}_{\mathrm{l}}^{\mathrm{salt}}(m) = \mathrm{k_B}T l_{\mathrm{B}}(ml_{\mathrm{ss}})\tau_{\mathrm{ss}}^2 \\
\times \Bigg[\ln(\kappa m l_{\mathrm{ss}}) - \ln(\pi/2) + \gamma - \frac{\kappa m l_{\mathrm{ss}}}{2} {}_1F_2\left(1/2, \begin{pmatrix} 1 \\ 3/2 \end{pmatrix}, \left(\frac{\kappa m l_{\mathrm{ss}}}{2\pi}\right)^2\right) \\
+ \frac{1}{2}\left(\frac{\kappa m l_{\mathrm{ss}}}{\pi}\right)^2 {}_2F_3\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3/2 \\ 3/2 \\ 2 \end{pmatrix}, \left(\frac{\kappa m l_{\mathrm{ss}}}{2\pi}\right)^2\right) \\
+ \frac{1}{\kappa m l_{\mathrm{ss}}}\left(1 - \exp(-\kappa m l_{\mathrm{ss}}) + \kappa m l_{\mathrm{ss}}\Gamma(0, \kappa m l_{\mathrm{ss}})\right) \Bigg] \,,
\end{aligned} \tag{4.3}$$

with $l_{\mathrm{B}} = e_0^2/(\mathrm{k_B}T 4\pi\varepsilon_0\varepsilon_{\mathrm{r}})$ the Bjerrum length, which in water has a value of roughly 7 Å, $\kappa^{-1} = \sqrt{\varepsilon_0\varepsilon_{\mathrm{r}}\mathrm{k_B}T/(2\mathrm{N_A}e_0^2 I)}$ the Debye screening length, ε_0 the vacuum dielectric constant, $\varepsilon_{\mathrm{r}} \approx 80$ the relative dielectric constant of water [141], $I = 1/2(\rho_{\mathrm{a}} z_{\mathrm{a}}^2 + \rho_{\mathrm{c}} z_{\mathrm{c}}^2)$

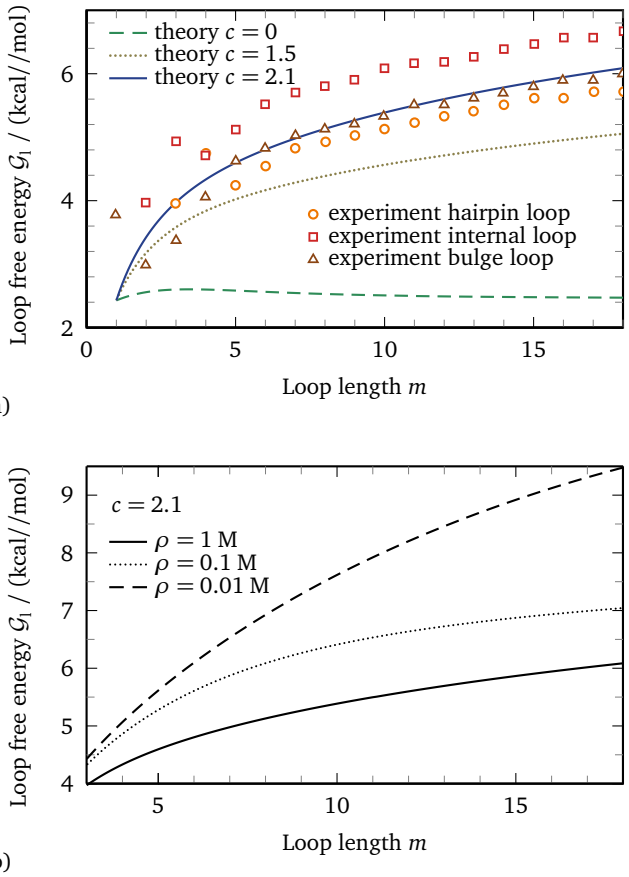


Figure 4.1.: (a) Free energy of a loop as a function of the number of segments m for different loop exponents $c = 0, 1.5, 2.1$ (lines) and for NaCl concentration $\rho = 1$ M. Symbols denote experimental values for various types of loops (hairpin, bulge, internal) for $\rho = 1$ M NaCl [136, 139]. $\mathcal{G}_l^{\text{init}}$ is obtained by fitting $\mathcal{G}_l(m)$, eq. (4.1), to the experimental data for $c = 2.1$ and $\rho = 1$ M. The same salt concentration is used for plotting the curves with $c = 0, 1.5$. (b) Salt dependence of the free energy of loops as a function of the number of segments m for different salt concentrations $\rho = 1$ M, 0.1 M, 0.01 M and loop exponent $c = 2.1$ according to eq. (4.1).

the ionic strength, ρ_a/ρ_c and z_a/z_c the concentration and the valency of the anions/cations, N_A the Avogadro constant, e_0 the elementary charge, $\gamma \approx 0.58$ Euler's constant, $\Gamma(a,x)$ the incomplete gamma function, and ${}_pF_q$ the generalized hypergeometric functions [115]. To estimate the line charge density $\tau_{ss} = -\eta/l_{ss}$ we employ Manning's counterion condensation theory [138] predicting

$$\eta = \min(1, l_{ss}/(l_B z_c)) \,. \tag{4.4}$$

Eq. (4.3) amounts to a ground state approximation of the electrostatic contribution to the free energy of a loop. This is rationalized by the fact that the electrostatic interaction is screened and decays exponentially over the Debye length, which is roughly $\kappa^{-1} = 1$ nm for 100 mM salt solution. However, typical distances between bases in a loop are of the order of the helix diameter $d = 2$ nm or larger. Therefore, we expect electrostatic interactions to be basically independent of the global configuration of a loop, which justifies both the ground state approximation and our additive ansatz, where ion effects and conformational contributions decouple, see eq. (4.1). In the appendix, see eq. (B.5), we give an interpolation formula for eq. (4.3) involving no hypergeometrical functions.

The last term in eq. (4.1) is the loop initiation free energy $\mathcal{G}_{\mathrm{l}}^{\mathrm{init}}$. As we are employing a logarithmic loop entropy, eq. (4.2), we cannot use the standard literature value for $\mathcal{G}_{\mathrm{l}}^{\mathrm{init}}$, which was optimized for a different loop parameterization [112, 122]. Therefore, a new value $\mathcal{G}_{\mathrm{l}}^{\mathrm{init}}$ is obtained by fitting $\mathcal{G}_{\mathrm{l}}(m)$, given by eq. (4.1), to experimental data using $c = 2.1$ in $\mathcal{G}_{\mathrm{l}}^{\mathrm{conf}}(m)$ and the salt concentration $\rho = 1$ M in $\mathcal{G}_{\mathrm{l}}^{\mathrm{salt}}(m)$, see fig. 4.1a. On this plot, we show experimentally determined free energies for terminal, internal and bulge loops as a function of the loop size. The differences between different loop types are quite small. Therefore, we do not distinguish between those loop types in the theory and consequently fit a single parameter $\mathcal{G}_{\mathrm{l}}^{\mathrm{init}}$ to the data, which turns out to be $\mathcal{G}_{\mathrm{l}}^{\mathrm{init}} = 1.9$ kcal/mol for $T = 300$ K, see appendix B.6. In fig. 4.1a the fitted $\mathcal{G}_{\mathrm{l}}(m)$ for the loop exponent $c = 2.1$ is depicted by the solid line; the other lines illustrate the effect of different loop exponents on the loop free energy according to eq. (4.1) using the same value for $\mathcal{G}_{\mathrm{l}}^{\mathrm{init}}$. In fig. 4.1b the salt dependence of $\mathcal{G}_{\mathrm{l}}(m)$ is depicted.

4.2.2. Free energy of a helix

The free energy of a helix

$$\mathcal{G}_{\mathrm{h}} = \mathcal{G}_{\mathrm{h}}^{\mathrm{stack}} + \mathcal{G}_{\mathrm{h}}^{\mathrm{init}} + \mathcal{G}_{\mathrm{h}}^{\mathrm{term}} + \mathcal{G}_{\mathrm{h}}^{\mathrm{salt}} \tag{4.5}$$

depends on the sequence $\{b_i\}$, which consists of the four nucleotides $b_i = \mathrm{C}, \mathrm{G}, \mathrm{A}, \mathrm{U}$. The stacking free energy $\mathcal{G}_{\mathrm{h}}^{\mathrm{stack}}$ is based on experimentally determined parameters incorporating the base pairing free energy as well as the stacking free energy between neighboring base pairs. In the standard notation, $g_{\mathrm{h}}^{\mathrm{stack}}[(b_i, b_j), (b_{i+1}, b_{j-1})]$ is the contribution of the two neighboring, stacked base pairs (b_i, b_j) and (b_{i+1}, b_{j-1}) to $\mathcal{G}_{\mathrm{h}}^{\mathrm{stack}}$. The explicit values for the enthalpic and entropic parts are given in appendix B.6. We use the expanded nearest neighbor model [112, 122] to calculate the base pairing and stacking contributions of a helical section ranging from base pair (i, j) through $(i + h, j - h)$ and obtain

$$\mathcal{G}_{\mathrm{h}}^{\mathrm{stack}} = \sum_{h'=1}^{h} g_{\mathrm{h}}^{\mathrm{stack}}[(b_{i+h'-1}, b_{j-h'+1}), (b_{i+h'}, b_{j-h'})] \,. \tag{4.6}$$

The initiation and termination free energies in eq. (4.5) take into account weaker pairing energies of AU or GU base pairs at the ends of the helix. We use the standard literature values for $\mathcal{G}_{\mathrm{h}}^{\mathrm{init}}$ and $\mathcal{G}_{\mathrm{h}}^{\mathrm{term}}$ [112, 122] and summarize the explicit values in appendix B.6.

Increasing the salt concentration increases the stability of a helix: First, counterions condense on the negatively charged backbone and reduce the electrostatic repulsion and, second, the diffuse counterion cloud surrounding the charged molecule screens the interaction. We model the two strands of a helix as two parallel rods at distance $d = 2$ nm interacting via a Debye-Hückel potential characterized by the screening

length κ^{-1}. The electrostatic interaction energy per nucleotide with the other strand is given by

$$g_{\mathrm{h}}^{\mathrm{DH}}(\rho) = \mathrm{k_B}T\tau_{\mathrm{ds}}^2 l_{\mathrm{ds}} l_{\mathrm{B}} \int_{-\infty}^{\infty} \frac{\exp(-\kappa\sqrt{d^2+z^2})}{\sqrt{d^2+z^2}} dz = 2\mathrm{k_B}T\tau_{\mathrm{ds}}^2 l_{\mathrm{ds}} l_{\mathrm{B}} \mathrm{K}_0(\kappa d)\,. \quad (4.7)$$

$l_{\mathrm{ds}} = 3.4\,\text{Å}$ is the helical rise per base pair of double-stranded RNA (dsRNA) and $\mathrm{K}_0(\kappa d)$ is the zeroth order modified Bessel function of the second kind. Again, we employ Manning's theory [138] to calculate the line charge density $\tau_{\mathrm{ds}} = -\eta/l_{\mathrm{ds}}$, with $\eta = \min(1, l_{\mathrm{ds}}/(l_{\mathrm{B}} z_{\mathrm{c}}))$. The reference state for the salt correction of the pairing free energy is at temperature $T = 300\,\mathrm{K}$ with monovalent salt concentration $\rho = 1\,000\mathrm{mM}$, as the experimental pairing free energies $g_{\mathrm{h}}^{\mathrm{stack}}$ were determined at this concentration. The free energy shift for a helix with h nucleotides due to electrostatic interactions is then

$$\mathcal{G}_{\mathrm{h}}^{\mathrm{salt}} = h(g_{\mathrm{h}}^{\mathrm{DH}}(\rho) - g_{\mathrm{h}}^{\mathrm{DH}}(1000\,\mathrm{mM}))\,. \quad (4.8)$$

The use of Debye-Hückel theory to incorporate salt effects enables to include the overall dependence on temperature and salt concentration but involves several approximations. First, we are using Manning's counterion condensation theory to obtain the actual line charge density of ssRNA and dsRNA [138]. However, Manning condensation is known to underestimate the line charge at increasing salt concentration and therefore favors the bound state [142]. Second, when calculating the electrostatic energy of a loop we effectively use a ground state approximation and neglect conformational fluctuation effects. Third, when two ssRNA strands come together to form a helix the line charge density increases for the distance between two bases decreases. The salt dependence of the work to decrease the distance between two bases from $l_{\mathrm{ss}} = 6.4\,\text{Å}$ to $l_{\mathrm{ds}} = 3.4\,\text{Å}$ is neglected. This approximation favors the unbound state. Therefore, it is very important to validate the model we employ, which we do by detailed comparison with experimental data. From the favorable comparison with experiments we tentatively conclude that the various errors partially cancel and the resulting expression for the salt influence is quite accurate. We would like to point out that after determining $\mathcal{G}_{\mathrm{l}}^{\mathrm{init}}$ in eq. (4.1) no further fitting is done and only standard literature values are used.

Our theory is able to consider variations of the salt concentration as well as of the temperature, which makes it suitable to study RNA melting at various salt concentrations in a consistent way. However, since our approach is solely based on mean field theory it might break down in the case of multivalent ions, where correlations become important. Also, ion specific effects, which are important for the divalent ions such as Mg^{2+} [143], are not considered in our approach.

4.2.3. Response of the molecule to an external stretching force

In atomic force microscope or optical tweezers experiments it is possible to apply a stretching force F to the two terminal bases of the molecule. We model the stretching

response of the M non-nested backbone bonds, see fig. 3.1, with the freely jointed chain (FJC) model. A non-nested bond is defined as a backbone bond, which is neither part of a helix nor part of a loop. It is outside all secondary structure elements and therefore contributes to the extension observed in force spectroscopy experiments. The force dependent contribution to the free energy per non-nested monomer is given by

$$g_{\mathrm{f}}^{\mathrm{FJC}} = \mathcal{G}_{\mathrm{f}}^{\mathrm{FJC}}/M = -\mathrm{k_B}T\frac{l_{\mathrm{ss}}}{b_{\mathrm{ss}}}\ln\left(\frac{\sinh(\beta b_{\mathrm{ss}}F)}{\beta b_{\mathrm{ss}}F}\right), \tag{4.9}$$

where $\beta = 1/(\mathrm{k_B}T)$ is the inverse thermal energy and $b_{\mathrm{ss}} = 1.9$ nm is the Kuhn length of ssRNA [93] (we used the Kuhn length of ssDNA as the corresponding ssRNA data is not known to us). The stretching response of one non-nested monomer to an external force is then given by

$$x^{\mathrm{FJC}}(F) = -\frac{\mathrm{d}g_{\mathrm{f}}^{\mathrm{FJC}}}{\mathrm{d}F} = l_{\mathrm{ss}}\mathcal{L}(\beta F b_{\mathrm{ss}}) = l_{\mathrm{ss}}\left(\coth(\beta F b_{\mathrm{ss}}) + 1/(\beta F b_{\mathrm{ss}})\right), \tag{4.10}$$

$\mathcal{L}$ is the Langevin function. Electrostatic effects on the stretching response are considered to be small and hence are neglected [26, 71].

4.3. Partition function

So far we showed how to calculate the free energy of one given secondary structure. The next step is to enumerate all possible secondary structures in order to obtain the partition function, which allows to study the thermodynamics of the system. We extend our theory for homopolymeric RNA in chapter 3, eq. (3.3), and include sequence effects as well as a more elaborate helix parameterization. In our notation, the canonical partition function $Q_{i,j}^{M}$ of a sub-strand from base i at the 5'-end through j at the 3'-end depends on the number of non-nested backbone bonds M [1, 95, 97], see fig. 3.1. Neglecting tertiary interactions – and in particular pseudoknots – the recursion relations for $Q_{i,j}^{M}$ can be written as

$$Q_{i,j+1}^{M+1} = \frac{\nu_{\mathrm{f}}(M+1)}{\nu_{\mathrm{f}}(M)}\left[Q_{i,j}^{M} + \sum_{k=i+M+1}^{j-N_{\mathrm{loop}}} Q_{i,k-1}^{M} Q_{k,j+1}^{0}\right] \tag{4.11a}$$

and

$$Q_{k,j+1}^{0} = \sum_{h=1}^{(j-k-N_{\mathrm{loop}})/2} \exp\left(-\beta\mathcal{G}_{\mathrm{h}(k+h,j+1-h)}^{(k,j+1)}\right) \times \sum_{m=1}^{j-k-1-2h} Q_{k+1+h,j-h}^{m}\frac{\exp\left(-\beta\mathcal{G}_{\mathrm{l}}(m+2)\right)}{\nu_{\mathrm{f}}(m)}. \tag{4.11b}$$

An illustration of the recursion relation is depicted in fig. 3.3 on page 33. Eq. (4.11a) describes elongation of an RNA structure by either adding an unpaired base (first term) or by adding an arbitrary sub-strand $Q^0_{k,j+1}$ that is terminated by a helix. Eq. (**??**) constructs $Q^0_{k,j+1}$ by closing structures with m non-nested bonds, which are summed up in $Q^m_{k+1+h,j-h}$, by a helix of length h. $N_{\text{loop}} = 3$ is the minimum number of bases in a terminal loop. $v_{\text{f}}(M)$ denotes the number of configurations of a free chain with M links and drops out by introducing the rescaled partition function $\tilde{Q}^M_{i,j} = Q^M_{i,j}/v_{\text{f}}(M)$ and will not be considered further. $\mathcal{G}_{\text{h}}{}^{(k,j+1)}_{(k+h,j+1-h)}$ is the free energy of a helix beginning with base pair $(k, j+1)$ and ending with base pair $(k+h, j+1-h)$ according to eq. (4.5). $\mathcal{G}_{\text{l}}(m+2)$ is the free energy of a loop consisting of $m+2$ segments as given by eq. (4.1). $\mathcal{G}_{\text{l}}$ and $\mathcal{G}_{\text{h}}$ contain all interactions discussed in the previous section. Eq. (4.11) allows to compute the partition function in polynomial time ($\mathcal{O}(N^4)$). Further, it is possible to employ non-linear functions for $\mathcal{G}_{\text{l}}(m)$ and $\mathcal{G}_{\text{h}}(h)$; for instance, $\mathcal{G}_{\text{l}}(m)$ is strongly non-linear by virtue of eqs. (4.2) and (4.3).

The unrestricted partition function of the entire RNA, where the number of non-nested backbone bonds is allowed to fluctuate, is given by

$$Z_N = \sum_{M=0}^{N} \exp(-\beta g_{\text{f}}^{\text{FJC}} M) \tilde{Q}^M_{0,N} \tag{4.12}$$

and contains the influence of force via $g_{\text{f}}^{\text{FJC}}$ defined in eq. (4.9). The partition function Z_N contains all secondary structure interactions, but neglects pseudoknots and other tertiary interactions. Although one might miss some details, this approximation is known to work very well [29] and yields reliable predictions for the stability of single stranded nucleic acid structures [144].

Using the same ideas, we determine the minimum free energy (mfe) and the mfe structure. The mfe structure, is defined as the secondary structure, which gives the largest contribution to the partition function. Since the mfe cannot be derived from the partition function itself, it has to be determined from a slightly modified set of recursion relations, where the main idea is to replace the sums in eq. (4.11) by maximum-operators, see appendix B.4. The mfe structure is then obtained by backtracking the steps that lead to the mfe.

4.4. Salt dependence of melting curves

In this section we calculate melting curves for different salt concentrations by applying eqs. (4.1) and (4.5), which include our novel salt dependent free energy parameterization. In fig. 4.2 we compare experimental results [123, 124] with our predictions for the heat capacity of yeast tRNA-phe; the sequence is given in appendix B.7. The heat capacity is readily obtained by

$$C = T \frac{\partial^2 \text{k}_\text{B} T \ln Z_N}{\partial T^2}\,, \tag{4.13}$$

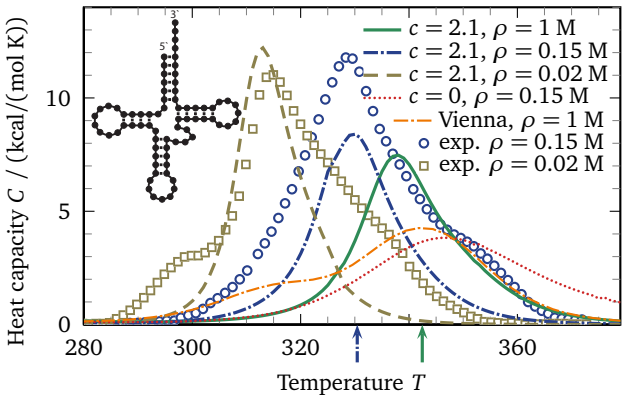
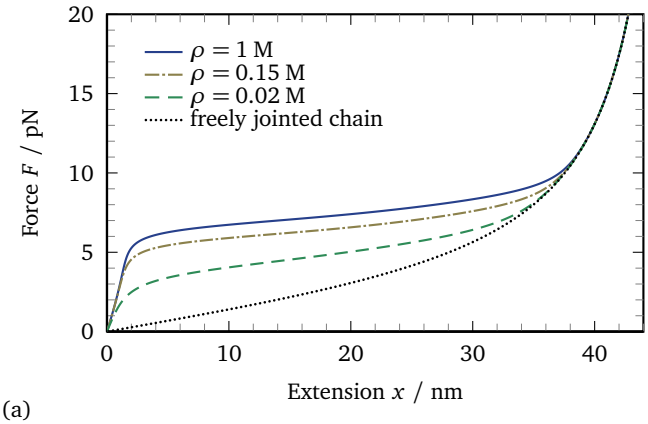
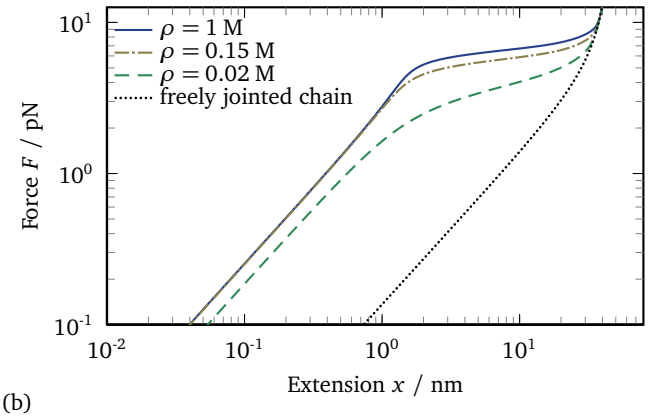


Figure 4.2.: Melting curve of the 76 bases long tRNA-phe of yeast; the predicted minimum free energy structure at $\rho = 1\,\mathrm{M}$, $T = 300\,\mathrm{K}$, $c = 2.1$ is shown and is identical to the experimentally observed secondary structure, cf. fig. 1.2b. Symbols denote experimental melting curves for NaCl concentrations $\rho = 20\,\mathrm{mM}$ (squares) and 150 mM (circles) [123]. Our predictions for different salt concentrations are depicted by the dashed (20 mM), dash-dotted (150 mM), and solid (1 M) lines. The respective arrows indicate melting temperatures obtained by experiments of another group [124]. The dotted line shows our prediction for $\rho = 150\,\mathrm{mM}$ and $c = 0$ and exemplifies that a non-zero loop exponent is responsible for rendering the transition more cooperative, which is also observed experimentally; for 150 mM and $c = 0$ the melting temperature is at higher temperatures since the energy parameters are optimized for $c = 2.1$. The orange dash-dotted curve is the prediction of the Vienna package, which uses a linearized multi-loop entropy corresponding to $c = 0$ and $\rho = 1\,\mathrm{M}$. This is to be compared to our prediction for $c = 2.1$ and $\rho = 1\,\mathrm{M}$: while the melting temperatures are similar, the cooperativity, i. e. the widths of the peaks are different due to different loop exponents.

where Z_N is the unrestricted partition function of the RNA at zero force, eq. (**??**). In all our calculations, we use the same literature parameter set for the stacking and pairing free energy $g_{\mathrm{h}}^{\mathrm{stack}}$. No additional fit parameter enters except the loop initialization free energy $\mathcal{G}_{\mathrm{l}}^{\mathrm{init}}$, which is determined in fig. 4.1a from a separate experimental data set. The salt dependence of the experimentally observed melting temperatures are reproduced well, compare fig. 4.2. We also plot a melting curve for loop exponent $c = 0$ and NaCl concentration $\rho = 150\,\mathrm{mM}$, which exhibits a far less cooperative transition than observed in the corresponding curve with $c = 2.1$. Finally, we compare our prediction for $\rho = 1\,\mathrm{M}$ and $c = 2.1$ with the prediction of RNAheat in the Vienna Package [91], which uses a linearized multi-loop entropy amounting to $c = 0$ in our framework. The predicted melting temperatures are almost identical. However, the widths of the peaks in both melting curves differ and the melting profile for $c = 2.1$ is more peaked. Taking all these observations together leads to the conclusion that only a combined use of logarithmic loop entropy (characterized by a non-zero loop exponent) and salt dependent free energy corrections leads to a correct prediction of melting curves. The additional features in the experimental data, e. g. the shoulder at lower



(a)



(b)

Figure 4.3.: Salt dependence of stretching curves of tRNA-phe for different salt concentrations $\rho =$ 20 mM, 150 mM, 1 M. Increasing salt concentration stabilizes the secondary structure due to screening of the electrostatic interaction. The dotted line is the theoretical prediction for the force extension curve of a freely jointed chain, eq. (4.10). The deviation of the predicted curves for RNA from the FJC curve is due to the presence of secondary structure. The observed plateau force is due to the rupture of the secondary structure. We show the force extension curves in (a) a linear plot and (b) a double logarithmic plot, indicating that the force extension curve is linear in the low-force regime, before the secondary structure is ruptured apart.

temperatures and the increased width of the experimental curves might be attributed to tertiary structure rearrangements, which are not captured by our approach, or to melting occurring in multiple stages.

4.5. Salt dependence of stretching curves

Apart from temperature, force is an important variable to study denaturation of RNA molecules [93–95, 125, 132, 145–151]. In fig. 4.3 we show the salt dependence of stretching curves for yeast tRNA-phe. The stretching curves have been obtained by describing the behavior of the M non-nested backbone bonds, see fig. 3.1, with the freely jointed chain (FJC) model, see eq. (4.10),

$$x(F) = \mathrm{k_B}T \frac{\partial \ln Z_N}{\partial F} = \mathrm{k_B}T \frac{\partial \ln Z_N}{\partial g_\mathrm{f}^\mathrm{FJC}} \frac{\partial g_\mathrm{f}^\mathrm{FJC}}{\partial F} = M x^\mathrm{FJC}(F)\,, \tag{4.14}$$

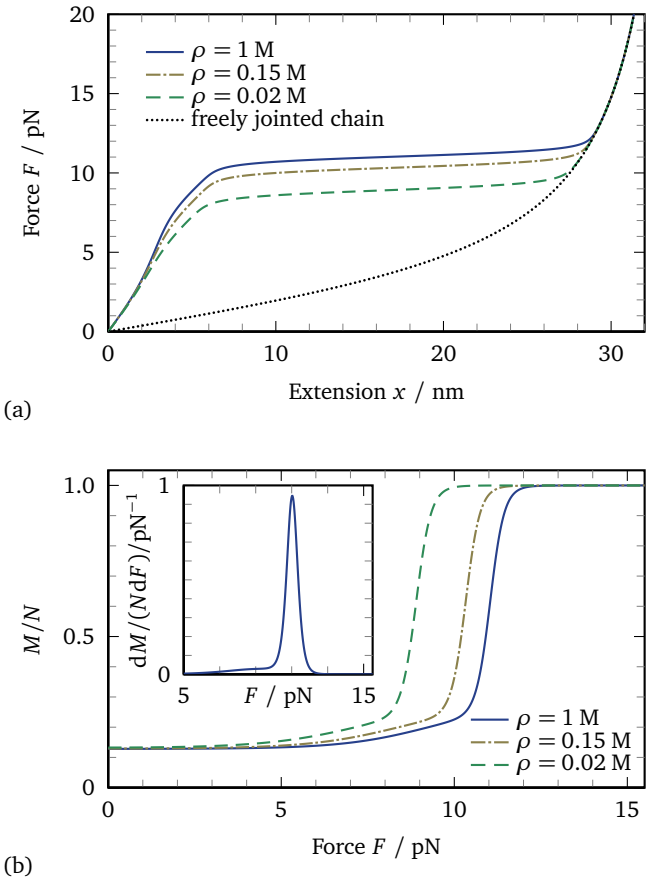


Figure 4.4.: (a) Salt dependence of stretching curves of the 56 bases long RNA hairpin P5ab [125] for different salt concentrations $\rho = 20$ mM, 150 mM, 1 M. Increasing salt concentration stabilizes the secondary structure due to screening of the electrostatic interaction. The dotted line is the force extension curve of an FJC, eq. (4.10). (b) The fraction of non-nested segments as a function of force. One observes that the unzipping of the hairpin happens in two stages, which is visible as a shoulder for 5 pN $\lesssim F \lesssim$ 6 pN - 10 pN (exact values depend on the salt concentration) and a successive cooperative transition. The inset shows the derivative $\mathrm{d}M/(N\mathrm{d}F)$ for $\rho = 1$ M, where the first transition is visible as a shoulder at $F \approx 8$ pN. The sharp peak at $F = 11$ pN is the rupture of the complete helix.

where we used the definition of the number of non-nested backbone segments

$$M = -\mathrm{k_B}T \frac{\partial \ln Z_N}{\partial g_\mathrm{f}^\mathrm{FJC}} \,. \tag{4.15}$$

As for the melting curves one observes that increasing salt concentration stabilizes the structure leading to higher unfolding forces. All curves converge in the large force limit to a freely jointed chain of the length of the whole RNA molecule $(N-1)l_\mathrm{ss}$, where $N = 76$ is the number of bases in the chain. The deviation for small forces from this theoretical prediction is due to the secondary structure of RNA, which is present at small forces and which becomes disrupted at forces $F \gtrsim 5$ pN. In fig. 4.4a we show the force extension curve of the P5ab hairpin [125]; the sequence is given in appendix B.7. Apart from the salt dependence of the force extension curve one observes that the unzipping of the helix happens in two stages. This is seen best by considering the fraction of non-nested segments and its derivative, fig. 4.4b. The first stage is a smooth unzipping of the first three base pairs up to the bulge loop visible as a shoulder at $F \approx 8$pN in the derivative. The second stage is a sharp transition where the rest of the hairpin is unzipped. In fig. 4.5 we show mfe predictions for the

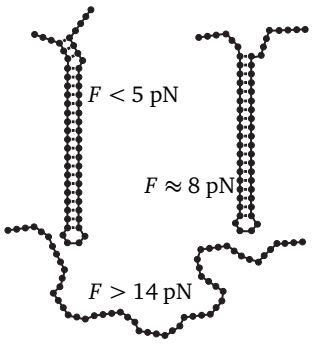


Figure 4.5.: Predicted minimum free energy structures of the hairpin P5ab at different forces, see also appendix B.4. For $F < 5$ pN the hairpin is in the native state with all base pairs intact. At $F \approx 8$ pN the first helix, consisting of three base pairs and bounded by the bulge loop, is ruptured. This causes the first smooth transition. Forces $F \gtrsim 14$ pN lead to the unzipping of the whole hairpin in a very cooperative fashion.

secondary structure at different forces for $\rho = 1$ M NaCl. For $F < 5$ pN we correctly predict the experimentally observed native state with all base pairs intact [125]. For forces $F \approx 8$ pN an intermediate state appears, where the first three base pairs are unzipped up to the bulge loop. Denaturation is observed for $F \gtrsim 14$ pN.

The native structure of the P5ab hairpin contains the stacked pairs $\begin{smallmatrix}\mathrm{GG}\\ \mathrm{AA}\end{smallmatrix}$ – bp(17,42) and bp(18,41) [125]. For this stack, no free energy parameters are available and we use the parameters for the stack $\begin{smallmatrix}\mathrm{GG}\\ \mathrm{UU}\end{smallmatrix}$, instead. However, other parameterizations for this stack work equally well and reproduce the experimental transition force within errors, see fig. 4.6.

4.6. Phase diagrams of the P5ab RNA hairpin

With the tools established in the previous sections we are now able to study phase diagrams of RNA. We consider the P5ab hairpin, which is a well studied system [125, 131, 132, 152, 153]. In fig. 4.7b the phase diagram in the F-ρ plane is shown for $T = 298$ K, 300 K, 320 K and $c = 2.1$. The phase boundary is defined as the force where half of the helical section is unzipped. In particular we exclude the three unpaired bases at the 5'- and the four bases at the 3'-end, see fig. 4.5, and obtain the condition $M - 7 = (N - 7)/2$ for the transition point. This position is depicted by an arrow in fig. 4.7a. Below the phase boundary the hairpin is stable, above the molecule is denatured and completely open. In fig. 4.7b we additionally include experimental results by Liphardt et al. [125] and observe a perfect match with our results. It is important to note that this transition is not a phase transition in the strict statistical mechanics sense, but just a crossover. A true phase transition is defined as a non-analyticity of the free energy, which can only occur for an infinite system with long-range interactions [1]; this has been studied in chapters 2 and 3. The three-dimensional phase space we are considering is spanned by temperature, force, and salt concentration. In figs. 4.8 and 4.9 we show slices in the F-T and in the T-ρ plane. The phase boundary for the F-T plane is determined the same way as in the F-ρ plane,

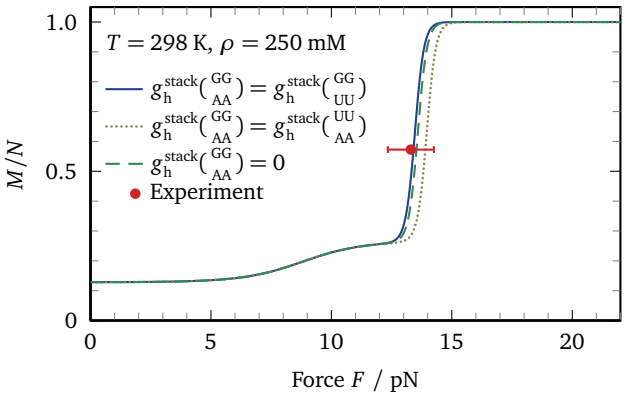


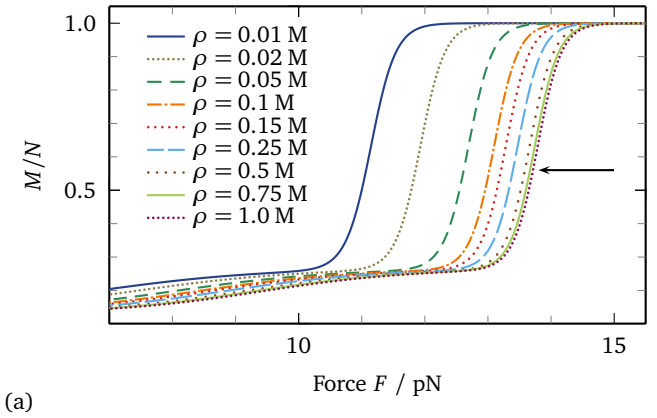
Figure 4.6.: The effect of different parameterizations for the free energy parameters for $^{GG}_{AA}$ on the denaturation curve is marginal. Here, the fraction of non-nested backbone bonds is plotted against the force for the P5ab hairpin and $T = 298\,\mathrm{K}$, $\rho = 250\mathrm{mM}$, $c = 2.1$. The solid line is obtained with the parameters used in this chapter $g_h^{stack}(^{GG}_{AA}) = g_h^{stack}(^{GG}_{UU})$. The dotted line is obtained by using $g_h^{stack}(^{GG}_{AA}) = g_h^{stack}(^{UU}_{AA})$, whereas the dashed line is for $g_h^{stack}(^{GG}_{AA}) = 0$. All three curves coincide and differ only at the transition exhibiting a slightly different transition force, which is, however, identical to the experimentally observed force within errors [125]. The values of the free energy parameters are given in appendix B.6.

yet with varying temperature and fixed salt concentration. The phase boundary in the T-ρ plane is determined differently: heat capacity curves as a function of temperature are calculated for different salt concentrations. The position of the peaks in the heat capacity curves (one is depicted by an arrow in fig. 4.9a determine the phase diagram in fig. 4.9b. Therefore slight differences between the phase diagrams in figs. 4.7, 4.8 on the one hand and fig. 4.9 on the other hand may arise.

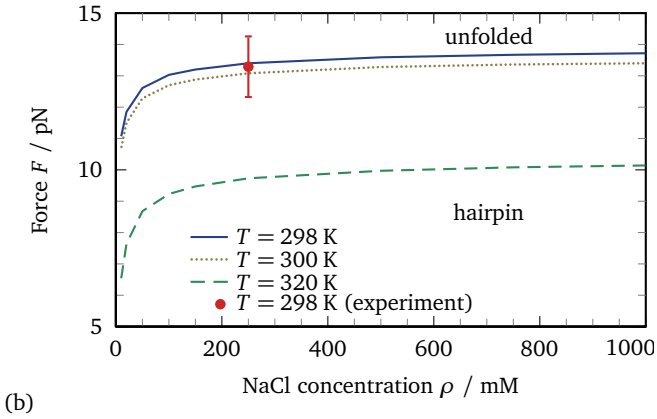
We observe that for large salt concentrations the denaturation forces and temperatures are rather independent of the salt concentration, see figs. 4.7 and 4.9. Only when the Debye screening length κ^{-1} is of the order of the typical length scales of RNA, which is the case for $\rho \lesssim 100\,\mathrm{mM}$, a marked dependence on the salt concentration is observed.

4.7. Conclusions

We construct a theory for RNA folding and melting that includes the effects of monovalent salt, loop entropy, and stretching forces. Our theory is based on salt and temperature dependent modifications of the free energies of RNA helices and loops that include electrostatic interactions on the linear Debye-Hückel level and conformational fluctuation effects via the asymptotic, non-linear expression for the entropy of loop formation. Decreasing salt concentration is shown to generally destabilize RNA folds and to lower denaturation temperatures and forces. The predictions are in excellent

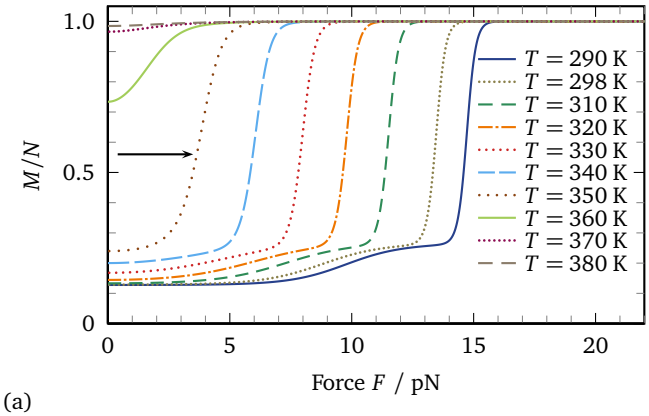


(a)

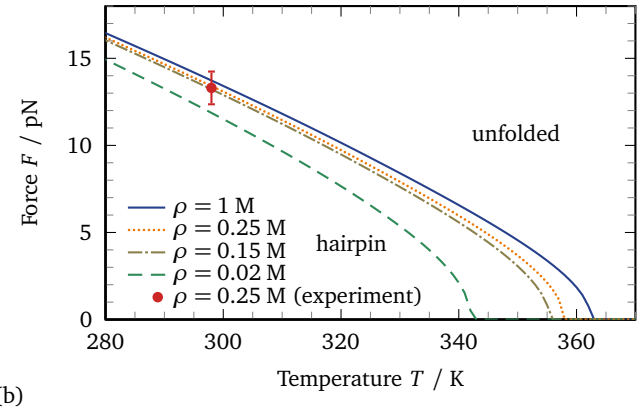


(b)

Figure 4.7.: (a) Fraction of non-nested segments of the P5ab hairpin as a function of force for different salt concentrations and constant temperature $T = 298$ K. The position of the crossover, which is defined as the point where $M - 7 = (N - 7)/2$, i. e. $M/N = 0.56$ (indicated by the arrow), determines the phase diagram. (b) Phase diagram of the P5ab hairpin in the F-ρ plane for different temperatures $T = 298$ K, 300 K, 320 K. Below the curve the RNA is in the hairpin phase, above the RNA is denatured and completely open. The symbol at $\rho = 250$ mM, $F = 13.3$ pN, and $T = 298$ K denotes the experimental data by Liphardt et al. [125] and coincides perfectly with our prediction.

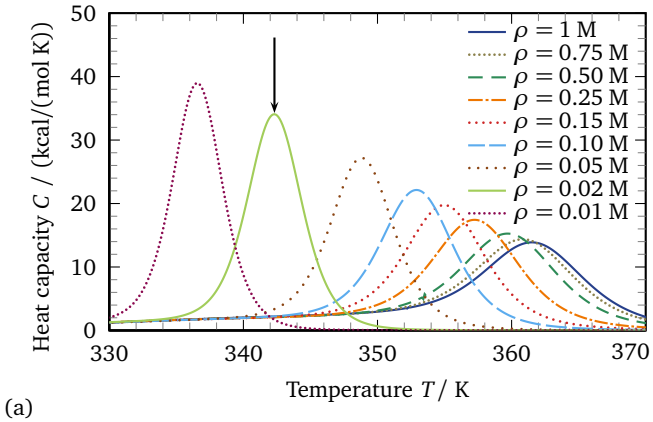


(a)

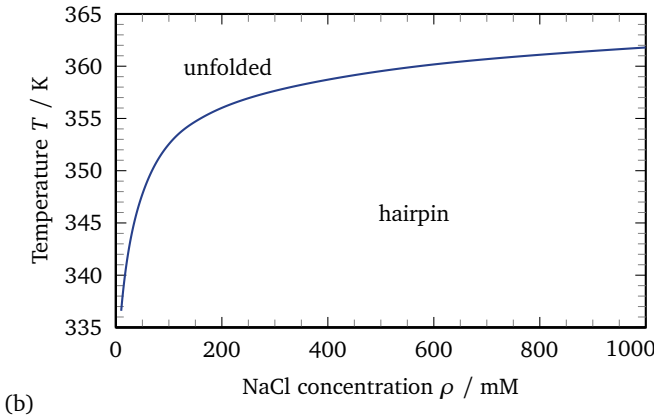


(b)

Figure 4.8.: (a) Fraction of non-nested segments of the P5ab hairpin as a function of force for different temperatures and constant salt concentration $\rho = 250$ mM. The position of the crossover (arrow, $M/N = 0.56$) determines the phase boundary. With increasing temperature a decrease of the denaturation force is observed. Above the melting temperature $T_m \approx 358$ K the molecule is always in the denatured state. (b) Phase diagram of the P5ab hairpin in the F-T plane. Below the curve the RNA is in the native hairpin phase, above the RNA is denatured and completely open. The symbol denotes experimental values [125].



(a)



(b)

Figure 4.9.: (a) Heat capacity curves for different salt concentrations and constant temperature. The position of the peaks moves to the right with increasing salt concentration. The positions of the peaks, denoted exemplarily for one curve by the arrow, determine the phase diagram. (b) Phase diagram of the P5ab hairpin in the T-ρ plane. Below the curve the RNA is in the native hairpin phase, above the RNA is denatured and completely open.

agreement with experimental data as shown for two different scenarios, namely the heat capacity curves for the thermal denaturation of tRNA-phe and the response of the P5ab RNA hairpin to an external pulling force.

Due to the usage of the linear Debye-Hückel approximation in conjunction with the Manning condensation concept, our approach is limited to monovalent salt and neglects ion-specific effects. Electrostatic nonlinear and correlation effects could in principle be taken into account by more advanced modeling using variational approaches [142], while ion-specific effects could be straightforwardly included using effective interactions between different ions and RNA bases [154]. More complex phenomena involving multivalent ions such as Mg^{2+} could in principle be modeled by allowing for a few tertiary contacts, which is left for future studies.

We find that for a proper description of RNA melting curves, correct modeling of the loop entropy is crucial. A non-zero loop exponent leads to an increased cooperativity of the melting transition and thus makes the heat capacity curve narrower in good agreement with experimental results. We conclude that for a correct description of RNA denaturation thermodynamics both loop entropy and salt effects are important and should be included in standard structure and melting curve prediction software.

CHAPTER 5

REPTATION DYNAMICS IN POLYMER GLOBULES: DIFFUSION VS. PULLING SIMULATIONS

5.1. Introduction

5.1.1. Motivation

The previous chapters covered equilibrium properties of nucleic acids. However, conformational dynamics of nucleic acids and polymers in general play a crucial role in biological systems. For example, translocation of polymers such as RNA requires bond breaking and drastic conformational rearrangements to accommodate the geometrical constraints imposed by a pore [97, 155, 156]. Unfolding of biopolymers in force spectroscopy experiments induces huge conformational changes on the polymer [157], which may occur in a dissipative fashion [158, 159]. RNA sequences known as riboswitches also experience conformational changes upon binding of small metabolites [160]. All of these transitions necessitate spatial rearrangements of the molecules and reptational dynamics of polymers within a collapsed globule [161]. The time scale on which these changes happen is influenced by the medium, particularly the solvent viscosity, and by the polymer itself through internal interactions. Therefore, if one desires to understand the dynamics of processes such as protein folding [162–165], packing of DNA in the chromosome [166–169], polymer collapse [170–172], or adsorption [8, 173, 174], all dissipation and viscous effects have to be considered [175].

While solvent effects are due to the viscosity of the medium, there are a number of effects that together give rise to internal friction behavior. Local interactions such as conformational transitions of backbone bonds [162, 176], entanglement effects and excluded volume interactions in polymer systems, degrees of freedom orthogonal to the reaction coordinate [177], and the breakage and reformation of cohesive bonds [163, 178, 179] all lead to dissipation in systems out of equilibrium. For globular homopolymers, proteins in the molten globule phase [88], and disordered intermediates during protein folding [180], these effects manifest themselves as a rough free

energy landscape with many competing and intermediate states [181, 182]. The resulting bond breakage and reformation is prevalent during the folding process, leading directly to the concept of an effective internal viscosity [183–185]. There have been a large number of coarse-grained simulation studies on the force induced unfolding of proteins [186], globular polymers [170, 173, 187–189], and the diffusion of knots along a stretched chain [190–192]. Cohesive interactions between polymer monomers have been shown to lead to a phase transition from a liquid-like to a solid-like globule for long enough chains [7, 193–200].

5.1.2. Outline

In this chapter we study the effect of internal friction on two model systems, which both can be realized experimentally. We perform Brownian dynamics simulations on a homopolymer. Attractive interactions are modeled with a Lennard-Jones potential, where the cohesive strength ε is varied. First, we study the diffusion of an unknotted globule, which forms on a polymer held at constant extension smaller than the contour length. This simulation is conducted under equilibrium conditions and no external forces are applied. For $\varepsilon < \varepsilon_s$, where $\varepsilon_s \approx 2.5 k_B T$ depending on the globule size, normal diffusion of the globule is observed and characterized by the diffusivity D_G, which is a direct measure of the internal viscosity since motion of the globule requires internal rearrangements. We observe that the diffusion constant is proportional to the reciprocal globule size $D_G \propto N_G^{-1}$. The diffusion constant shows a marked dependence on the cohesive strength ε, as well. For $\varepsilon > \varepsilon_s$, no diffusion is observed on the simulation time scales and the globule is stuck in a single conformation. This reflects the change of the internal dynamics going from liquid-like ($\varepsilon < \varepsilon_s$) to solid-like ($\varepsilon > \varepsilon_s$). Support for this interpretation is given by our second set of simulations, where we measured the dissipated work during stretching and unraveling the globule with a finite velocity. This is the same setup as studied by us before, yet with significantly longer chains and slower pulling velocities [183], and allows to study the liquid-solid transition. Like in our equilibrium simulations, we observe two different regimes. For large $\varepsilon > \varepsilon_s$, the force extension curve is characterized by huge fluctuations, which abruptly cease when $\varepsilon < \varepsilon_s$. These extreme fluctuations in the force extension curve are present at small extensions and hence large globules. Pulling decreases the number of monomers inside the globule causing the fluctuations to vanish once N_G is below a certain threshold. Therefore, reducing the number of monomers N_G in the globule by stretching the chain drives the system from the solid into the liquid state. In the liquid state we perform extensive simulations to obtain the internal viscosity. We show that in the liquid regime the internal friction is extensive, meaning that the dissipated work per monomer scales linearly with the chain length N and the pulling velocity v, $\Delta W/N \propto Nv$. We also show that this scaling extends into size regimes beyond what has been studied previously [183]. For the reptational dynamics this means that in the liquid-like regime, a Rouse-type model with a local friction is valid. Clearly, internal friction effects will dramatically influence the time scales of reptational dynamics. In

the solid-like regime, no definite conclusion on the reptation mode is possible from the simulations we performed.

5.2. Brownian dynamics

We model the homopolymer by N freely jointed beads of radius a interacting with a potential $U(\{\boldsymbol{r}^{(N)}\})$, which depends on the set of the positions of the individual beads $\{\boldsymbol{r}^{(N)}\}$. The position $\boldsymbol{r}^{(i)}$ of the i^{th} bead obeys the overdamped Langevin equation [201, 202]

$$\frac{\partial \boldsymbol{r}^{(i)}}{\partial t} = -\mu_0 \boldsymbol{\nabla}_{\boldsymbol{r}^{(i)}} U(\{\boldsymbol{r}^{(N)}\}) + \mu_0 \boldsymbol{f}^{(i)}(t)\,, \tag{5.1}$$

where $\mu_0 = 1/(6\pi\eta_0 a)$ is the Stokes mobility of a sphere with radius a in a solvent with viscosity η_0. $\boldsymbol{f}^{(i)}(t)$ is the random force acting on the i^{th} bead. The components of the random force $f_\alpha^{(i)}(t)$, $\alpha = x, y, z$, satisfy the Stokes-Einstein relation

$$\langle f_\alpha^{(i)}(t)\, f_\beta^{(j)}(t')\rangle = \frac{2\mathrm{k_B}T}{\mu_0}\, \delta_{ij}\delta_{\alpha\beta}\delta(t-t') \tag{5.2}$$

and are unbiased $\langle f_\alpha^{(i)}(t)\rangle = 0$. Hydrodynamic interactions are neglected since we are interested in the internal friction of the polymer caused by the monomer interactions. We introduce dimensionless quantities and express energies in units of the thermal energy $\mathrm{k_B}T$, lengths in units of the bead radius a, and times in units of $\tau = a^2/(\mu_0 \mathrm{k_B}T)$, which is the characteristic diffusion time of a single bead. Using the dimensionless quantities $\tilde{t} = t/\tau$, $\tilde{\boldsymbol{r}} = \boldsymbol{r}/a$, $\tilde{\boldsymbol{f}} = \boldsymbol{f}/(\mathrm{k_B}T/a)$, and $\tilde{U} = U/(\mathrm{k_B}T)$ eq. (5.2) reads

$$\frac{\partial \tilde{\boldsymbol{r}}^{(i)}}{\partial \tilde{t}} = -\boldsymbol{\nabla}_{\tilde{\boldsymbol{r}}^{(i)}} \tilde{U}(\{\tilde{\boldsymbol{r}}^{(N)}\}) + \tilde{\boldsymbol{f}}^{(i)}(\tilde{t})\,. \tag{5.3}$$

For the Brownian dynamics simulation we use a discretized version of eq. (5.2) with a time step $\Delta t = 0.0005\tau$ and obtain the coordinates at time $t = n\Delta t$ [203]

$$\tilde{\boldsymbol{r}}^{(i)}(n+1) = \tilde{\boldsymbol{r}}^{(i)}(n) + \left(-\boldsymbol{\nabla}_{\tilde{\boldsymbol{r}}^{(i)}} \tilde{U}(\{\tilde{\boldsymbol{r}}^{(N)}\}) + \tilde{\boldsymbol{f}}^{(i)}(n)\right) \Delta\tilde{t} \tag{5.4}$$

and

$$\langle \tilde{f}_\alpha^{(i)}(n)\, \tilde{f}_\beta^{(j)}(n')\rangle = \frac{2}{\Delta\tilde{t}}\, \delta_{ij}\delta_{\alpha\beta}\delta_{nn'}\,. \tag{5.5}$$

5.3. Diffusivity of a globule along a periodic chain: equilibrium simulations

In a preceding work, the internal viscosity η_G of a homopolymeric globule has been determined by measuring the dissipated energy due to non-equilibrium pulling at finite

speed [183]. We showed that for moderate attractive interactions ε the internal friction is the dominant dissipative effect as long as the majority of the monomers are part of the globule. However, as the globule is unraveled the size of the globule decreases and more and more beads in the linker sections dissipate energy due to solvent friction. These effects had to be subtracted in order to obtain the internal globule friction. Here, we introduce a new system, where we perform equilibrium simulations and the size of the globule stays constant. We study the diffusive behavior of a globule that forms on a chain, which is held at constant extension smaller than the contour length. As motion of the globule necessitates internal rearrangements via reptation moves, the diffusion constant of the globule is a measure of the internal friction. Therefore, the microscopic internal dynamics inside the globule manifests itself as a macroscopic and experimentally observable quantity: the diffusivity of the globule $D_\mathrm{G} \propto 1/\eta_\mathrm{G}$. This system can be realized experimentally by studying a polymer held at constant extension in an optical or magnetic tweezers setup [107]. The cohesive force can be varied by changing the solution conditions, whereas the size of the globule can be varied by changing the trap distance.

5.3.1. Model

Description of the system

The diffusivity D_G of a globule on an extended chain is a measure of the viscosity η_G inside the globule since they are related via $\eta_\mathrm{G} \propto 1/D_\mathrm{G}$. To model this system we consider a polymer held at a fixed extension $L = 50a$ in the x-direction, which is smaller than the contour length $L_\mathrm{c} = 2a(N-1)$ of the polymer. A globule will form due to the attractive Lennard-Jones interaction between the monomers, see fig. 5.1. To eliminate finite size effects we introduce periodic boundary conditions in x-direction, which are implemented via the minimum image convention [204]: the components of the vector pointing from $\boldsymbol{r}^{(i)}$ to $\boldsymbol{r}^{(j)}$ are given by

$$r_x^{(ij)} = \mathcal{D}(r_x^{(i)}, r_x^{(j)}; L) = ((r_x^{(j)} - r_x^{(i)} + 3L/2) \mod L) - L/2 \tag{5.6a}$$

$$r_{y/z}^{(ij)} = r_{y/z}^{(j)} - r_{y/z}^{(i)} \,. \tag{5.6b}$$

We use a box size $L = 50a$ in all simulations. The potential energy has four contributions

$$U = U_\mathrm{b} + U_\mathrm{r} + U_\mathrm{LJ} + U_\mathrm{tr} \,. \tag{5.7}$$

U_b and U_r are the bond potential acting between neighboring monomers. The backbone bonds are modeled by harmonic potentials

$$U_\mathrm{b} = \frac{\kappa}{2} \sum_{i=1}^{N-1} (r^{(ii+1)} - 2a)^2 \,, \tag{5.8}$$

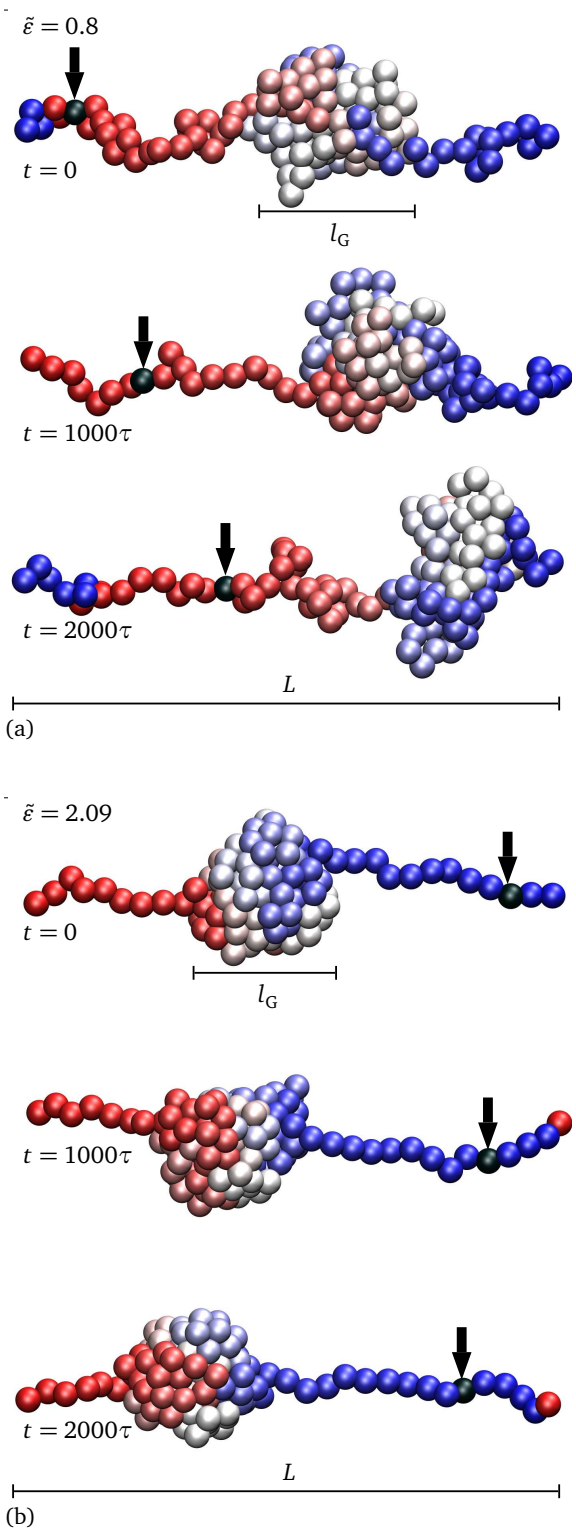


Figure 5.1.: The diffusion of a globule on a stretched polymer is simulated. The diffusivity D_G, the number of monomers N_G inside the globule, the extension of the globule l_G, and the fluctuations of the linkers (stretched part) depend on the cohesive strength ε, eq. (5.3.1). Snapshots at different times for (a) $\tilde{\varepsilon} = 0.8$ and (b) $\tilde{\varepsilon} = 2.09$ are shown. To prevent motion of the linkers and model a polymer held at fixed extension the monomer i_{tr} in the middle of the linkers (indicated by the arrow) is trapped by a harmonic potential, eq. (5.3.1). Periodic boundary conditions in a box of length L using the minimum image convention are employed to model an infinite polymer. The color coding indicates the index of a monomer along a chain.

with $\kappa = 200 k_B T/a^2$ and $r^{(ij)} = |\boldsymbol{r}^{(ij)}|$, see eq. (4.11a). As periodic boundary conditions are employed, the polymer forms a closed ring, which is achieved by connecting the first and last monomer by

$$U_r = \frac{\kappa}{2}(r^{(N1)} - 2a)^2 \,. \tag{5.9}$$

The monomer cohesion and excluded volume interactions are modeled with a Lennard-Jones potential

$$U_{LJ} = \varepsilon \sum_{i=1}^{N} \sum_{j=1}^{i-1} \left(\left(\frac{2a}{r^{(ij)}} \right)^{12} - 2 \left(\frac{2a}{r^{(ij)}} \right)^{6} \right) . \tag{5.10}$$

$\varepsilon = 0$ models an ideal phantom chain without excluded volume interactions and attractive interactions between the monomers. For $\varepsilon > 0$ the first term in eq. (5.3.1) accounts for the repulsive excluded volume interaction at short separations, whereas the second term is responsible for cohesion. For $0 \leq \varepsilon < \varepsilon_G \approx 0.5$ the polymer is in the swollen state and no globule exists. Increasing ε above ε_G causes the polymer to collapse and a globule forms [7, 25, 170, 183, 193–200]. The position of the globule transition ε_G depends on the system size [193]. For even larger values $\varepsilon > \varepsilon_s$ a solid phase appears.

U_{tr} is an external trapping potential for individual monomers in the linker in order to model e. g. optical tweezers. As we are interested in the motion of the globule and not of the complete chain we have to prevent the linker from moving. The linker is the stretched part that does not belong to the globule, fig. 5.1. Therefore, we introduce a harmonic trap potential

$$U_{tr} = \frac{\kappa_{tr}}{2}(\boldsymbol{r}^{(i_{tr})} - \boldsymbol{R})^2 \,, \tag{5.11}$$

which is located at $\boldsymbol{R}$ and acts on bead i_{tr} that is in the middle of the linker, with stiffness $\kappa_{tr} = 10 k_B T/a^2$. The trapped beads in the moment of the snapshots shown in fig. 5.1 are indicated by arrows. The index of the trapped bead i_{tr} depends on the position of the globule and which beads belong to the globule. The exact definition of the globule is described in the next section. Due to globule motion the index of the bead in the middle of the linker might change from i_{tr} at time t to i'_{tr} at some later time t'. If this happens we update the x-coordinate of the trap position to the new position $R_x(t') = R_x(t) + 2a\,\mathcal{D}(i_{tr}, i'_{tr}; N)$, eq. (4.11a), and the trapping force then acts on bead i'_{tr}. This setup allows to study the diffusion of the globule along the x-axis.

The effect of using periodic boundary conditions rather than long linkers to each side is fourfold. First, the simulation is sped up as the system is smaller. Second, the globule is stabilized as the configurational space for the unfolded system is reduced as only fluctuations up to a wavelength of the order of the simulation box are possible. Third, motion of the globule as a whole without internal reptation is reduced. Fourth, knots may not form as the polymer forms a closed ring.

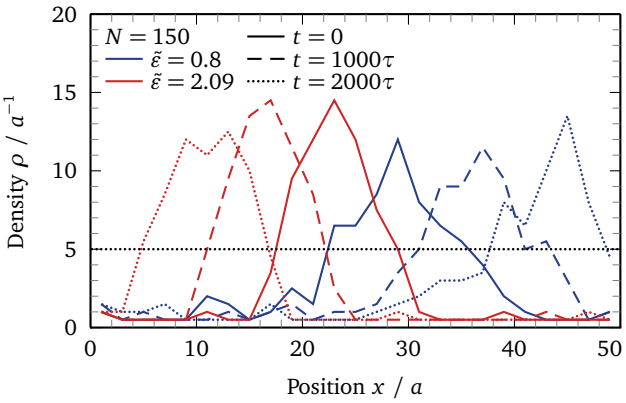


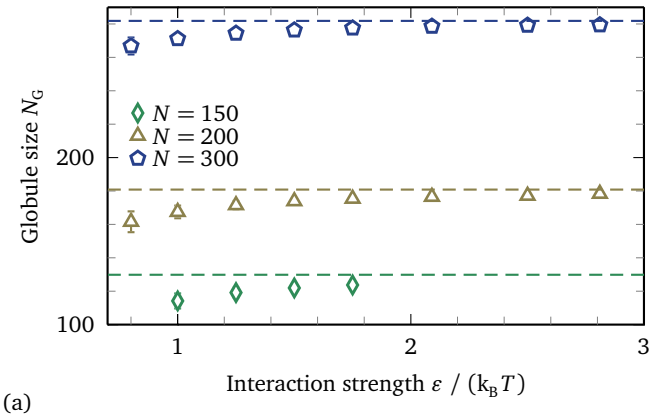
Figure 5.2.: The density profile of the globule depends on the cohesive strength ε. The globule is defined as a region where the monomer density is $\rho > 5/a$ (indicated by the horizontal dotted line). The density profiles correspond to the snapshots shown in fig. 5.1.

Definition of the globule

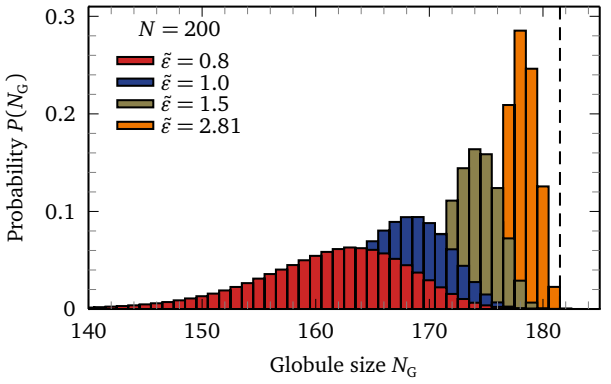
We partition the x-axis into bins of width $l_b = 2a$ and measure the monomer density ρ_k in the k^{th} bin, $k = 1, \ldots, L/l_b$. The globule is defined as the region, where the monomer density projected on the x-axis is $\rho_k > 5/a$. In fig. 5.2 the density profiles of the snapshots in fig. 5.1 are shown. If at the edge of the globule the density profile is not monotonous and $\rho_k < 5/a$, but $\rho_{k-1}, \rho_{k+1} > 5/a$, we account for such cases by adding these bins to the globule, too. This ensures that we end up with a list of bins that are connected. For small $\tilde{\varepsilon} < 1$ and small $N < 150$ it is possible that – according to the above definition – more than one globule exists or that the globule is smeared out over the complete simulation box, however we will not consider simulations in which this occurs. Such complications are never observed for $\tilde{\varepsilon} > 1$ or $N > 150$.

Definition of the index of the trapped bead

Let us first define the index of the bead i_r, which is at the right edge of the globule, and the index of the bead i_l, which is at the left edge of the globule. i_r is obtained by picking an arbitrary monomer inside the globule and moving along the chain with increasing monomer index. i_r is the largest index that is still in a bin belonging to the globule. We also check for loops, which leave the globule and return again: If a loop occurs, we add all monomers of the loop to the globule even if they lie in a bin outside the globule. i_l is defined in the same way yet by decreasing the index. The index of the central bead i_c in the middle of the globule and the number of monomers N_G inside the globule are defined as $i_c = (i_l + i_r)/2$, $N_G = i_r - i_l + 1$ if $i_l < i_r$ and $i_c = (i_l - N + i_r)/2$, $N_G = i_l - i_r + 1$ if $i_l > i_r$. The index of the central bead i_c yields the index of the trapped bead $i_{tr} = i_c + N/2$, which therefore depends on the motion of the globule, see fig. 5.1.



(a)



(b)

Figure 5.3.: (a) The number of monomers N_G inside the globule increases weakly with the interaction strength ε until it finally levels off. The limiting values are depicted by the horizontal broken lines and are given by the root of eq. (5.3.2) describing a spherical globule and tightly stretched linkers. Error bars denote the standard deviations of the distributions. (b) Probability distribution of the number of monomers N_G inside the globule for chain length $N = 200$ and different cohesive strengths $\tilde{\varepsilon} = 0.8, 1, 1.5, 2.81$. For small ε considerable size fluctuations are observed, which decrease upon increasing ε. Further, increasing ε increases N_G up to the limiting value given by eq. (5.3.2) (broken line).

5.3.2. Results

Number of monomers inside the globule

The number of monomers N_G inside the globule depends on the chain length N and on the cohesive strength ε, fig. 5.3a. Increasing ε increases N_G up to a limiting value which is given by the root of

$$N_G^{1/3} + (N - N_G) = L/(2a)\,. \tag{5.12}$$

Eq. (5.3.2) describes a spherical globule consisting of N_G monomers with a tightly stretched linker with $N - N_G$ monomers. As the limiting value of our simulations coincides with the predicted value of eq. (5.3.2), as shown in fig. 5.3a, our definition of the globule is justified. For large ε, the linkers are tightened and stretched as the energetic gain of a monomer joining the globule outweighs the entropy of a loosely fluctuating linker. For small ε, monomers are not tightly bound to the globule as can be seen qualitatively from fig. 5.1. Therefore size fluctuations of the globule are more substantial for small cohesive strengths and decrease upon increasing ε, fig. 5.3b.

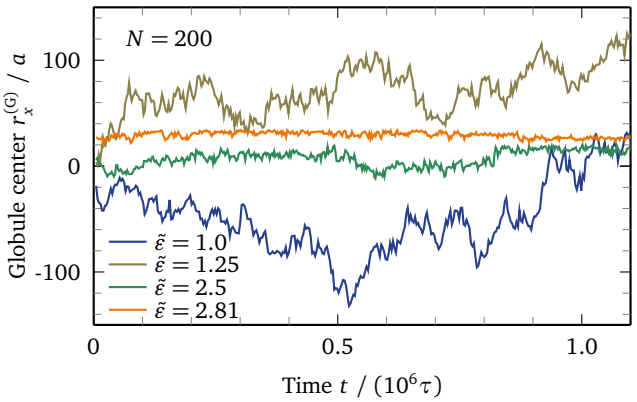


Figure 5.4.: Trajectories of the x-coordinate of the center of the globule $r_x^{(G)}$ for $N = 200$ and interaction strengths $\tilde{\varepsilon} = 1, 1.25, 2.5, 2.81$, eq. (5.3.2). The smaller the attractive interaction between the monomers the more mobile the globule is. If ε is large, the globule is frozen in a single conformation and does not move.

Definition of the center of the globule

Since we are considering a ring-like polymer subject to periodic boundary conditions in the x-direction, the center of mass of the globule has to be defined carefully. We introduce the non-periodic polymer coordinates $\{\hat{\boldsymbol{r}}^{(N)}\}$, which are not restricted to the first simulation cell. They are defined recursively starting from the first monomer with $\hat{\boldsymbol{r}}^{(1)} = \boldsymbol{r}^{(1)}$ and

$$\hat{\boldsymbol{r}}^{(i+1)} = \hat{\boldsymbol{r}}^{(i)} + \boldsymbol{r}^{(ii+1)}\,, \tag{5.13}$$

for $i \geq 1$, where $\boldsymbol{r}^{(ii+1)}$ is given by eq. (4.11a). The center of the globule $\boldsymbol{r}^{(G)}$ is calculated by using the non-periodic polymer coordinates, eq. (5.3.2),

$$\boldsymbol{r}^{(G)} = \frac{1}{N_G} \sum_{i=i_l}^{i_r} \hat{\boldsymbol{r}}^{(i)}\,. \tag{5.14}$$

Periodic boundary conditions may introduce jumps of the size of the box L in the trajectory $\boldsymbol{r}^{(G)}(t)$. We remove those jumps by connecting the value of any quantity $q(t)$ at time t, which is subject to periodic boundary conditions, with the value $q(t-\Delta t)$ in the previous time step. This is done by recording $q(t-\Delta t) + \mathcal{D}(q(t), q(t-\Delta t); L)$ instead of $q(t)$. It is important to note that information is neither lost nor added by applying this operation, since only multiples of the box size are added to the quantity.

Diffusivity of the globule

This setup enables us to study the motion of the globule in space in a fashion that is coupled to its internal reptation dynamics. In fig. 5.4 trajectories of the x-coordinate of the center of the globule $r_x^{(G)}$, eq. (5.3.2), is shown for $N = 200$ with various cohesive strengths ε. Fig. 5.4 demonstrates that the diffusivity decreases with increasing ε. As the monomers become more attractive it is more difficult for the globule to rearrange

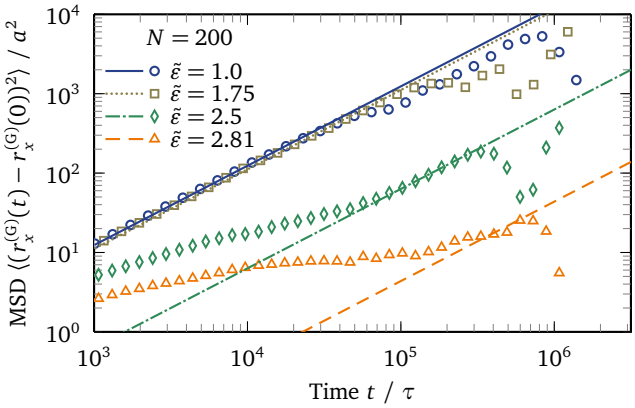


Figure 5.5.: The mean squared displacement (MSD) of the center of the globule for $N = 200$ and $\tilde{\varepsilon} = 1, 1.75, 2.5, 2.81$ is calculated from the trajectories, fig. 5.4, using eq. (5.3.2). Symbols denote the measured MSD from our simulations, lines the corresponding linear fits for $\mathrm{MSD}(t) \gtrsim 10$. For $\tilde{\varepsilon} < 2.5$, the globule exhibits normal diffusion and the diffusivity D_G decreases as ε increases. For $\tilde{\varepsilon} = 2.5$, normal diffusive behavior is almost reached, whereas for $\tilde{\varepsilon} = 2.81$ the diffusion time scale is larger than the simulation time and hence no diffusion is observed.

internally and hence to move. This effect is further accented by the increase in globule size N_G, see fig. 5.3, which also decreases the mobility of the globule.

To quantify these observations we calculate the mean squared displacement (MSD) of $r_x^{(\mathrm{G})}$. For normal diffusive behavior one expects the MSD to scale linearly with time and to be characterized by the diffusion constant D_G

$$\mathrm{MSD}(t) = \left\langle \left(r_x^{(\mathrm{G})}(t) - r_x^{(\mathrm{G})}(0)\right)^2 \right\rangle = \frac{1}{T}\int_0^T \left(r_x^{(\mathrm{G})}(t+t') - r_x^{(\mathrm{G})}(t')\right)^2 \mathrm{d}t' = 2D_\mathrm{G}t \ . \quad (5.15)$$

MSD curves for $N = 200$ and various cohesive strengths are shown in fig. 5.5 on a double logarithmic plot. The MSD curves are fitted with linear functions for $\mathrm{MSD}(t) \gtrsim 10$ in order to obtain the diffusion constant D_G. For small $\varepsilon < \varepsilon_\mathrm{s}$, normal diffusion is observed with D_G decreasing as ε increases. However, as can be seen in fig. 5.5 for $N = 200$ and $\tilde{\varepsilon} = 2.5$, the normal diffusive regime only occurs at very long time scales and for $\tilde{\varepsilon} = 2.81$ is barely reached on the time scales of our simulations. We attribute this to a change of the internal dynamics of the globule. We speculate that, while for small ε reptation is the dominant process for rearranging the globule, this process is suppressed for $\varepsilon > \varepsilon_\mathrm{s}$. Therefore, the only way for the globule to rearrange might be to dissolve – at least partly – and refold into a different configuration. As a consequence the time scale characterizing the internal dynamics for $\varepsilon > \varepsilon_\mathrm{s}$ should become comparable to the time scale on which the globule dissolves. This dissolution time scale is huge as it scales exponentially with εN_G and is beyond our simulation time. For that reason we observe only stuck globules for large cohesive strengths,

which remain in a single conformation. The fitted diffusion constants are shown in fig. 5.6 and compared with the ideal system, where internal friction is absent and the globule and the linker move independently. The diffusivity of the ideal system is estimated by the Rouse diffusion constant of N'_G monomers [201]

$$D_0 = \frac{\mu_0 k_B T}{N'_G} . \tag{5.16}$$

$N'_G \leq N_G$ is the reduced number of monomers that actually have to move on average when the globule is displaced by some distance x; the center of the $N - N'_G$ linker monomers remains stationary. N'_G can be estimated by continuing the linker through the globule and subtracting the number of monomers, which belong to this linker, from N_G, see fig. 5.7 for an illustration,

$$N'_G = N_G - (N - N_G)\frac{l_G}{L - l_G} . \tag{5.17}$$

In fig. 5.6a a pronounced dependence of the diffusion constant D_G on the globule size N_G is observed. The diffusivity decreases as the internal interactions ε increase. This is due to the coupling of the internal dynamics of the globule to the overall motion of the globule, because in our simulation setup the linker is fixed and the globule can only move if the chain reptates through the globule. Slowing down the internal dynamics by increasing ε thus reduces the mobility of the globule. For small ε one observes $D_G \approx D_0$, implying that internal friction is unimportant. However, increasing ε causes increasing deviations between the mobility of the globule and the ideal system until finally D_G drops to zero as the liquid-solid transition is approached. To get rid of the size dependence we show the rescaled diffusivity D_G/D_0 in fig. 5.6b. For $\varepsilon \to 0$ the rescaled diffusivity approaches unity indicating that internal friction is unimportant. The rescaled diffusivity exhibits, within error, no dependence on the size of the globule. Only as ε approaches the solid regime deviations between different system sizes become observable. We conclude that the internal friction is extensive and scales with the size of the globule

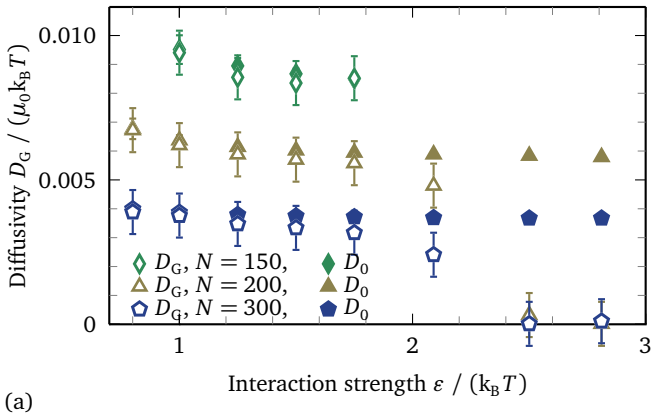
$$D_G \propto 1/N'_G \tag{5.18}$$

as will be corroborated by our non-equilibrium simulations in section 5.4.

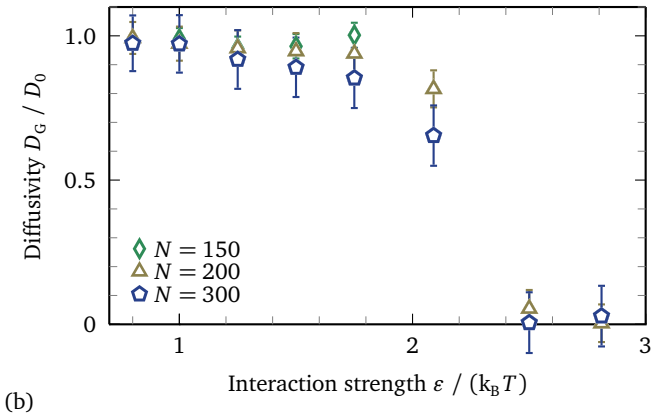
5.4. Forced unfolding of globules: non-equilibrium simulations

5.4.1. Model

We now discuss the results of the previous section in the context of non-equilibrium denaturation of homopolymeric globules [183]. In a distinct set of simulations, the



(a)



(b)

Figure 5.6.: (a) Diffusion constant D_G of the center of the globule (open symbols) as obtained from linear fits to the MSD curves, fig. 5.5. D_G decreases with increasing N and ε. At $\tilde{\varepsilon} = \tilde{\varepsilon}_s \approx 2.5$ a transition of the internal dynamics occurs and the diffusion constant drops to zero. D_G is compared to the ideal Rouse diffusion constant D_0 (solid symbols) of a globule with N'_G monomers, eqs. (5.3.2) and (5.3.2), which can move freely and independent of the linkers. Therefore, D_0 is the diffusion constant of a system where internal friction is absent and does not impede the motion of the globule. For small ε one observes $D_G \approx D_0$ and internal friction is negligible. However, increasing ε slows down the internal dynamics and hence the macroscopic motion of the globule. This causes deviations from the ideal system. (b) Rescaling the diffusion constant of the center of the globule D_G/D_0 removes the N dependence. Only weak dependence on the system size is observable. Therefore the diffusion constant scales as $1/N'_G$ and the internal friction is extensive. By approaching the liquid-solid transition ε_s, a dependence on N_G sets in as ε_s depends on N_G.

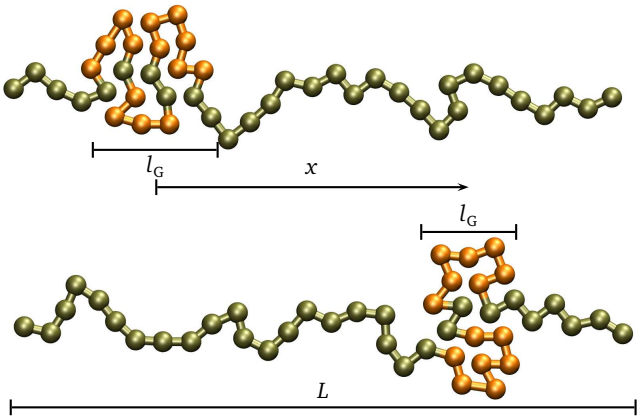


Figure 5.7.: Illustration of the concept and calculation of the reduced number of monomers N'_G in the globule. For a displacement of the globule by a distance x only a reduced number of monomers have to move, here $N'_G = 14$ (shown in orange). Therefore, the linker is continued through the globule and the monomers belonging to this linker are subtracted from N_G yielding N'_G, see eq. (5.3.2).

ends of a globule are trapped at $\boldsymbol{r}^{(1)}(t) = -\boldsymbol{R}(t)$ and $\boldsymbol{r}^{(N)}(t) = \boldsymbol{R}(t)$. The traps are moved at constant speed $R_x(t) = R_x^{(\min)} + vt$, with $R_x^{(\min)} = aN/10$ and $R_{y/z} = 0$, up to a maximal extension $R_x^{(\max)} = a(N-1)$. The force acting on these terminal beads is measured. The backbone bonds are modeled by a harmonic potential U_b, eq. (5.3.1). Excluded volume and cohesive interactions are again modeled by a Lennard-Jones potential U_{LJ}, eq. (5.3.1). Eq. (5.2) is used to integrate the Langevin equation. We no longer employ periodic boundary conditions, however still prevent knot formation by introducing a potential, which mimics two repulsive bars that extend from the first/last bead to the left/right along the x-axis

$$U_k = \begin{cases} \sum_{i=2}^{N-1} \left(\left(2a/\hat{\rho}^{(i)}\right)^{12} - 2\left(2a/\hat{\rho}^{(i)}\right)^{6} + 1 \right) & \text{if } \hat{\rho}^{(i)} < 2a \text{ and } |r_x^{(i)}| > R_x \\ 0 & \text{else,} \end{cases} \tag{5.19}$$

$\hat{\rho}^{(i)} = \sqrt{r_y^{(i)2} + r_z^{(i)2}}$. U_k does not affect the stretching response. Two different protocols are used to obtain the initial configurations to investigate the role of configurational history on the globule pulling response. We record force extension curves for various chain lengths $N = 50, 100, 200, 300$, cohesive strengths $0 \leq \tilde{\varepsilon} \leq 4.1$, and pulling velocities $\tilde{v} = v/(a/\tau) = 0.001, 0.0045, 0.01, 0.0225, 0.045$ going significantly beyond our previous work, where the largest system was $N = 100$, the slowest velocity was $\tilde{v} = 0.0045$, and only un-annealed initial configurations are used. For each parameter set twenty stretching cycles are simulated.

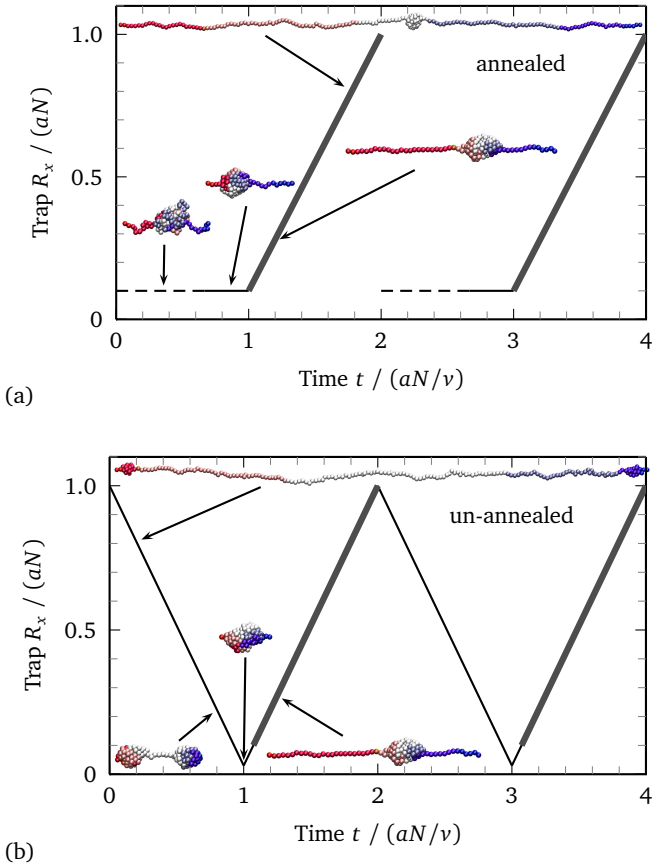


Figure 5.8.: Illustration of the two pulling protocols. (a) Preparation of the annealed structures. After a long equilibrium run with $\tilde{\varepsilon} = 0.8$ and $R_x = 0.1$ (broken line) the globule is equilibrated at the cohesive strength at which the pulling curve is recorded (solid line). The subsequent pulling cycle is depicted by the thick gray line. (b) Preparation of the un-annealed structures. The traps are moved from complete extension to $R_x = 0.03aN$ to ensure that one single globule forms. Without pausing the pulling cycle starts. Force extension curves are recorded for $0.1 < R_x/(aN) < 1$ (thick gray line).

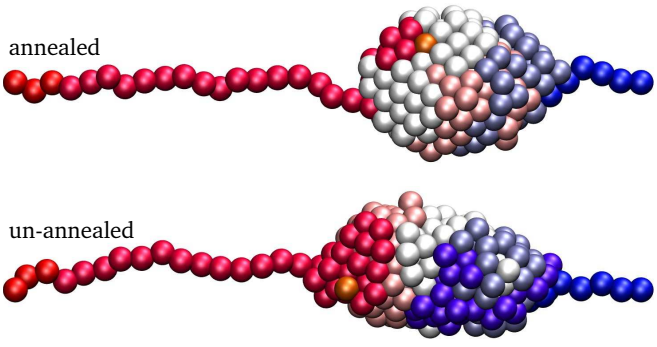


Figure 5.9.: Typical initial configurations for the set of annealed structures (top) and of the un-annealed globules (bottom) for $\tilde{\varepsilon} = 2.91$ and $\tilde{\nu} = 0.001$. Especially for larger ε the lately collapsed, un-annealed structures are rather ordered. The color coding indicates the index of a monomer along a chain.

Annealed initial structures

Annealed initial structures are obtained by performing an equilibrium annealing simulation with cohesive strength $\tilde{\varepsilon} = 0.8$ and fixed trap position $R_x = 0.1aN$. Every $t = 20000\tau$ a structure is recorded, which is subsequently equilibrated for $t = 10000\tau$ using the cohesive strength at which the pulling simulation is to be conducted. The resulting structure is used as one initial configuration for the subsequent pulling cycle. In fig. 5.8a the pulling protocol is illustrated and in fig. 5.9 typical initial configurations are depicted.

Un-annealed initial structures

This set of initial structures is obtained by starting from an extended configuration and moving the traps from $R_x^{(\max)} = a(N-1)$ to $R_x = 0.03aN$. Typically, we observe the formation of one globule for $\tilde{v} < 0.01$ and the formation of two globules near the traps for $\tilde{v} > 0.01$, which merge at small extension, see snapshots in fig. 5.8b. Without pausing, the traps are subsequently extended to $R_x^{(\min)} = 0.1aN$, where the actual pulling cycle starts and force extension curves are recorded. During the preparation the traps are moved with the same velocity v with which the force extension curve is recorded, fig. 5.8b. For larger cohesive strengths ε, non-equilibrium ordered structures prevail as initial configurations for the subsequent pulling cycle, see fig. 5.9 for an illustration. These structures might be of relevance when studying the dynamics and packing of the DNA chromatin structure [166–168].

5.4.2. Liquid-solid transition

In fig. 5.10, averaged stretching curves of twenty pulling cycles for the set of annealed structures are shown for various ε and $\tilde{v} = 0.001$, $N = 300$. Beyond the globule transition, $\tilde{\varepsilon} > \tilde{\varepsilon}_G \approx 0.5$ [25, 183, 193], a ε dependent force plateau is observed. The plateau force F_p increases as ε increases and is associated with the equilibrium free energy per unit length of globule formation, $F_p a \propto \varepsilon - \varepsilon_G$ [183]. For larger extensions $x/L_c \approx 0.8$ a dip in the force extension curve appears, which is the signature of two-state behavior characteristic of globule dissolution [159]. Once the globule is disrupted, the force extension curve becomes independent of the cohesive strength ε and follows the trace of an extensible freely jointed chain (broken line)

$$x/L_c = \coth(2aF/(k_B T)) + k_B T/(2aF) + F/(2a\kappa)\,. \tag{5.20}$$

A phantom chain ($\tilde{\varepsilon} = 0$) coincides perfectly with eq. (5.4.2) indicating that pulling happens under equilibrium conditions. The curve for $\tilde{\varepsilon} = 2.5$ features a maximum at $x/L_c \approx 0.3$, whose origin will be discussed in the next paragraph.

Fig. 5.11 shows force extension traces of the two sets of initial configurations for $N = 300$, $\tilde{v} = 0.001$, and strong cohesive forces $\tilde{\varepsilon} = 2.08, 2.5, 2.91$. The thick curves depict the average of twenty pulling curves. The thin curves are examples

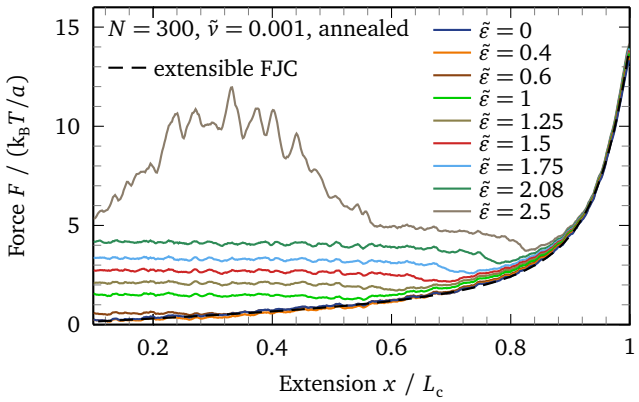


Figure 5.10.: Force extension curves are shown for constant velocity $\tilde{v} = 0.001$, various cohesive strengths, $N = 300$, and annealed initial configurations. The curves are averages of twenty pulling cycles. Above the globule transition $\tilde{\varepsilon} > \tilde{\varepsilon}_G \approx 0.5$ a force plateau followed by a dip in the force extension curve is observed. These two features indicate the existence of a globule at small extensions and the dissolution of the globule at larger extensions. The broken line depicts the theoretically expected force extension trace of an extensible freely jointed chain for $\tilde{\varepsilon} = 0$, eq. (5.4.2). For $\tilde{\varepsilon} = 2.5$ the averaged pulling curve exhibits a marked maximum at small extensions due to large fluctuations in the individual force extension curves, which indicate the onset slow internal dynamics, see fig. 5.11, where averaged and individual force extension curves for $\tilde{\varepsilon} > 2$ are shown.

of individual curves for a given parameter set and different initial configurations. In figs. 5.11a-c pulling curves starting from the set of un-annealed configurations are shown. Increasing ε leads to an increasing plateau force as already observed [183], see fig. 5.10. For $\tilde{\varepsilon} \geq 2.5$ some stronger fluctuations of the force are observed, but the force extension traces are qualitatively similar to the curves for $\tilde{\varepsilon} \leq 2.08$. The situation is vastly different for pulling curves starting from the set of annealed initial configurations, figs. 5.11d-f. Again, for small cohesive strengths $\tilde{\varepsilon} \leq 2.08$ the pulling curves are smooth and no strong fluctuations occur, cf. fig. 5.10. However, increasing the cohesive strength further leads to pronounced fluctuations in the force extension curve. This is due to a transition of the internal dynamics from liquid-like to solid-like [7, 183, 195, 198–200, 205]. Since the un-annealed initial configurations are rather ordered – especially for large ε – the globules are easily unwound without inducing the pronounced fluctuations, which are observed for the annealed, rather disordered initial configurations and $\varepsilon > \varepsilon_s$. The liquid-solid transition occurs for the un-annealed structures, too, but has no signature in the pulling setup.

The dependence of the liquid-solid transition on the globule size N_G is illustrated in fig. 5.12. There, the force is plotted versus the number of monomers N_G inside the globule. N_G is calculated via eq. (5.3.2) using the extension x of the polymer instead of L. One notices that the huge fluctuations cease once the globule is below a certain size, which depends on ε. This feature is independent of the chain length as curves with

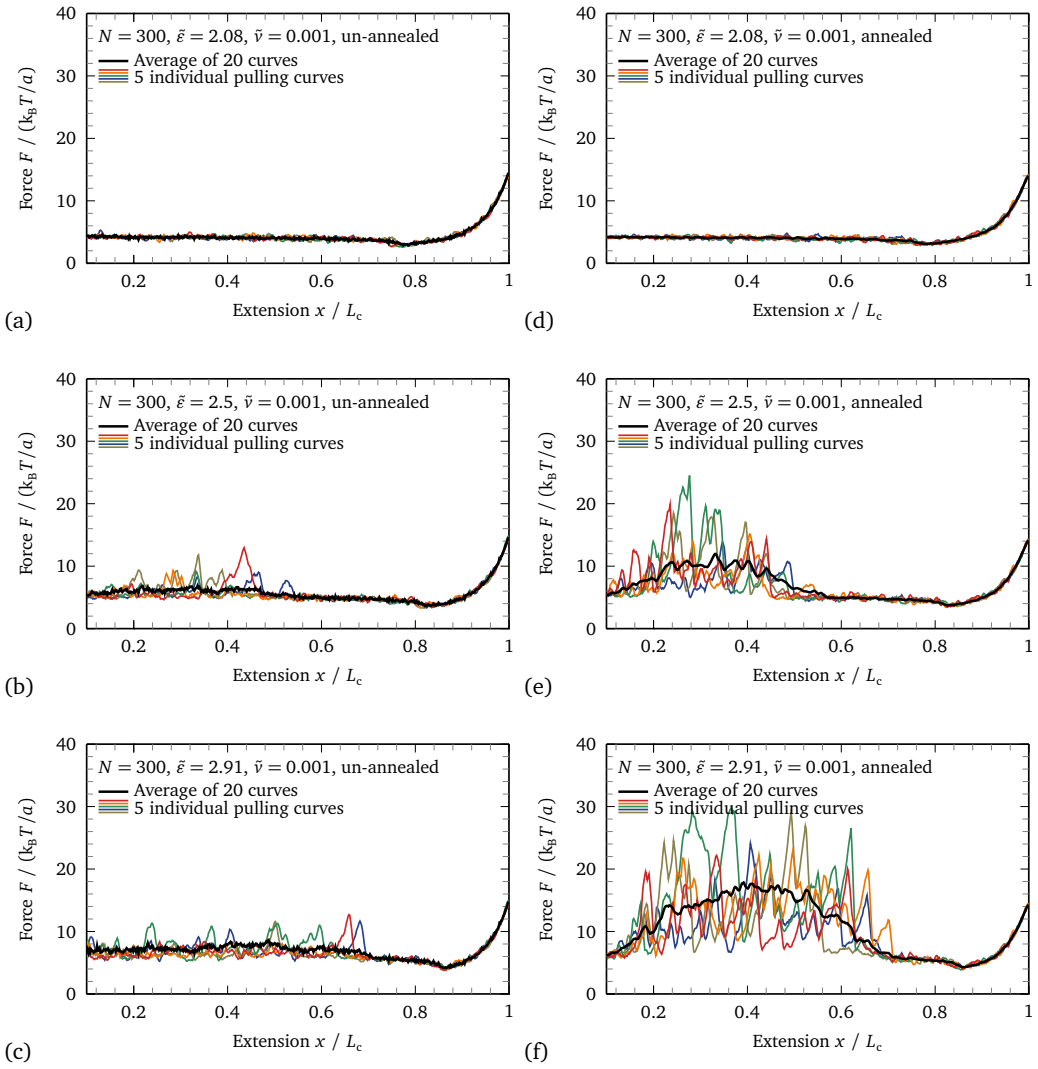


Figure 5.11.: Typical pulling curves for $N = 300$ and $\tilde{v} = 0.001$ with (a-c) un-annealed and (d-f) annealed structures as initial configurations. The thin lines are individual force extension traces, whereas the thick black line is the average of twenty pulling curves. The un-annealed structures exhibit rather smooth pulling curves and no qualitative difference between different cohesive strengths ε is observed. However, the annealed structures feature marked fluctuations above a certain threshold ε_s, which cease as the globule size decreases. We attribute these fluctuations to a liquid-solid transition, which dramatically changes the internal dynamics. This liquid-solid transition occurs in the un-annealed structures, too, but shows no clear signature in the pulling curves as those structures are rather ordered and hence are easily unwound.

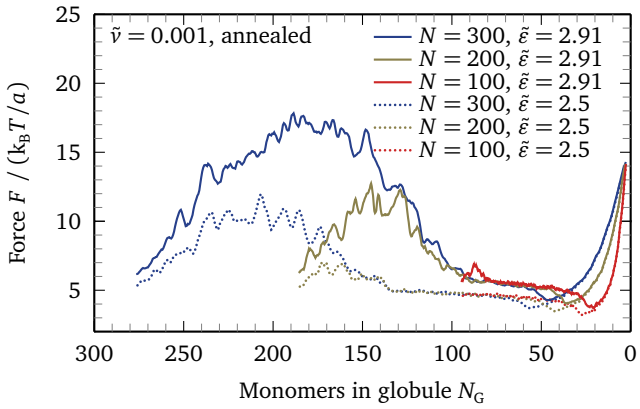


Figure 5.12.: The internal dynamics and the position of the liquid-solid transition depend on the number of monomers N_G, eq. (5.3.2), inside the globule and the cohesive strength ε. Here, averaged force extension curves in the vicinity of the liquid-solid transition are shown for the annealed set of initial configurations. The solid state and the slow internal dynamics are recognized as huge fluctuations in the pulling force, which lead to the peak in the averaged curve. Once N_G is below a certain threshold (≈ 100 for $\tilde{\varepsilon} = 2.91$, ≈ 140 for $\tilde{\varepsilon} = 2.5$) the fluctuations cease and the globule is driven into the liquid state, which is recognized as the collapse of the curves at the transition for different N and equal ε. The dependence of the liquid-solid transition on N_G will be shown more explicitly in ref. [7].

equal ε but different N coincide at the transition, but solely depends on N_G [7, 193–200]. The cohesive strength ε_s at which this liquid-solid transition occurs depends on the size of the globule N_G, but is independent of the pulling velocity within the range of velocities studied. In fig. 5.13 we see that even with the highest pulling velocity no large fluctuations in the force extension curve are induced and the globule remains in the liquid phase for $\tilde{\varepsilon} = 2.08$. Therefore, it is not a mere non-equilibrium pulling feature but an indication of a change in the internal dynamics. The transition observed here is the same transition that induces the abrupt change of the globule diffusivity, fig. 5.6, since both the equilibrium globule diffusion and the non-equilibrium pulling simulations are ultimately related to the time scale describing the motion of monomers inside the globule.

5.4.3. Internal friction

Dissipated energy scales linearly with the velocity below the liquid-solid transition

For small $\varepsilon < \varepsilon_s$ the internal dynamics are fast and monomers are mobile inside the globule. Therefore, no significant difference between the two different sets of initial configurations is observed.

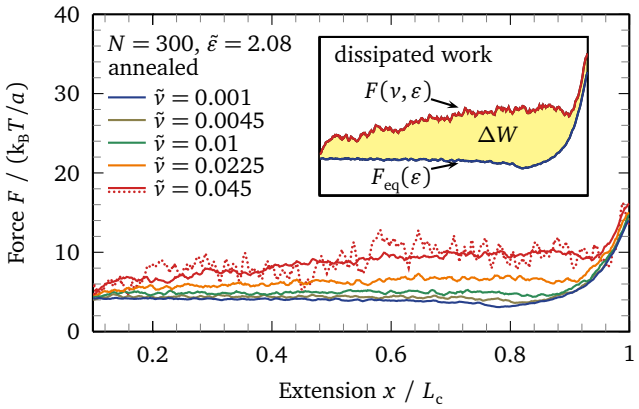


Figure 5.13.: Averaged pulling curves for various pulling velocities with $N = 300$, $\tilde{\varepsilon} = 2.08$, and annealed initial configurations. The friction force increases with increasing pulling velocity. The curves with the slowest pulling velocities $\tilde{v} = 0.001, 0.0045$ almost coincide, indicating that one has reached nearly the equilibrium pulling limit. The dotted curve is one individual pulling curve for $\tilde{v} = 0.045$. The inset illustrates the definition of the dissipated work, eq. (5.4.3), which is the shaded area between the two curves.

In fig. 5.13 pulling curves for various pulling velocities v are shown. We observe that even for a cohesive strength as large as $\tilde{\varepsilon} = 2.08$ and polymers as long as $N = 300$, the two slowest pulling curves $\tilde{v} = 0.001, 0.0045$ almost coincide. This indicates that force extension curves for $\tilde{v} = 0.001$ are already very good approximations for equilibrium curves. Increasing velocity leads to increasing energy dissipation. There are two major mechanisms leading to dissipation: solvent friction and internal friction. These dissipation mechanisms dominate at different parts of the pulling curve. For small extensions, when most of the monomers are part of the globule, internal friction dominates. Towards the end of the curve, the globule is markedly smaller and the solvent friction term dominates the pulling curve. The dissipated work $\Delta W(v, \varepsilon)$ is defined as [183]

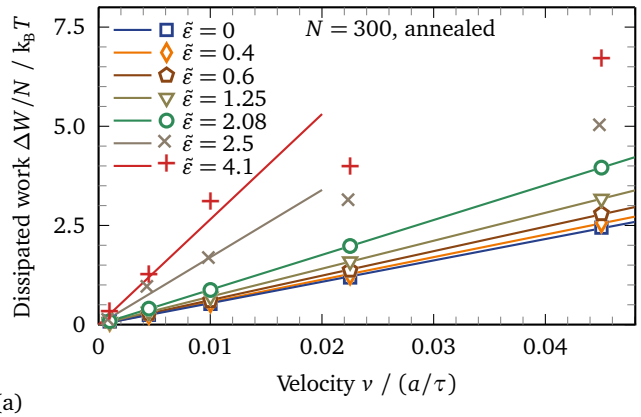
$$\Delta W(v, \varepsilon) = W(v, \varepsilon) - W_{eq}(\varepsilon)\,, \tag{5.21}$$

where the work done by one trap is generally defined as

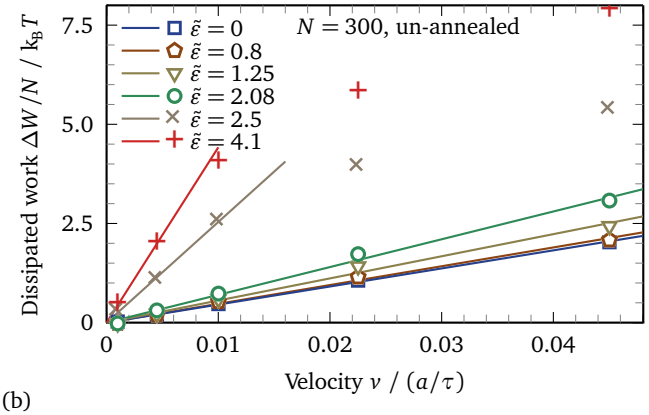
$$W = -\int_{R_x^{(min)}}^{R_x^{(max)}} F(x)\mathrm{d}x\,, \tag{5.22}$$

see inset of fig. 5.13 for an illustration. The equilibrium work $W_{eq}(\varepsilon)$ is obtained from extrapolating $W(v, \varepsilon)$ to $v \to 0$.

In fig. 5.14 we show the dissipated work ΔW as a function of the velocity v for $N = 300$ and various ε obtained from the pulling curves. Below the liquid-solid transition $\tilde{\varepsilon} < \tilde{\varepsilon}_s \approx 2.5$ the dissipated work scales linearly with v, which is elucidated further



(a)

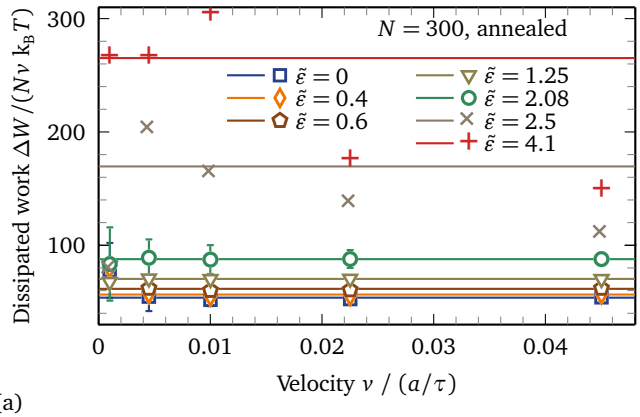


(b)

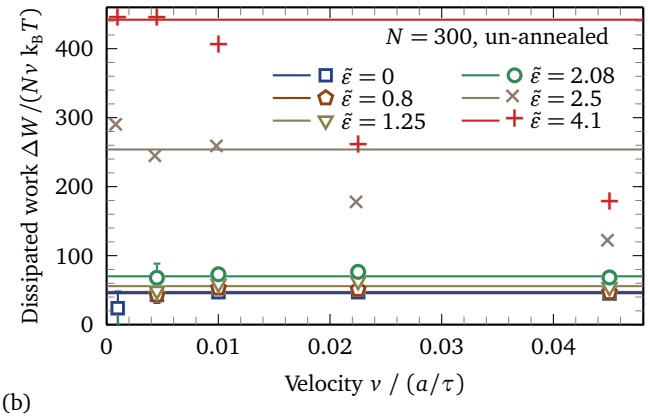
Figure 5.14.: Dissipated work per monomer $\Delta W/N$ as a function of the pulling velocity v for $N = 300$ and different cohesive strengths ε starting from the (a) annealed and the (b) un-annealed set of initial configurations. Symbols depict simulation data, lines show linear fits to the data, see eq. (5.4.3). Below the globule transition, $\tilde{\varepsilon} < \tilde{\varepsilon}_G \approx 0.5$, the curves collapse and the attractive forces become almost negligible. Below the liquid-solid transition, $\tilde{\varepsilon} < \tilde{\varepsilon}_s \approx 2.5$, our simulations are carried out in the linear response regime as the curves are linear in v in the whole range of velocities studied. Above the liquid-solid transition the linear scaling breaks down. ΔW is slightly lower for the simulations starting from the un-annealed set of initial configurations as the globule might not be completely equilibrated and still rather ordered. In fig. 5.16 the fluctuations of the dissipated work are depicted.

in fig. 5.15, where $\Delta W/v$ is plotted. Therefore, our simulations are conducted in the experimentally relevant linear response regime. Above the liquid-solid transition this scaling breaks down. Further, we observe that below the globule transition, $\tilde{\varepsilon} < \tilde{\varepsilon}_G \approx 0.5$, the dissipated work is almost independent of the cohesive strength. For the curves starting from the un-annealed set of initial configurations, fig. 5.15b, the dissipated work is virtually independent of the cohesive strength as long as $\tilde{\varepsilon} \leq 0.8$. As curves with $\tilde{\varepsilon} = 0$ (phantom chain) and $0 < \tilde{\varepsilon} \leq 0.8$ (with self avoidance) coincide, monomer-monomer attraction and topological constraints, e. g. entanglements, are negligible for small values of the cohesive strength. This effect is less pronounced in the simulations starting with the annealed set of initial configurations, where topology starts to play a role even for small ε. Knots, which might arise due to bond crossing, are not observed in our simulations. The fluctuations of the dissipated work reveal again the liquid-solid transition. In fig. 5.16 the sample standard deviation

$$\sigma_{\Delta W} = \sqrt{k/(k-1)}\,\sqrt{\langle \Delta W^2 - \langle \Delta W \rangle^2 \rangle} \tag{5.23}$$



(a)



(b)

Figure 5.15.: Dissipated work per monomer rescaled with pulling velocity, $\Delta W/(Nv)$, as a function of the pulling velocity v for $N = 300$ and different cohesive strengths ε starting from the (a) annealed and the (b) un-annealed set of initial configurations. Symbols denote simulation results, lines depict $\tilde{\Gamma}$ as obtained from linear fits, see eq. (5.4.3), to the data, see fig. 5.14. For $\varepsilon < \varepsilon_s$, our simulations are carried out in the linear response regime as there is no v dependence after rescaling. Above the liquid-solid transition fluctuations are huge, the linear scaling in v breaks down, and significant deviations from the linear fit occur. For $\tilde{\varepsilon} = 0$, 2.08 error bars indicate the standard deviations of the twenty measurements of $\Delta W(v, \varepsilon)$. The fluctuations of the work necessary to stretch the polymer are small below the liquid-solid transition and huge above; see fig. 5.16, where the standard deviations are plotted.

of $k = 20$ measurements of ΔW for $\tilde{v} = 0.001$ is plotted against ε. For $\varepsilon < \varepsilon_s$, $\sigma_{\Delta W}/N$ is independent of N and ε. As one crosses from the liquid into the solid phase at $\varepsilon = \varepsilon_s$ the fluctuations increase. Fig. 5.16 clearly demonstrates that ε_s depends on the system size N; explicitly, $\tilde{\varepsilon}_s \approx 4$ for $N = 50$, $\tilde{\varepsilon}_s \approx 2.9$ for $N = 100$, $\tilde{\varepsilon}_s \approx 2.3$ for $N = 200$, and $\tilde{\varepsilon}_s \approx 2.1$ for $N = 300$.

Friction force scales linearly with N and friction is extensive

We already observed that the internal friction and the dissipated work ΔW scale linearly with the velocity v. Therefore, we extract the velocity dependence and express the friction force F_{fr} in terms of the friction coefficient $\Gamma(\varepsilon, N)$

$$F_{fr}(v, \varepsilon, N) = F(v, \varepsilon, N) - F_{eq}(\varepsilon, N) = v\, \Gamma(\varepsilon, N)\,, \tag{5.24}$$

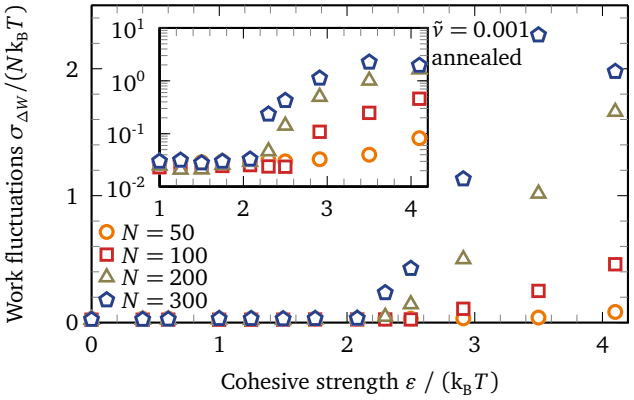


Figure 5.16.: The standard deviation per monomer $\sigma_{\Delta W}/N$ of the dissipated work, see fig. 5.14, as obtained from twenty pulling cycles increases abruptly at the liquid-solid transition. For $\varepsilon < \varepsilon_s$, $\sigma_{\Delta W}/N$ is independent of the chain length N and small. The fluctuations of the dissipated work increases significantly when $\varepsilon > \varepsilon_s$, where ε_s decreases for increasing N. The data is obtained from simulations with $\tilde{\nu} = 0.001$ and annealed initial configurations.

where F_{eq} is the equilibrium force extension curve. Analogous to the Stokes friction of a sphere, we define the friction coefficient to scale like [183]

$$\Gamma(\varepsilon, N) \propto \eta_G(\varepsilon) a N_G^{\gamma} , \tag{5.25}$$

where $\eta_G(\varepsilon)$ is the internal viscosity, which depends on ε. The exponent γ describes the friction mechanism at work during unraveling the globule. Two limits can be distinguished: first, if $\gamma = 0$ the friction force is independent of the globule size N_G and only a finite number of monomers, which does not scale with N_G, contribute to dissipation. We term this case local, intensive friction. This scenario does not seem compatible with a reptation mechanism, where the entire chain would have to move through the globule. Second, $\gamma = 1$ describes the situation, where a finite fraction of the globule, that is proportional to N_G, or even the entire chain rearranges and hence contributes to the friction force. We term this case global, extensive friction and this scenario would be equivalent to reptation. Integrating the friction force, eq. (5.4.3), yields the scaling of the dissipated work

$$\Delta W(v, \varepsilon) = -\int F_{fr}(v, \varepsilon, N) \mathrm{d}x \propto \eta_G a^2 N^{\gamma+1} v , \tag{5.26}$$

where we assume $N_G \approx (L_c - 2R_x)/(2a)$. In fig. 5.17 $\Delta W/(N^2 v)$ is plotted and all curves with equal $\varepsilon < \varepsilon_s$ coincide. This implies that the dissipated energy per monomer $\Delta W/N$ is extensive, $\gamma = 1$, meaning that a finite fraction of the globule rearranges during pulling and not only the few monomers, which are about to be pulled out of the globule. Again, this scaling behavior breaks down above the liquid-solid transition and beyond the linear response regime.

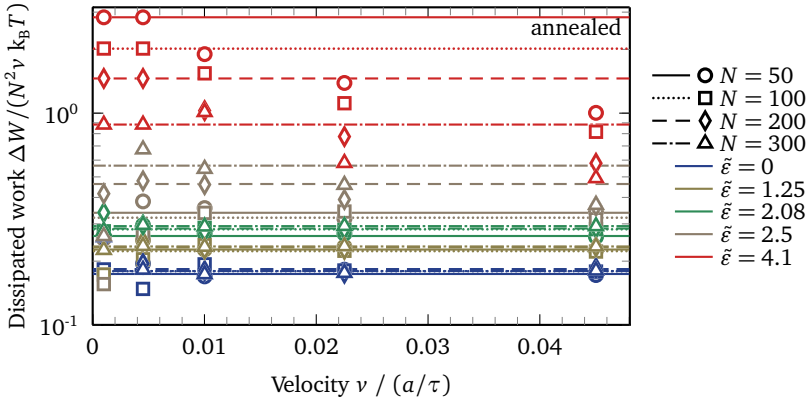


Figure 5.17.: Rescaled dissipated work per monomer $\Delta W/(N^2 v)$ as a function of the pulling velocity v for different N and cohesive strengths ε starting from the annealed set of initial configurations. Curves with equal ε and different N collapse for $\varepsilon < \varepsilon_s$. This shows that the dissipated work per monomer is an extensive function and $\Delta W \propto N^2 v$. Again, this scaling behavior breaks down above the liquid-solid transition.

Internal viscosity

To compare the internal viscosity quantitatively we fit functions linear in v to the dissipated work,

$$\Delta W(v,\varepsilon) = \tilde{\Gamma}(\varepsilon,N)\, v\,, \tag{5.27}$$

with $\tilde{\Gamma}(\varepsilon,N) \propto \eta_G(\varepsilon) N^{\gamma+1}$. As can be seen in fig. 5.14, $\Delta W \propto v$ for all velocities studied if $\varepsilon < \varepsilon_s$. Consequently we include all velocity data to fit the linear function. However, for $\varepsilon > \varepsilon_s$ a marked deviation from linear behavior is observed. There, we fit only to the slowest velocities, where the data still scales linearly with v. The relative excess viscosity

$$\frac{\eta_G(\varepsilon)}{\eta_G(0)} - 1 = \frac{\tilde{\Gamma}(\varepsilon,N)}{\tilde{\Gamma}(0,N)} - 1 = \frac{\Delta W(\varepsilon)}{\Delta W(0)} - 1 \tag{5.28}$$

is shown in fig. 5.18 for annealed and un-annealed initial configurations. This way the numerical prefactors in eq. (5.4.3) are eliminated and we are able to compare all different sets of parameters. The excess viscosities coincide for different N, which indicates that all $\tilde{\Gamma}(\varepsilon,N)$ exhibit the same N dependence. In fig. 5.18a we additionally plot the excess viscosities as obtained from the equilibrium globule diffusion simulations $\eta_G(\varepsilon)/\eta_G(0) - 1 = D_0/D_G(\varepsilon) - 1$, see fig. 5.6b. The excess viscosity from the equilibrium simulations is slightly lower which might be due to underestimating D_0, eq. (5.3.2), or additional dissipative mechanisms in the non-equilibrium pulling simulations: sometimes it may happen that the whole globule is pulled through the solvent as an entanglement is not resolved quickly enough, which causes additional dissipation. We also compare our data to the effective friction constant of a particle diffusing

in a periodic potential $U_\mathrm{p} = \theta\varepsilon/2\cos(\pi x/(2a))$ with amplitude $\theta\varepsilon$. θ is a scaling factor for the amplitude of the corrugated and may be viewed as a fitting parameter. This potential mimics the energy landscape the globule experiences if it is translated along the x-axis. The solution of this one-dimensional diffusion problem yields an effective viscosity [181]

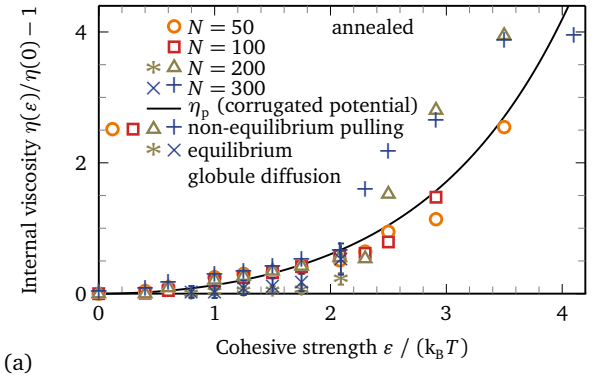
$$\frac{\eta_\mathrm{p}(\varepsilon)}{\eta(0)} = \mathrm{I}_0^2\left(\frac{\theta\varepsilon}{2\mathrm{k_B}T}\right) \tag{5.29}$$

for times larger than the typical time during which the particle is trapped in one minimum, $t \gg \tau\exp(\theta\varepsilon/(\mathrm{k_B}T))$. $\mathrm{I}_0(z)$ is the zeroth order modified Bessel function with the asymptotic limits $\mathrm{I}_0(z) \sim 1+z^2/4$ for $z \ll 1$ and $\mathrm{I}_0(z) \sim \mathrm{e}^z/\sqrt{2\pi z}$ for $z \gg 1$ [115]. As can be seen from figs. 5.18 the rescaled internal viscosity, eq. (5.4.3), is reproduced very well by the simple scaling prediction eq. (5.4.3) for $\theta = 1$. Our results compare excellently to our previous results where smaller globules $N \leq 100$ have been considered [183].

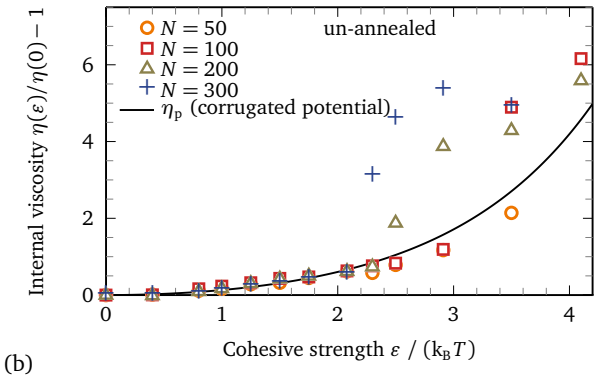
In summary, we have observed two different dynamical regimes: a solid-like regime, characterized by huge fluctuations in the force extension trace, and a liquid-like regime. The position ε_s of the transition between these two regimes depends on the size of the globule N_G and the cohesive strength ε, cf. figs. 5.12 and 5.16. In the liquid regime we have shown that the dissipated work per monomer is extensive and a finite fraction of the globule contributes to the internal friction, fig. 5.17. Further, by rescaling the internal viscosity we showed that the dependence of the internal friction on the cohesive strength ε is described very well by the diffusion of a single particle in a corrugated potential. We extended our previous results [183] to significantly larger systems and are able to show that – below the liquid-solid transition – the history, i. e. the preparation of the initial structures, does not influence the scaling results, see fig. 5.18.

5.5. Conclusions

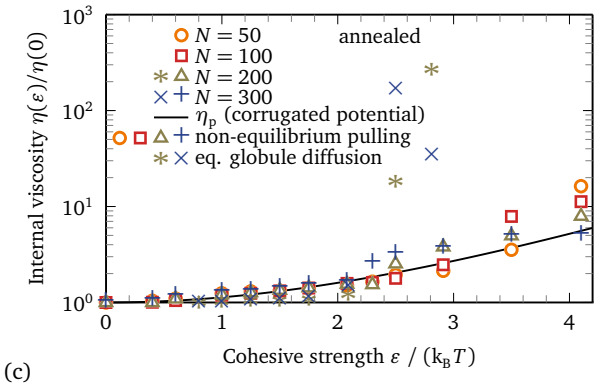
We have shown that two dynamic regimes exist for a collapsed homopolymer depending on the cohesive strength ε and the size of the globule N_G. For small $\varepsilon < \varepsilon_\mathrm{s}$ and small N_G the monomers inside the globule are rather mobile and the chain is in a liquid state. The internal friction is studied by considering the diffusion of a globule along an extended chain and obtaining the dissipated work from non-equilibrium stretching simulations in the linear response regime. We find that the internal friction is extensive and thus scales linearly with N_G. The internal friction increases as ε increases until the liquid-solid transition is reached. The signature of the solid state is $D_\mathrm{G} = 0$ for the equilibrium globule diffusion setup and huge fluctuations in the force extension curves for the non-equilibrium pulling setup. The solid state is characterized by very slow internal dynamics. This has implications for protein folding, since our results suggest that a quick collapse of a large hydrophobic core of the protein may result in a kinetically trapped and misfolded protein, which would not be able to fold into



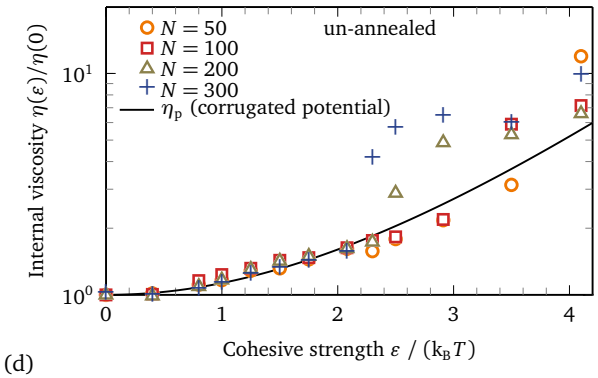
(a)



(b)



(c)



(d)

Figure 5.18.: Rescaled excess viscosity $\eta_G(\varepsilon)/\eta_G(0) - 1$, eq. (5.4.3), as a function of the cohesive strength ε for (a), (c) annealed (b), (d) un-annealed initial configurations. The viscosity $\eta_G(\varepsilon)$ is obtained from linear fits to the dissipated work as a function of the velocity v. As the data points for different N coincide, all $\tilde{\Gamma}(\varepsilon, N)$ exhibit the same N dependence. The solid line is the prediction of the excess viscosity of a Brownian particle in a sinusoidal potential, see eq. (5.4.3). In (a), (c) the excess viscosity as obtained from our equilibrium simulations, see fig. 5.6b, $\eta_G(\varepsilon)/\eta_G(0) - 1 = D_0/D_G(\varepsilon) - 1$ is shown. The liquid-solid transition is recognized as the deviation of the simulation results from the simple scaling relation eq. (5.4.3).

the native structure due to very slow internal dynamics for too large cohesive energies. The system can be driven into the liquid state by removing monomers from the globule through chain extension.

CHAPTER 6

SUMMARY AND OUTLOOK

In this work the statistical mechanics of nucleic acids and bio-polymers have been studied. We demonstrated that coarse graining is an excellent tool both to study the thermodynamics of these systems and to improve theories, on which many biotechnological applications rely.

In chapter 2, we introduced a three-state model for double stranded DNA under tension that distinguishes between native B-DNA, overstretched S-DNA, and molten M-segments. At the same time it correctly accounts for the entropy of molten loops, characterized by the exponent $\hat{c}$ in the asymptotic expression $S \sim -\hat{c} \ln m$ for the entropy of a inter-strand loop of length m. Force extension curves were exactly derived employing a generalized Poland-Scheraga approach and compared to experimental data. We simultaneously fitted to force-extension data at room temperature and to the denaturation phase transition at zero force and were able to establish a global phase diagram in the force-temperature plane. Under a stretching force, the effects of the stacking energy, entering as a domain-wall energy between paired and unpaired bases, and the loop entropy are disentangled. Therefore we could estimate the loop exponent $\hat{c}$ independently from the precise value of the stacking energy. The fitted value for $\hat{c}$ turned out to be small, suggesting that nicks dominate the experimental force extension traces of natural DNA.

In chapter 3 we developed a theory for secondary structures in single stranded homopolymeric nucleic acids that accounts for the logarithmic entropy $-c \ln m$ of loops within one strand and allows to study chains under external force. In the thermodynamic limit of a long strand, the chain displays a phase transition between a low temperature / low force compact (folded) structure and a high temperature / high force molten (unfolded) structure. The influence of c on phase diagrams, critical exponents, melting, and force extension curves was derived analytically. For vanishing pulling force, only for the limited range of loop exponents $2 < c \lesssim 2.479$ a melting transition is possible; for $c \leq 2$ the chain is always in the folded phase and for $c \gtrsim 2.479$ always in the unfolded phase. A force induced melting transition with singular behavior is possible for all loop exponents $c \lesssim 2.479$ and can be observed

experimentally by single molecule force spectroscopy. We discussed the influence of these findings on the hybridization or denaturation of double stranded nucleic acids. The Poland-Scheraga model for nucleic acid duplex melting does not allow base pairing between nucleotides on the same strand in denatured regions of the double strand. If the sequence allows these intra-strand base pairs, we showed that for a realistic loop exponent $c \approx 2.1$ pronounced secondary structures appear inside the single strands. This leads to a higher melting temperature of the duplex than predicted by the Poland-Scheraga model. Further, these secondary structures renormalize the effective loop exponent $\hat{c}$, which characterizes the weight of a denatured region of the double strand.

In chapter 4 the theory of the previous chapter has been extended to correctly account for sequence and salt effects. We calculated the folding free energies and minimum free energy configurations for real sequences. Salt effects entered in two ways: first, we derived salt induced modifications of the free energy parameters for describing base pairing and, second, we included the electrostatic free energy for loop formation. Both effects were modeled on the Debye-Hückel level including counterion condensation. We validated our theory for two different RNA sequences: For tRNA-phe, the resultant heat capacity curves for thermal denaturation at various salt concentrations accurately reproduce experimental results. For the P5ab RNA hairpin, we derived the global phase diagram in the three-dimensional space spanned by temperature, stretching force, and salt concentration. Excellent agreement with the experimentally determined critical unfolding force is obtained. We showed that for a proper description of RNA melting and stretching, both salt and loop entropy effects are needed.

In chapter 5 we studied the reptational dynamics of polymer globules by Brownian dynamics simulations. Equilibrium simulations, where we investigated the diffusional reptation in a globule, were compared with non-equilibrium simulations, where we unfolded the globule by pulling the ends with prescribed velocity. A strong dependence of the internal dynamics on the Lennard-Jones interaction strength ε and the globule size N_G is observed. We found two distinct dynamical regimes: a liquid-like regime with fast internal dynamics and a solid-like regime with slow internal dynamics. The position $\varepsilon_s \approx 2.5 k_B T$ of the transition depends on N_G. We characterized these regimes using diffusion and pulling simulations to explore each regime's internal friction characteristics. In the liquid-like regime ($\varepsilon < \varepsilon_s$) a moderate dependence on ε is observed and the dissipation exhibits extensive scaling $\propto N_G$. The solid regime is characterized by slow internal dynamics suppressing diffusion of the globule along the chain and inducing huge fluctuations in the force extension curve.

As always, answers to questions evoke other – new and interesting – questions. We saw that secondary structure formation inside denatured regions of double stranded nucleic acids plays an important role and should exhibit marked sequence dependence. The next logical step would be to extend existing theories for DNA melting – such as the Poland-Scheraga model [27, 28] or the Peyrard-Bishop model [51] – to allow for intra-strand base pairing. Experimentally the effect of intra-strand secondary structure

formation might be studied with the setup of Léger et al. [35], who were able to apply both torque and stretching force on DNA molecules.

The prediction of RNA structures from sequences and the design of sequences yielding specific structures is an active field of research [110, 119]. The theory presented in this work relies on hierarchical folding and on the secondary structure of RNA being independent of the tertiary structure. Although this is a very good starting point, it is the interplay of secondary and tertiary structure, which ultimately shapes the molecule. Therefore future work should include this by allowing for pseudoknots [206], specific salt interactions [143], or by combining coarse grained secondary structure prediction with all-atom molecular dynamics simulations.

Knowledge on the internal dynamics of proteins, nucleic acids, and other bio-polymers is essential for understanding folding or processes, which involve structural reorganization of molecules. Through our Brownian dynamics simulations we demonstrated that force extension curves depend strongly on the – annealed or un-annealed – structure of the polymer globule. The influence of structure may also be studied by introducing specific interactions between monomers, which would compel the polymer to assume a certain conformation. The dynamics inside solid-like globules have been shown to be fundamentally different from the dynamics of liquid-like globules. Therefore, it would be interesting to further explore the mechanisms, which lead to this difference, how they influence bio-polymer systems, and how nature circumvents the restrictions imposed by the slow dynamics.

The statistical mechanics of bio-polymers and bio-physics in general is a rich field of research, where theorists and experimentalists closely interact. Theoreticians are constantly devising new models, improving existing techniques, and employing perpetually increasing simulation system sizes. Theoretical physics enables a close up view of nature and of the mechanisms at work, which might not yet be accessible experimentally. On the other hand, experimentalists are able to grab and manipulate single molecules, resolve molecular structures on an Ångström scale, and observe dynamics faster than picoseconds. Experiments are our enhanced eyes, ears, and hands, with which we can experience and manipulate the world, test theories, and discover new phenomena.

This thesis is my contribution to this inspiring scientific community.

Chapter 7

Danksagung

An dieser Stelle möchte ich den Menschen danken, ohne die diese Arbeit nicht möglich gewesen wäre. Mein besonderer Dank gilt meinem Doktorvater Herrn Prof. Roland Netz, von dem ich viel lernen durfte, der mir große Freiheit bei meiner Forschung ließ und mir neben zahlreichen Konferenzteilnahmen auch Aufenthalte in Santa Barbara, Paris und Istanbul ermöglichte. Vor allem werde ich aber auch die vielen gemeinsamen Unternehmungen des Lehrstuhls – Winterschulen, Sommerfeste, Wandertage und Weihnachtsfeiern – in bester Erinnerung behalten. Ich bedanke mich ganz herzlich bei unserer Sekretärin, Frau Sonja Ortner, der guten Seele des Lehrstuhls, die den Laden am Laufen hält.

Auf keine Fall unerwähnt bleiben dürfen meine beiden Bürokollegen, die ich im Laufe der Zeit verschlissen habe – Herr Dr. Sebastian Fischer und Herr Yann von Hansen. Ich bedanke mich für die wunderbare Zeit, die unzähligen vollgeschmierten Tafeln und die intensiven Diskussionen. Herrn Prof. Dominik Horinek und Herrn Felix Sedlmeier danke ich unter anderem für die Administration des Rechnerclusters, dem Dank ihrer Arbeit nur fakultätsweite Stromausfälle oder kaputte Klimaanlagen etwas anhaben konnten. Dankeschön an Frau Dr. Maria Fyta, die mich neben Yann und Felix auf den letzten Metern zur Anfertigung dieses Manuskripts unterstützt hat. Ich möchte mich beim kompletten Lehrstuhl T37 bedanken: für die tolle Atmosphäre, die vielen gemeinsamen Erlebnisse sowie die zahlreichen “fachlichen” Auseinandersetzungen, die häufig am Kicker oder beim BZFlagspielen endeten. Dank geht auch an Herrn Prof. Andreas Bausch für die stete Unterstützung und ohne den ich heute nicht Biophysik machen würde.

Für die erfolgreiche Zusammenarbeit bei Teilprojekten dieser Dissertation bedanke ich mich bei Herrn Prof. Hans-Jürgen Kreuzer und Herrn Douglas Staple (DNA), Herrn Prof. Alfredo Alexander-Katz und Herrn Charles Sing (Polymerglobule) sowie Herrn Paul Näger und Herrn Prof. Henri Orland (RNA), dem ich auch für den Aufenthalt an seinem Institut danke.

Ich bedanke mich beim Elitenetzwerk Bayern für die großzügige Förderung im Rahmen des Doktorandenkollegs CompInt als auch beim Leibniz-Rechenzentrum für die zur Verfügung gestellte Rechenzeit.

Die letzten, viel zu kurzen, Zeilen gehören meiner Familie – meiner Frau Kathrin, meinen Eltern Brigitte und Fritz, meinem Bruder Stefan und seiner Frau Aline: Vergelt’s Gott für Euren Rückhalt und alles was ich von Euch erfahren und lernen durfte. Ohne Euch gäbe es diese Arbeit vermutlich nicht.

Parts of this thesis have been published in or submitted to peer-reviewed journals, or manuscripts for such contributions are in preparation:

Thomas R. Einert, Paul Näger, Henri Orland, and Roland R. Netz. Impact of loop statistics on the thermodynamics of RNA folding. *Physical Review Letters*, 101(4): 048103, 2008.

Thomas R. Einert, Paul Näger, Henri Orland, and Roland R. Netz. Loop parameterization and RNA secondary structure folding. In Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty, Walter Nadler, and Olav Zimmermann, editors, *From Computational Biophysics to Systems Biology (CBSB08)*, volume 40 of *NIC Series*, pages 201–204, 2008.

Thomas R. Einert, Douglas B. Staple, Hans-Jürgen Kreuzer, and Roland R. Netz. A three-state model with loop entropy for the over-stretching transition of DNA. *Biophysical Journal*, 99(2): 578–587, 2010.

Andreas Serr, Christian Sendner, Florian Müller, Thomas R. Einert, and Roland R. Netz. Single polymer adsorption in shear: flattening versus hydrodynamic lift and surface potential corrugation effects. *Europhysics Letters*, 92: 38002, 2010.

Thomas R. Einert and Roland R. Netz. Theory for RNA folding, stretching, and melting including loops and salt. To be published, 2010.

Thomas R. Einert, Henri Orland, and Roland R. Netz. Secondary structures of homopolymeric single-stranded nucleic acids including force and loop entropy: implications for hybridization. To be published, 2010.

Thomas R. Einert, Charles E. Sing, Alfredo Alexander-Katz, and Roland R. Netz. Internal friction of homopolymeric systems studied by diffusion and non-equilibrium unfolding of globules. To be published, 2010.

Charles E. Sing, Thomas R. Einert, Roland R. Netz, and Alfredo Alexander-Katz. Probing structural transitions in polymer globules with dynamic properties. To be published, 2010.

Other publications, which are not related to the present thesis:

Thomas Einert, Jörg Schilling Peter Lipowsky, Mark J. Bowick, and Andreas R. Bausch. Grain boundary scars on spherical crystals. *Langmuir*, 21(26): 12076–12079, 2005.

APPENDIX A

APPENDIX – DNA

A.1. Gibbs free energy of a worm-like chain

To calculate the worm-like chain Gibbs free energy $g_i^{\mathrm{WLC}}(F)$, we first calculate the Helmholtz free energy $H_i^{\mathrm{WLC}}(x)$ of a worm-like chain (WLC) in the fixed extension ensemble (Helmholtz ensemble). After that we Legendre transform $H_i^{\mathrm{WLC}}(x)$ in order to obtain $G_i^{\mathrm{WLC}}(F)$, which is the thermodynamic potential of a WLC in the fixed force ensemble (Gibbs ensemble). As we will see, $G_i^{\mathrm{WLC}}(F)$ is extensive with respect to the chain's contour length nl_i, l_i is the length of one segment, and thus we can define the Gibbs free energy per segment $g_i^{\mathrm{WLC}}(F) = G_i^{\mathrm{WLC}}(F)/n$. We introduce the dimensionless variables

$$\tilde{x}_i = \frac{x}{nl_i}\,, \qquad \tilde{F}_i = F\beta\xi_i\,, \qquad \tilde{H}_i^{\mathrm{WLC}} = H_i^{\mathrm{WLC}}\frac{\beta\xi_i}{nl_i}\,, \qquad \tilde{G}_i^{\mathrm{WLC}} = G_i^{\mathrm{WLC}}\frac{\beta\xi_i}{nl_i}\,, \tag{A.1}$$

where ξ_i is the persistence length of DNA in state $i = \mathrm{B}, \mathrm{S}, \mathrm{M}$. The interpolation formula of Marko and Siggia [26] for the force extension curve of the WLC

$$\tilde{F}_i(\tilde{x}_i) = \frac{1}{4(1-\tilde{x}_i)^2} + \tilde{x}_i - \frac{1}{4} \tag{A.2}$$

is used to calculate the free energy in the Helmholtz ensemble

$$\tilde{H}_i^{\mathrm{WLC}}(\tilde{x}) = \int_0^{\tilde{x}} \tilde{F}(\tilde{x}')\mathrm{d}\tilde{x}' + \tilde{H}_{i,0}^{\mathrm{WLC}} = \frac{\tilde{x}^2(2\tilde{x}-3)}{4(\tilde{x}-1)} + \tilde{H}_{i,0}^{\mathrm{WLC}}\,. \tag{A.3}$$

$\tilde{H}_{i,0}^{\mathrm{WLC}}$ is a free energy offset that accounts for the fact that even in the absence of an external force the free energies of chains consisting of B-, S-, or M-segments are not the same. In fact, this constant is not easy to calculate but can be dropped in the

following since it can be adsorbed into the free energy contributions g_i^0, cf. eq. (2.2) in the main text. Now, we invert eq. (A.1)

$$\tilde{x}_i(\tilde{F}_i) = 1 - \left[\frac{\frac{4\tilde{F}_i}{3} - 1}{\left(2 + \sqrt{4 - \left(\frac{4\tilde{F}_i}{3} - 1\right)^3}\right)^{1/3}} + \left(2 + \sqrt{4 - \left(\frac{4\tilde{F}_i}{3} - 1\right)^3}\right)^{1/3} \right]^{-1} \tag{A.4}$$

and perform a Legendre transformation to obtain the Gibbs free energy of a WLC in the Gibbs ensemble

$$\tilde{G}_i^{\mathrm{WLC}}(\tilde{F}_i) = \tilde{H}_i^{\mathrm{WLC}}(\tilde{x}_i(\tilde{F}_i)) - \tilde{F}_i \tilde{x}_i(\tilde{F}_i)\,. \tag{A.5}$$

Eqs. (A.1) and (A.1) lead to

$$\begin{aligned} g_i^{\mathrm{WLC}}(F) &= \frac{1}{n}\frac{nl_i}{\beta\xi_i}\tilde{G}^{\mathrm{WLC}}(F\beta\xi_i) \\ &= \frac{l_i}{\beta\xi_i}\left(\frac{\tilde{x}_i\left(F\beta\xi_i\right)^2\left(2\tilde{x}_i\left(F\beta\xi_i\right)-3\right)}{4(\tilde{x}_i\left(F\beta\xi_i\right)-1)} - F\beta\xi_i\tilde{x}_i\left(F\beta\xi_i\right)\right)\,, \end{aligned} \tag{A.6}$$

with $\tilde{x}_i(\tilde{F}_i)$ from eq. (A.1).

A.2. Explicit form of the grand canonical partition function of the three-state model for DNA

In the main text the expression

$$\mathcal{Z} = \sum_{k=0}^{\infty} \boldsymbol{\nu}^{\mathrm{T}} \cdot (\mathbf{M}_{\mathrm{PS}}\mathbf{V}_{\mathrm{PS}})^k \mathbf{M}_{\mathrm{PS}} \cdot \boldsymbol{\nu} = \boldsymbol{\nu}^{\mathrm{T}} \cdot (\mathbf{1} - \mathbf{M}_{\mathrm{PS}}\mathbf{V}_{\mathrm{PS}})^{-1}\mathbf{M}_{\mathrm{PS}} \cdot \boldsymbol{\nu}\,, \tag{A.7}$$

has been derived for the grand canonical partition function. The matrices are given by

$$\mathbf{M}_{\mathrm{PS}} = \begin{pmatrix} \mathcal{Z}_{\mathrm{B}} & 0 & 0 \\ 0 & \mathcal{Z}_{\mathrm{S}} & 0 \\ 0 & 0 & \mathcal{Z}_{\mathrm{M}} \end{pmatrix}\,, \quad \mathbf{V}_{\mathrm{PS}} = \begin{pmatrix} 0 & \mathrm{e}^{-\beta V_{\mathrm{BS}}} & \mathrm{e}^{-\beta V_{\mathrm{BM}}} \\ \mathrm{e}^{-\beta V_{\mathrm{SB}}} & 0 & \mathrm{e}^{-\beta V_{\mathrm{SM}}} \\ \mathrm{e}^{-\beta V_{\mathrm{MB}}} & \mathrm{e}^{-\beta V_{\mathrm{MS}}} & 0 \end{pmatrix}\,, \quad \boldsymbol{\nu} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\,. \tag{A.8}$$

1 is the unity matrix. Evaluating eq. (A.2) yields the grand canonical partition function of the three-state model

$$\begin{aligned}\mathcal{Z} = \Bigg[& \mathcal{Z}_\mathrm{B} + \mathcal{Z}_\mathrm{S} + \mathcal{Z}_\mathrm{M} \\ & + \left(\mathrm{e}^{-\beta V_\mathrm{BS}} + \mathrm{e}^{-\beta V_\mathrm{SB}}\right)\mathcal{Z}_\mathrm{B}\mathcal{Z}_\mathrm{S} + \left(\mathrm{e}^{-\beta V_\mathrm{SM}} + \mathrm{e}^{-\beta V_\mathrm{MS}}\right)\mathcal{Z}_\mathrm{S}\mathcal{Z}_\mathrm{M} + \left(\mathrm{e}^{-\beta V_\mathrm{MB}} + \mathrm{e}^{-\beta V_\mathrm{BM}}\right)\mathcal{Z}_\mathrm{M}\mathcal{Z}_\mathrm{B} \\ & + \Big(\mathrm{e}^{-\beta(V_\mathrm{BS}+V_\mathrm{SM})} + \mathrm{e}^{-\beta(V_\mathrm{SM}+V_\mathrm{MB})} + \mathrm{e}^{-\beta(V_\mathrm{MB}+V_\mathrm{BS})} + \mathrm{e}^{-\beta(V_\mathrm{BM}+V_\mathrm{MS})} + \mathrm{e}^{-\beta(V_\mathrm{MS}+V_\mathrm{SB})} \\ & \quad + \mathrm{e}^{-\beta(V_\mathrm{SB}+V_\mathrm{BM})} - \mathrm{e}^{-\beta(V_\mathrm{BS}+V_\mathrm{SB})} - \mathrm{e}^{-\beta(V_\mathrm{SM}+V_\mathrm{MS})} - \mathrm{e}^{-\beta(V_\mathrm{MB}+V_\mathrm{BM})}\Big)\mathcal{Z}_\mathrm{B}\mathcal{Z}_\mathrm{S}\mathcal{Z}_\mathrm{M}\Bigg] \\ \times & \Bigg[1 - \mathrm{e}^{-\beta(V_\mathrm{BS}+V_\mathrm{SB})}\mathcal{Z}_\mathrm{B}\mathcal{Z}_\mathrm{S} - \mathrm{e}^{-\beta(V_\mathrm{SM}+V_\mathrm{MS})}\mathcal{Z}_\mathrm{S}\mathcal{Z}_\mathrm{M} - \mathrm{e}^{-\beta(V_\mathrm{MB}+V_\mathrm{BM})}\mathcal{Z}_\mathrm{M}\mathcal{Z}_\mathrm{B} \\ & - \left(\mathrm{e}^{-\beta(V_\mathrm{BS}+V_\mathrm{SM}+V_\mathrm{MB})} + \mathrm{e}^{-\beta(V_\mathrm{BM}+V_\mathrm{MS}+V_\mathrm{SB})}\right)\mathcal{Z}_\mathrm{B}\mathcal{Z}_\mathrm{S}\mathcal{Z}_\mathrm{M}\Bigg]^{-1},\end{aligned} \tag{A.9}$$

which depends on the grand canonical partition functions $\mathcal{Z}_i$ of the different regions.

A.3. Origin of the logarithmic loop entropy

In the simplest case a polymer can be described as an ideal random walk with step length b in $d = 3$ dimensions. When speaking of polymers b is also called Kuhn length. Let us first consider a one-dimensional random walk. The probability to be after N steps at point $x = nb$ is given by the binomial distribution

$$\begin{aligned}P(x) &= \frac{N!}{((N+n)/2)!((N-n)/2)!}\cdot p^{(N+n)/2}\cdot(1-p)^{(N-n)/2} \\ &= \binom{N}{(N+n)/2}\cdot p^{(N+n)/2}\cdot(1-p)^{(N-n)/2},\end{aligned} \tag{A.10}$$

where $p = 1/2$ is the probability to move to the right and $(1-p)$ the probability to move to the left. By virtue of the central limit theorem the binomial distribution can be approximated by the normal distribution if N is large

$$P(x) \simeq \frac{1}{\sqrt{Nb^2 2\pi}}\mathrm{e}^{-\frac{1}{2}\frac{x^2}{Nb^2}}. \tag{A.11}$$

Now, we consider a random walk in three dimensions, where there are $N/3$ steps in each spatial direction. Since the steps in each of the three directions are independent from each other, the probability distribution for being at point $\boldsymbol{R} = (x, y, z)$ after N steps is

$$P(\boldsymbol{R}) = P(x)P(y)P(z) = \frac{1}{(Nb^2 2\pi/3)^{3/2}}\mathrm{e}^{-\frac{3}{2}\frac{R^2}{Nb^2}}. \tag{A.12}$$

The probability of a closed random walk, i. e. a loop, is the probability to return to the origin after N steps

$$P(0) = \frac{1}{(N b^2 2\pi/3)^{3/2}} \,. \tag{A.13}$$

The entropy difference between an ideal polymer forming a loop and an unconstrained polymer is therefore given by

$$\Delta S_{\text{loop}}(N) = \mathrm{k_B} \ln P(0) \sim \mathrm{k_B} \ln N^{-\hat{c}} \,, \tag{A.14}$$

where we introduced the loop exponent $\hat{c} = \hat{c}_{\text{ideal}} = d/2 = 3/2$ [74].

If one considers self-avoiding polymers the loop exponent increases, $\hat{c}_{\text{SAW}} = d\nu \simeq 1.76$ with $\nu \simeq 0.588$ in $d = 3$ dimensions [74]. Hence the entropic penalty for closing a loop increases. However, helices emerging from the loop increase $\hat{c}$ further. In the asymptotic limit of long helical sections, renormalization group predicts $\hat{c}_l = d\nu + \sigma_l - l\sigma_3$ for a loop with l emerging helices [75, 102] where $\sigma_l = \varepsilon l(2-l)/16 + \varepsilon^2 l(l-2)(8l-21)/512 + \mathcal{O}(\epsilon^3)$ in an $\epsilon = 4-d$ expansion. One obtains $c_1 = 2.06$ for terminal loops in hairpin structures, $c_2 = 2.14$ for internal loops and $c_4 = 2.16$ for a loop with four emerging helices [1].

The same analysis holds if one considers different topologies, namely peeling off one strand from the other starting from a nick. In this case one can show that $\hat{c} = 0$ for an ideal polymer and $\hat{c} = 0.092$ for a self avoiding polymer [75, 102].

Application of force to loops also alters the value of the loop exponent, which has been shown by Hanke et al. [58].

A.4. Polylogarithm

The polylogarithm or Jonquiére's function is defined on the open unit disc in the complex plane by the series [76, 115]

$$\mathrm{Li}_a(z) = \sum_{n=0}^{\infty} \frac{z^n}{n^a} \,. \tag{A.15}$$

The polylogarithm for different values of a is depicted in fig. A.1. The definition of the polylogarithm can be generalized by analytic continuation on the whole complex plane. $\mathrm{Li}_a(z)$ has two branch points with respect to z, namely $z = 1$ and $z = \tilde{\infty}$ (complex infinity). Note that $\mathrm{Li}_a(1) = \zeta_a$ is the Riemann ζ-function.

By virtue of eq. (A.4) one can show that the derivative of $\mathrm{Li}_a(z)$ is given by

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathrm{Li}_a(z) = \frac{1}{z}\mathrm{Li}_{a-1}(z) \,. \tag{A.16}$$

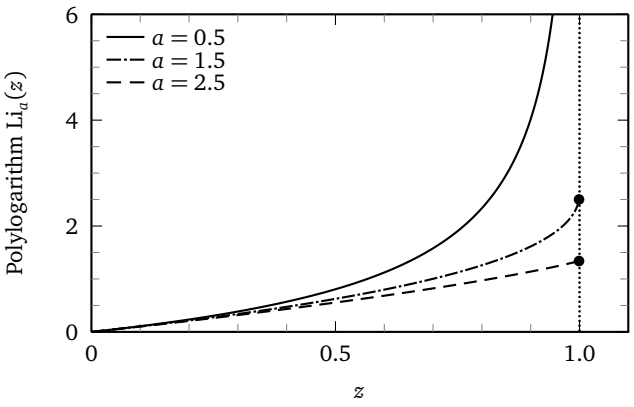


Figure A.1.: Plot of the polylogarithm $\mathrm{Li}_a(z)$ on the real axis for three different values of the parameter $a = 0.5,\ 1.5$ and 2.5. The polylogarithm diverges at $z = 1$ for $a \leq 1$. For $a > 1$ the value of $\mathrm{Li}_a(z = 1) < \infty$ is finite. For $a \leq 2$ the polylogarithm possesses a divergent slope at $z = 1$.

The expansion of $\mathrm{Li}_a(z)$ around $z_0 = 1$ for $z \notin [1, \infty)$ is given by [76]

$$\mathrm{Li}_a(z) = \zeta_a + \zeta_{a-1}(z-1) + \frac{1}{2}\left(\zeta_{a-2} - \zeta_{a-1}\right)(z-1)^2 + \ldots \\ + (1-z)^{a-1}\left(\Gamma(1-a) + \frac{1}{2}(1-a)\Gamma(1-a)(1-z) + \ldots\right) \quad \text{(A.17)}$$

From this expansion or also from eq. (A.4), the limiting behavior of the polylogarithm and its derivative at $z = 1$ can be deduced

$$\lim_{z\to 1} \mathrm{Li}_a(z) = \begin{cases} \infty & \text{for } 0 < a \leq 1, \\ \zeta_a & \text{for } a > 1, \end{cases} \quad \text{(A.18)}$$

$$\lim_{z\to 1} \frac{\mathrm{d}}{\mathrm{d}z}\mathrm{Li}_a(z) = \begin{cases} \infty & \text{for } a \leq 2, \\ \zeta_{a-1} & \text{for } a > 2. \end{cases} \quad \text{(A.19)}$$

A.5. Singularity analysis of generating functions

In this appendix we are trying to shed some light on the general relationship between the asymptotic expansion of a generating function f near its dominant singularities, i. e. the singularities which are closest to the origin, and the asymptotic expansion of the function's coefficients f_n in its series expansion $f(z) = \sum_n f_n z^n$. It will turn out that for $n \to \infty$ the coefficients f_n are solely determined by the dominant singularities [78,

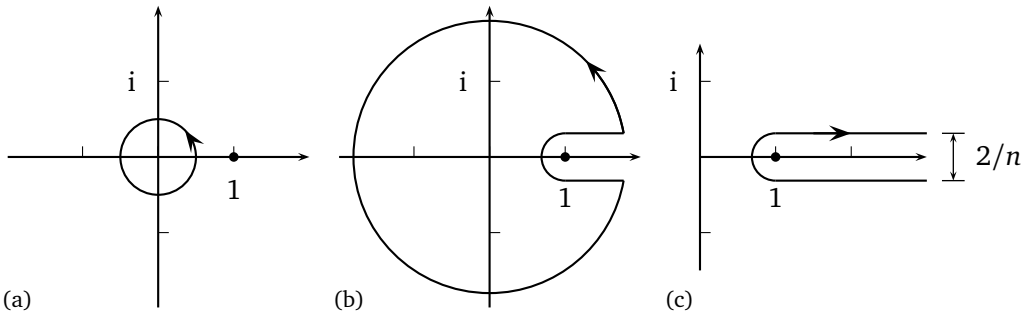


Figure A.2.: The contours (a) $\mathcal{C}_0$, (b) $\mathcal{C}_1$ and (c) $\mathcal{H}(n)$ used for calculating the coefficients used in the proof of proposition A.5.1. As the only singularity is located at $z = 1$ (denoted by the dot) and since the integrand in eq. (A.5) vanishes for $|z| \to \infty$, the integration along these three contours yields the same result.

79]. In statistical mechanics the generating function f is generally referred to as the grand canonical partition function whereas the coefficients f_n are referred to as the canonical partition functions

Proposition A.5.1. *Let α be an arbitrary complex number in $\mathbb{C} \setminus \{0, -1, -2, -3, \ldots\}$. The coefficient of z^n in*

$$f(z) = (1-z)^{-\alpha} \tag{A.20}$$

admits for large n a full asymptotic expansion in descending powers of n. We will use for the coefficient of z^n in a power series of a function $f(z)$ the common notation $f_n = [z^n]f(z)$. The coefficient of z^n behaves for large n like

$$[z^n]f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} \left(1 + \sum_{k=1}^{\infty} \frac{e_k}{n^k}\right) , \tag{A.21}$$

where e_k is a polynomial in α of degree $2k$. In particular

$$e_1 = \frac{1}{2}\alpha(\alpha-1) \tag{A.22}$$

$$e_2 = \frac{1}{24}\alpha(\alpha-1)(\alpha-2)(3\alpha-1) \tag{A.23}$$

$$\vdots$$

Proof. First, we express the coefficient $f_n = [z^n](1-z)^{-\alpha}$ as a complex contour integral via Cauchy's formula

$$f_n = \frac{1}{2\pi \mathrm{i}} \int_{\mathcal{C}} dz \frac{(1-z)^{-\alpha}}{z^{n+1}} , \tag{A.24}$$

where $\mathcal{C}$ is a small enough contour around the origin, so that all singularities of f lie outside. In our case the only singularity of f is located at $z = 1$. For instance we can start with $\mathcal{C} = \mathcal{C}_0$ which is a positively oriented circle with radius $1/2$ around the origin, $\mathcal{C}_0 = \{z \in \mathbb{C}, |z| = 1/2\}$, see fig. A.2a. Next, we deform this contour into a contour $\mathcal{C}_1$ such, that it does not cross the half line $\Re z \geq 1$. $\mathcal{C}_1$ consists of a large circle with radius $R > 1$ and a notch that comes back near to the left of $z = 1$, cf. fig. A.2b. The contribution of this large circle is of $\mathcal{O}(R^{-n-\alpha})$ and is therefore negligible for $R \to \infty$. Thus, we let R tend to infinity and are left with an integral representation for f_n where $\mathcal{C}_0$ has been replaced by a contour $\mathcal{H}$, see fig. A.2c, that can be parameterized as follows

$$\mathcal{H}(n) = \mathcal{H}^-(n) + \mathcal{H}^+(n) + \mathcal{H}^\circ(n)\,, \tag{A.25}$$

with

$$\mathcal{H}^-(n) = \{z \in \mathbb{C},\, z = w - \tfrac{\mathrm{i}}{n},\, w \in [1, \infty)\} \tag{A.26}$$

$$\mathcal{H}^+(n) = \{z \in \mathbb{C},\, z = w + \tfrac{\mathrm{i}}{n},\, w \in [1, \infty)\} \tag{A.27}$$

$$\mathcal{H}^\circ(n) = \{z \in \mathbb{C},\, z = 1 - \tfrac{\mathrm{e}^{\mathrm{i}\phi}}{n},\, \phi \in [-\tfrac{\pi}{2}, \tfrac{\pi}{2}]\}\,. \tag{A.28}$$

The contour starts slightly below the positive real axis at $+\infty$, winds clockwise in a semi-circle around $z = 1$ and ends at $+\infty$ slightly above the positive real axis. This type of contour is also called Hankel contour.

Let us perform a change of variable

$$z = 1 + \frac{t}{n} \tag{A.29}$$

and the integral in eq. (A.5) reads now

$$f_n = \frac{n^{\alpha-1}}{2\pi\mathrm{i}} \int_{\mathcal{H}} (-t)^{-\alpha} \left(1 + \frac{t}{n}\right)^{-n-1} \mathrm{d}t\,. \tag{A.30}$$

Next, we perform the expansion

$$\begin{aligned} \left(1 + \frac{t}{n}\right)^{-n-1} &= \mathrm{e}^{-(n+1)\log(1+t/n)} \\ &= \mathrm{e}^{-t} \left(1 + \frac{t^2 - 2t}{2n} + \frac{3t^4 - 20t^3 + 24t^2}{24n^2} \ldots\right)\,. \end{aligned} \tag{A.31}$$

Thus, the integrand in eq. (A.5) converges point-wise to $(-t)^{-\alpha} e^{-t}$ for $n \to \infty$. To exchange integration and execution of the limit we need to establish uniform convergence. From basic calculus it is known that a point-wise convergent series of functions converges uniformly on any bounded domain. Therefore we split the contour into two parts, namely into a part where $\Re t > t_0 > 1$ and a part where $\Re t \leq t_0 1$. It is clear

from eqs. (A.5), (A.5), and (A.5) that for large enough t_0 and large enough n the error that arises from neglecting the part with $\Re t > t_0$ can be made arbitrarily small. Thus we are left with a compact domain on which the series converges uniformly. Therefore one may exchange the integration and the limit. The integral in eq. (A.5) reads now

$$\int\limits_{\substack{\mathcal{H}\\ t\leq c}} (-t)^{-\alpha}\mathrm{e}^{-t}\left(1+\frac{t^2-2t}{2n}+\frac{3t^4-20t^3+24t^2}{24n^2}+\ldots\right). \tag{A.32}$$

Now we can use Hankel's representation of the Γ-function [115]

$$\frac{1}{\Gamma(s)}=\frac{1}{2\pi\mathrm{i}}\int_{\mathcal{H}}(-t)^{-s}\mathrm{e}^{-t}\mathrm{d}t\,. \tag{A.33}$$

A term of the form t^r/n^s in the expansion induces, by Hankel's formula, a term of the form $n^{-s}/\Gamma(\alpha-r)$. Since

$$\Gamma(\alpha-k)=\Gamma(\alpha)\frac{1}{(\alpha-1)(\alpha-2)\ldots(\alpha-k)} \tag{A.34}$$

the expansion in the above proposition follows. □

This can be generalized in a straight forward way to functions which do not possess the simple form depicted in eq. (A.5.1). It will turn out that only the dominant singularities determine the behavior of the coefficients $f_n=[z^n]f(z)$ for $n\to\infty$. Let us assume that $\mathcal{S}$ is the set of all singularities of the function $f(z)$, i. e. $\mathbb{C}\setminus\mathcal{S}$ is the maximal set on which f is holomorphic. A subset $\mathcal{S}_\mathrm{d}$ is called set of dominant singularities if and only if $|z_\mathrm{d}|<|z|$ and $|z_\mathrm{d}|=|z'_\mathrm{d}|$, $\forall z_\mathrm{d},z'_\mathrm{d}\in\mathcal{S}_\mathrm{d}$ and $\forall z\in\mathcal{S}\setminus\mathcal{S}_\mathrm{d}$. A singularity $z_\mathrm{d}\in\mathcal{S}_\mathrm{d}$ is called dominant singularity. We will recover a result very similar to that of the preceding proposition.

Like in the previous paragraph the prove of the following theorem relies on the integration on Hankel-type paths. In fact weaker conditions than those used before suffice and allow us to establish the asymptotics of the coefficients. For now let us assume that there is only one dominant singularity.

Definition A.5.1. Given two numbers ϕ and R, with $R>1$ and $0<\phi<\pi/2$, the open domain $\Delta(\phi,R)$ is defined as

$$\Delta(\phi,R)=\{z\in\mathbb{C},\ |z|<R,\ z\neq 1,\ |\arg(z-1)|>\phi\}\,. \tag{A.35}$$

Any domain is called Δ-domain if it is a $\Delta(\phi,R)$ for some R and ϕ, see fig. A.3. A function is called Δ-analytic if it is analytic in some Δ-domain.

Let

$$\mathcal{M}=\left\{(1-z)^{-\alpha},\ \alpha\in\mathbb{C}\right\}\,. \tag{A.36}$$

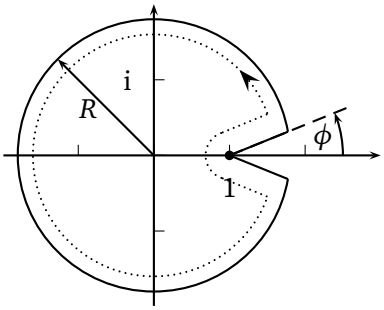


Figure A.3.: Definition of the Δ-domain. Everything that is inside the solid line belongs to $\Delta(\phi, R)$ according to definition A.5.1. The dotted line denotes the integration contour which is used in order to prove proposition A.5.2.

Proposition A.5.2. *Let $f(z)$ be an analytic function at 0, with a singularity at ξ, such that $f(z)$ can be continued to a domain of the form $\xi \cdot \Delta_0$, for a Δ-domain Δ_0, where $\xi \cdot \Delta_0$ is the image of Δ_0 by the mapping $z \mapsto \xi z$. Assume that there exist two functions σ, τ, where σ is a (finite) linear combination of functions in $\mathcal{M}$ and $\tau \in \mathcal{M}$, so that*

$$f(z) = \sigma(z/\xi) + \mathcal{O}(\tau(z/\xi)) \quad \text{as } z \to \xi \text{ in } \xi\Delta_0 \,. \tag{A.37}$$

Then, the coefficients of $f(z)$ satisfy the asymptotic estimate

$$f_n = [z^n] f(z) = \xi^{-n} \sigma_n + \mathcal{O}(\xi^{-n} \tau_n^*) \,, \tag{A.38}$$

where $\sigma_n = [z^n]\sigma(z)$ has its coefficients determined by proposition A.5.1 and $\tau_n^* = n^{a-1}$, if $\tau(z) = (1-z)^{-a}$. In particular, if $f(z) \sim K(\xi - z)^{-\alpha}$, where $K \in \mathbb{C}$ is a constant, then

$$f_n = [z^n] f(z) \sim \xi^{-n} \xi^{-\alpha} n^{\alpha-1} / \Gamma(\alpha) \,. \tag{A.39}$$

Proof. The proof follows the idea of the proof of proposition A.5.1 and will not be given here. □

Proposition A.5.2 can easily be generalized to the case where also logarithmic singularities appear. It may also be generalized to a finite number of isolated, dominant singularities. In the case of multiple singularities, the separate contributions to the coefficients of the generating function from each of the singularities, as given by the basic singularity analysis process, is summed up [78, 79].

A.6. Tables with model parameters of the three-state model for DNA

Table A.1.: Parameters for stretching curves at temperature $T = 293$ K.

Parameter	Symbol	λ-DNA	λ-DNA + DDP	Reference
segment length	l_B	0.34	0.34	Wenner et al. [80]
(nm)	l_S	0.61	0.60	fitted
	l_M	0.71	0.71	Hugel et al. [81]
persistence length	ξ_B	48	48	Wenner et al. [80]
(nm)	ξ_S	25	25	assumed
	ξ_M	3	3	Murphy et al. [83]
stretch modulus	κ_B	1.0	1.0	Wenner et al. [80]
(nN)	κ_S	$2 \cdot 9.4$	$2 \cdot 9.4$	Hugel et al. [81]
	κ_M	$2 \cdot 9.4$	$2 \cdot 9.4$	Hugel et al. [81]
chemical potential	g_B^0	0.0	0.0	defined
(10^{-20} J)	g_S^0	1.6	1.2	fitted
	g_M^0	2.4	2.8	fitted
interfacial energy	V_{BB}, V_{SS}, V_{MM}	0.0	0.0	incorporated in g_i^0
(10^{-20} J)	V_{BS}	1.2[a]	0.0	assumed
	V_{BM}	1.2[a]	0.0	assumed
	V_{SM}	0.0	0.0	

[a] disrupt stacked base pairs [84]

Table A.2.: Temperature dependence of parameters for DNA without DDP. The enthalpy h_M and entropy s_M of the denatured state are determined by fixing the denaturation temperature T_c for zero force at $T_c = 348$ K.

Parameter	Symbol	λ-DNA	Reference
persistence length	$\xi_B(T)$	$48/(T/293\text{ K})$	
(nm)	$\xi_S(T)$	$25/(T/293\text{ K})$	
	$\xi_M(T)$	$3/(T/293\text{ K})$	
chemical potential	$g_i^0(T) = h_i - Ts_i$		
enthalpy	h_B	0.0	
(10^{-20} J)	h_S	7.0[a]	Clausen-Schaumann et al. [46]
	h_M	15	fit for $\hat{c} = 0$
	h_M	16	fit for $\hat{c} = 3/2$
entropy	s_B	0.0	
(10^{-22} J/K)	s_S	1.8[a]	Clausen-Schaumann et al. [46]
	s_M	4.2	fit for $\hat{c} = 0$
	s_M	4.6	fit for $\hat{c} = 3/2$

[a] slightly rescaled

Appendix B

Appendix – RNA

B.1. The classical two-state Poland-Scheraga model

The Poland-Scheraga model describes a linear molecule whose segments can be in two different states, namely bound (B) or molten (M) [27, 28]. It is a special case of the three-state model for DNA [3] and is obtained by omitting the S-state and setting $\mathcal{Z}_S = 0$. The partition function of the two-state Poland-Scheraga model reads

$$\mathcal{Z}^{PS} = (1+\mathcal{Z}_M)\sum_{k=0}^{\infty}(\mathcal{Z}_B\mathcal{Z}_M)(1+\mathcal{Z}_B) - 1 = \frac{\mathcal{Z}_B + \mathcal{Z}_M + 2\mathcal{Z}_B\mathcal{Z}_M}{1-\mathcal{Z}_B\mathcal{Z}_M}\,, \tag{B.1}$$

where

$$\mathcal{Z}_B = \sum_{k=0}^{\infty}(zw)^k = \frac{zw}{1-zw} \tag{B.2}$$

and

$$\mathcal{Z}_M = \sum_{k=0}^{\infty}\frac{z^k}{k^{\hat{c}}} = \mathrm{Li}_{\hat{c}}(z) \tag{B.3}$$

are the grand canonical partition functions of regions in the respective states [3] with no applied force. Note that in eq. (B.1) the free energy of a molten region contains no extensive part as unbound bases outside a loop are defined as the ground state and hence all extensive parts accumulate in the free energy of bound B-segments $\varepsilon = \mathrm{k_B}T\ln w$. Finally, the grand canonical partition function, eq. (B.1), reads

$$\mathcal{Z}^{PS} = \frac{wz + (1+wz)\mathrm{Li}_{\hat{c}}(z)}{1-wz(1+\mathrm{Li}_{\hat{c}}(z))}\,. \tag{B.4}$$

The singularities are readily recognized as the branch point of the polylogarithm [115]

$$z_b^{PS} = 1 \tag{B.5}$$

and the pole $z_{\mathrm{p}}^{\mathrm{PS}}$ of the fraction in eq. (B.1), which is the root of the denominator

$$1 = w z_{\mathrm{p}}^{\mathrm{PS}}(1 + \mathrm{Li}_{\hat{c}}(z_{\mathrm{p}}^{\mathrm{PS}})) \,. \tag{B.6}$$

The pole and the branch point coincide at the critical point, which therefore is given by

$$w_{\mathrm{c}}^{\mathrm{PS}} = \frac{1}{1+\zeta_{\hat{c}}} \,. \tag{B.7}$$

B.2. Derivatives of κ

In the main text we derived the three constitutive equations determining the thermodynamic behavior of the system

$$\kappa(w,z) = 1 + \frac{w}{\kappa(w,z)}\mathrm{Li}_c(z\kappa(w,z)) \tag{B.8a}$$

$$\kappa(w,z_{\mathrm{b}}(w))^2 = w\mathrm{Li}_{c-1}(z_{\mathrm{b}}(w)\kappa(w,z_{\mathrm{b}}(w))) - w\mathrm{Li}_c(z_{\mathrm{b}}(w)\kappa(w,z_{\mathrm{b}}(w))) \tag{B.8b}$$

$$s z_{\mathrm{p}}(w,s)\kappa(w,z_{\mathrm{p}}(w,s)) = 1 \,. \tag{B.8c}$$

The function $\kappa(w,z)$ is the smallest positive root of eq. (4.11a). Derivatives of κ are obtained by implicit differentiation. To obtain the derivative of κ with respect to an arbitrary quantity one differentiates both sides of eq. (4.11a) with respect to this quantity and solves for the desired derivative. Explicitly,

$$\frac{\partial\kappa}{\partial z} = \frac{w\kappa\mathrm{Li}_{c-1}(z\kappa)}{z(\kappa^2 + w\mathrm{Li}_c(z\kappa) - w\mathrm{Li}_{c-1}(z\kappa))} \tag{B.9}$$

$$\frac{\partial\kappa}{\partial w} = \frac{\kappa\mathrm{Li}_c(z\kappa)}{\kappa^2 + w\mathrm{Li}_c(z\kappa) - w\mathrm{Li}_{c-1}(z\kappa)} \tag{B.10}$$

$$\frac{\partial\kappa}{\partial T} = -\frac{\kappa w}{T}\cdot\frac{\ln z\mathrm{Li}_{c-1}(z\kappa) + \ln w\mathrm{Li}_c(z\kappa)}{\kappa^2 + w\mathrm{Li}_c(z\kappa) - w\mathrm{Li}_{c-1}(z\kappa)} \,. \tag{B.11}$$

B.3. Expansion of the branch point for $\mathrm{T} < \mathrm{T}_{\mathrm{m}}$

To obtain the critical behavior for $T < T_{\mathrm{m}}$ at zero force, $s = 1$, we perform an asymptotic expansion of eqs. (4.11a) and (4.11a) around the critical point, where the branch point and the pole coincide. Thus, at the critical point all eqs. (4.11a) have to hold and we obtain the critical values exactly as

$$\kappa_{\mathrm{c}} = \frac{1}{2}\left(1 + \sqrt{1+4w\zeta_c}\right) \,, \tag{B.12a}$$

$$z_{\mathrm{c}} = 2\left(1 + \sqrt{1+4w\zeta_c}\right)^{-1} \,, \tag{B.12b}$$

$$w_{\mathrm{c}} = \frac{\zeta_{c-1} - \zeta_c}{(\zeta_{c-1} - 2\zeta_c)^2} \,. \tag{B.12c}$$

As an ansatz for $z_b(T)$ and $\kappa_b(T)$ we use a power series in $d = w/w_c - 1$, where $w = \exp(\varepsilon/(k_B T))$ is the Boltzmann weight of a base pair. To simplify notation, we define $\alpha = (c-2)^{-1}$ and write

$$\begin{aligned} z_b/z_c &\sim 1 + a_1 d + a_2 d^2 + \dots \\ &\quad + d^\alpha\left(a_\alpha + a_{\alpha+1} d + a_{\alpha+2} d^2 + \dots\right) + a_{2\alpha-1} d^{2\alpha-1} + \dots \end{aligned} \tag{B.13a}$$

$$\begin{aligned} \kappa_b/\kappa_c &\sim 1 + b_1 d + b_2 d^2 + \dots \\ &\quad + d^\alpha\left(b_\alpha + b_{\alpha+1} d + b_{\alpha+2} d^2 + \dots\right) + b_{2\alpha-1} d^{2\alpha-1} + \dots \end{aligned} \tag{B.13b}$$

Plugging the ansatz (4.11c) into eqs. (4.11a) and (4.11a) we can solve order by order. To do so, the series representation of the polylogarithm $\mathrm{Li}_\nu(x)$ around $x = 1$ is used [76]

$$\begin{aligned} \mathrm{Li}_\nu(x) &\sim \zeta_\nu - \zeta_{\nu-1}(1-x) + \frac{1}{2}(\zeta_{\nu-2} - \zeta_{\nu-1})(x-1)^2 + \dots \\ &\quad + (1-x)^{\nu-1}\left(\Gamma(1-\nu) + \frac{1}{2}(1-\nu)\Gamma(1-\nu)(1-x) + \dots\right). \end{aligned} \tag{B.14}$$

We obtain the coefficients for z_b

$$a_1 = -\frac{\zeta_c}{\zeta_{c-1}} \tag{B.15a}$$

$$a_2 = \frac{\zeta_c^2}{\zeta_{c-1}^3}(2\zeta_{c-1} - \zeta_c) \tag{B.15b}$$

$$a_3 = -\frac{\zeta_c^3}{\zeta_{c-1}^5}(5\zeta_{c-1}^2 - 6\zeta_{c-1}\zeta_c + 2\zeta_c^2) \tag{B.15c}$$

$$a_4 = \frac{\zeta_c^4}{\zeta_{c-1}^7}(14\zeta_{c-1}^3 - 28\zeta_{c-1}^2\zeta_c + 20\zeta_{c-1}\zeta_c^2 - 5\zeta_c^3) \tag{B.15d}$$

$$a_5 = -\frac{\zeta_c^5}{\zeta_{c-1}^9}(42\zeta_{c-1}^4 - 120\zeta_{c-1}^3\zeta_c + 135\zeta_{c-1}^2\zeta_c^2 - 70\zeta_{c-1}\zeta_c^3 + 14\zeta_c^4) \tag{B.15e}$$

$$\vdots$$

$$a_\alpha = 0 \tag{B.15f}$$

$$a_{\alpha+1} = -\frac{\Gamma(1-c) + \Gamma(2-c)}{\zeta_{c-1}}\left(-\frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}}\right)^{\alpha+1} \tag{B.15g}$$

$$\begin{aligned} a_{\alpha+2} &= \frac{\left(\Gamma(1-c) + \Gamma(2-c)\right)\left(\zeta_{c-1}^2 + 4(c-2)\zeta_{c-1}\zeta_c + (5-3c)\zeta_c^2\right)}{(c-2)\zeta_{c-1}^3} \\ &\quad \times \left(-\frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}}\right)^{\alpha+1} \end{aligned} \tag{B.15h}$$

$$\vdots$$

$$a_{2\alpha-1} = 0 \tag{B.15i}$$

$$\vdots$$

We obtain the coefficients for κ_b

$$b_1 = \frac{\zeta_c}{\zeta_{c-1}} \tag{B.16a}$$

$$b_2 = -\frac{\zeta_c^2}{\zeta_{c-1}^3}(\zeta_{c-1} - \zeta_c) \tag{B.16b}$$

$$b_3 = 2\frac{\zeta_c^3}{\zeta_{c-1}^5}(\zeta_{c-1} - \zeta_c)^2 \tag{B.16c}$$

$$b_4 = -5\frac{\zeta_c^4}{\zeta_{c-1}^7}(\zeta_{c-1} - \zeta_c)^3 \tag{B.16d}$$

$$b_5 = 14\frac{\zeta_c^5}{\zeta_{c-1}^9}(\zeta_{c-1} - \zeta_c)^4 \tag{B.16e}$$

$$\vdots$$

$$b_\alpha = -\left(-\frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}}\right)^\alpha \tag{B.16f}$$

$$\begin{aligned} b_{\alpha+1} = {} & \frac{\Gamma(1-c)+\Gamma(2-c)}{\zeta_{c-1}}\left(-\frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}}\right)^{\alpha+1} \\ & + \left(-\frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}}\right)^\alpha \frac{\zeta_{c-1}^2 - (c-2)\zeta_{c-1}\zeta_c - \zeta_c^2}{(c-2)\zeta_{c-1}^2} \end{aligned} \tag{B.16g}$$

$$\begin{aligned} b_{\alpha+2} = {} & \frac{\zeta_{c-1} - \zeta c}{2(c-2)^2 \cdot \Gamma(2-c)\zeta_{c-1}^4}\left(-\frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}}\right)^\alpha \\ & \times \Big(2(c-2)\Gamma(1-c)(\zeta_{c-1} - 2\zeta_c)\big(\zeta_{c-1}^2 + 2(c-2)\zeta_{c-1}\zeta_c + (5-3c)\zeta_c^2\big) \\ & + \Gamma(2-c)\big((c-3)\zeta_{c-1}^3 + (c-3)(4c-7)\zeta_{c-1}^2\zeta_c \\ & - (3c-5)(4c-9)\zeta_{c-1}\zeta_c^2 + (3c-5)(4c-7)\zeta_c^3\big)\Big) \end{aligned} \tag{B.16h}$$

$$\vdots$$

$$b_{2\alpha-1} = -\frac{\zeta_{c-2} - 3\zeta_{c-1} + 2\zeta_c}{(c-2)\Gamma(2-c)}\left(-\frac{\zeta_{c-1}^2 - 3\zeta_{c-1}\zeta_c + 2\zeta_c^2}{\Gamma(2-c)\zeta_{c-1}}\right)^{2\alpha-1} \tag{B.16i}$$

$$\vdots$$

Remarkably, the expansion of the product $z_b\kappa_b$ yields

$$z_b\kappa_b \sim 1 + (a_\alpha + b_\alpha)d^\alpha + (a_{\alpha+1} + b_1 a_\alpha + b_{\alpha+1} + a_1 b_\alpha)d^{\alpha+1} + \mathcal{O}(d^{\alpha+2}), \quad \text{(B.17)}$$

meaning that all integer powers d^n, for $n < \alpha$, vanish. n is the integer with $(c-2)^{-1} - 1 < n < (c-2)^{-1}$. This fact renders the transition of high order when $c \to 2$ and thus $\alpha \to \infty$. For the fraction of bound bases this can be seen directly by expanding the low temperature expression

$$\theta = 2\frac{\mathrm{Li}_c(z_b\kappa_b)}{\mathrm{Li}_{c-1}(z_b\kappa_b)} \quad \text{(B.18)}$$

using eqs. (B.3) and (B.3).

B.4. Minimum free energy structure prediction

The partition function involves the enumeration and contribution of all possible secondary structures and can be calculated exactly by the recursion relation eq. (4.11). From the partition function $Q^M_{i,j}$ the free energy of an RNA strand ranging from base i through j with M non-nested backbone bonds can be calculated

$$\mathcal{F}^M_{i,j} = -\mathrm{k_B}T \ln Q^M_{i,j} \,. \quad \text{(B.19)}$$

Another important quantity is the minimum free energy $\mathcal{E}$ (mfe) associated with the ground state structure, also called mfe structure. In analogy to the partition function we define the mfe of a substrand ranging from base i through j with M non-nested backbone segments as $\mathcal{E}^M_{i,j}$. RNA structure prediction relies on the assumption that the native – experimentally observed – structure is given by the mfe structure. Following the same idea of the recursive calculation of the partition function, see eq. (4.11), we can determine the minimum free energy of a sequence by replacing the sums by the min operator and obtain the recursion relations

$$\mathcal{E}^{M+1}_{i,j+1} = \min\left\{\mathcal{E}^M_{i,j}, \min_{k=i+1,\ldots,j}\left\{\mathcal{E}^M_{i,k-1} + \mathcal{E}^0_{k,j+1}\right\}\right\} \quad \text{(B.20a)}$$

$$\mathcal{E}^0_{k,j+1} = \min_{m=0,\ldots,j+1-k}\left\{\mathcal{E}^m_{k+1,j} + \varepsilon_{k,j+1}\right\} \,. \quad \text{(B.20b)}$$

$\varepsilon_{k,j+1}$ is the free energy of the base pair $(k, j+1)$. From the three-dimensional array $\mathcal{E}^M_{i,j}$ the mfe structure can now be obtained by a recursive backtracking algorithm. The following pseudo code initiates the backtracking

```
ℰ_buffer = ∞;
for( M = 0; M < N; M++ ){
 if( ℰ^M_{0,N} < ℰ_buffer ){
   ℰ_buffer = ℰ^M_{0,N};
```

```
    M_min = M;
  }
}

backtrack(0, N, M_min);
```

The function `backtrack(i, j, M)` performs the next backtracking step and calls itself recursively. Its pseudocode reads

```
backtrack(i, j, M){
  /* Check if (i, j, M) is an allowed triple,
     e.g. this includes
     -   M ≤ j-i
     -   i≤j
     -   if M == 0 then base i and base j
         must be complementary
     -   ...
    If (i, j, M) is not allowed, then E^M_{i,j} == ∞ */
  if( E^M_{i,j} < ∞){
    if( M == 0 ){
      /* i.e. base i and j are paired => add to list of
         base pairs */
      add_to_list_of_base_pairs(i, j);
      E_buffer = ∞;
      /* Determine the size m_min of the loop closed
         by the pair (i,j) */
      for( m = 0; m ≤ j-i; m++){
        if( E^m_{i+1,j-1} < E_buffer ){
          E_buffer = E^m_{i+1,j-1};
          m_min = m;
        }
      }
     /* Perform the next backtracking step */
     backtrack(i+1, j-1, m_min);
    }
    else if( M > 0 ){
      E_buffer = ∞;
      for( k = 0; k ≤ j; k++ ){
        /* Split the structure into two structures.
           The left structure ranges from i through k-1
           and has M-1 non-nested backbone bonds.
           The right structure ranges from k through j
           and has zero non-nested backbone bonds,
           i.e. it is either a single base (k==j) or is
           a closed structure terminated by the
           pair (k,j) */
        if( E^{M-1}_{i,k-1} + E^0_{k,j} < E_buffer ){
          E_buffer = E^{M-1}_{i,k-1} + E^0_{k,j};
```

```
          k_min = k;
        }
      }
      /* Perform next backtracking step
         on the left substructure */
     backtrack(i, k-1, M-1);
      /* Perform next backtracking step
         on the right substructure */
      if( k ≠ j ){
        backtrack(k, j, 0);
      }
    }
  }
}
```

B.5. Interpolation formulas

In order to make our results more user-friendly we give interpolation formulas for the salt corrections, which do not involve hypergeometric and Bessel functions as those might not be available for every user.

The salt correction for the free energy of loops contains two generalized hypergeometric functions

$$\begin{aligned}\mathcal{G}_{\mathrm{l}}^{\mathrm{salt}}(m) = \mathrm{k_B}T l_{\mathrm{B}}(m l_{\mathrm{ss}})\tau_{\mathrm{ss}}^2 \Bigg[& \ln(\kappa m l_{\mathrm{ss}}) - \ln(\pi/2) + \gamma \\ & - \frac{\kappa m l_{\mathrm{ss}}}{2} {}_1F_2\left(1/2, \begin{pmatrix}1\\3/2\end{pmatrix}, \left(\frac{\kappa m l_{\mathrm{ss}}}{2\pi}\right)^2\right) \\ & + \frac{1}{2}\left(\frac{\kappa m l_{\mathrm{ss}}}{\pi}\right)^2 {}_2F_3\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}3/2\\3/2\\2\end{pmatrix}, \left(\frac{\kappa m l_{\mathrm{ss}}}{2\pi}\right)^2\right) \\ & + \frac{1}{\kappa m l_{\mathrm{ss}}}\left(1 - \exp(-\kappa m l_{\mathrm{ss}}) + \kappa m l_{\mathrm{ss}}\Gamma(0, \kappa m l_{\mathrm{ss}})\right)\Bigg] . \end{aligned} \tag{B.21}$$

The sum of the two hypergeometric functions is well approximated by interpolating between the two asymptotic expansions for small and large argument

$$\begin{aligned} & -\frac{y}{2} {}_1F_2\left(1/2, \begin{pmatrix}1\\3/2\end{pmatrix}, \left(\frac{y}{2\pi}\right)^2\right) + \frac{1}{2}\left(\frac{y}{\pi}\right)^2 {}_2F_3\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}3/2\\3/2\\2\end{pmatrix}, \left(\frac{y}{2\pi}\right)^2\right) \\ & \approx \frac{1}{\frac{y^6}{(2\pi)^6}+1}\left(\frac{y^4}{36\pi^4} - \frac{y^3}{24\pi^2} + \frac{y^2}{2\pi^2} - \frac{y}{2}\right) \\ & \quad + \left(1 - \frac{1}{\frac{y^6}{(2\pi)^6}+1}\right)\left(\log\left(\frac{2\pi}{y}\right) - 1.96351\right) . \end{aligned} \tag{B.22}$$

The salt correction of the binding free energy of a base pair contains a modified Bessel function of the second kind, eq. (4.8),

$$\mathcal{G}_{\mathrm{h}}^{\mathrm{salt}}/h = g_{\mathrm{h}}^{\mathrm{DH}}(\rho) - g_{\mathrm{h}}^{\mathrm{DH}}(1\ \mathrm{M}) , \tag{B.23}$$

with

$$g_{\mathrm{h}}^{\mathrm{DH}}(\rho) = 2\mathrm{k_B}T\tau_{\mathrm{ds}}^2 l_{\mathrm{ds}} l_{\mathrm{B}} \mathrm{K}_0(\kappa d) . \tag{B.24}$$

Eq. (B.5) is well approximated by the heuristic formula

$$\mathcal{G}_{\mathrm{h}}^{\mathrm{salt}}/h \approx \mathrm{k_B}T\tau_{\mathrm{ds}}^2 l_{\mathrm{ds}} l_{\mathrm{B}}\left(b_1 + \frac{b_2 - b_3\ln(b_4\kappa)}{1 + b_5\kappa}\right) , \tag{B.25}$$

with the constants $b_1 = 0.0315171$, $b_2 = 35.0754$, $b_3 = 1.62292$, $b_4 = 1\,\mathrm{nm}$, $b_5 = 4.26381\,\mathrm{nm}$.

B.6. Free energy parameters

The enthalpy h and entropy s parameters used in our calculations are taken from reference [112, 122]. We summarize these parameters in the following tables. For instance, the entries in the row UA and the column GC, $h_{\mathrm{UA,GC}}$ and $s_{\mathrm{UA,GC}}$, give the enthalpy and entropy contribution due to the stacking of the two neighboring base pairs UA and GC, where U and C are located at the 5'-end. The two bottom rows contain the initiation and termination contribution to helices and loops. For instance, the total free enthalpy of the triple helix terminated by a hairpin loop of length $m = 6$, $\begin{smallmatrix}5'-\mathrm{CGA}-\\3'-\mathrm{GCU}-\end{smallmatrix} loop$, at 1 M NaCl is given by

$$\begin{aligned}\mathcal{G} &= \mathcal{G}_{\mathrm{h}}^{\mathrm{init}} + \mathcal{G}_{\mathrm{h}}^{\mathrm{stack}} + \mathcal{G}_{\mathrm{h}}^{\mathrm{term}} + \mathcal{G}_{\mathrm{l}}^{\mathrm{init}} + \mathcal{G}_{\mathrm{l}}^{\mathrm{conf}} \\ &= h_{\mathrm{h}}^{\mathrm{init}} - Ts_{\mathrm{h}}^{\mathrm{init}} + h_{\mathrm{CG,GC}} + h_{\mathrm{GC,AU}} - T(s_{\mathrm{CG,GC}} + s_{\mathrm{GC,AU}}) + h_{\mathrm{h}}^{\mathrm{term}} - Ts_{\mathrm{h}}^{\mathrm{term}} \\ &\quad + h_{\mathrm{l}}^{\mathrm{init}} - Ts_{\mathrm{l}}^{\mathrm{init}} - \mathrm{k_B}T \ln m^{-c} \,. \end{aligned} \tag{B.26}$$

Table B.1.: Enthalpy parameters for RNA [112, 122].

	Enthalpy h / (kcal/mol)					
	AU	UA	CG	GC	GU	UG
AU	−6.82	−9.38	−11.40	−10.48	−3.21	−8.81
UA	−7.69	−6.82	−12.44	−10.44	−6.99	−12.83
CG	−10.44	−10.48	−13.39	−10.64	−5.61	−12.11
GC	−12.40	−11.40	−14.88	−13.39	−8.33	−12.59
GU	−12.83	−8.81	−12.59	−12.11	−13.47	−14.59
UG	−6.99	−3.21	−8.33	−5.61	−9.26	−13.47
$h_{\mathrm{h}}^{\mathrm{init/term}}$	3.72	3.72	0.00	0.00	3.72	3.72
$h_{\mathrm{l}}^{\mathrm{init}}$	1.68	1.68	1.68	1.68	1.68	1.68

Table B.2.: Entropy parameters for RNA [112, 122].

	Entropy s / (10^{-3}kcal/(mol K))					
	AU	UA	CG	GC	GU	UG
AU	−19.0	−26.7	−29.5	−27.1	−8.6	−24.0
UA	−20.5	−19.0	−32.5	−26.9	−19.3	−37.3
CG	−26.9	−27.1	−32.7	−26.7	−13.5	−32.2
GC	−32.5	−29.5	−36.9	−32.7	−21.9	−32.5
GU	−37.3	−24.0	−32.5	−32.2	−44.9	−51.2
UG	−19.3	−8.6	−21.9	−13.5	−30.8	−44.9
$s_{\mathrm{h}}^{\mathrm{init/term}}$	10.5	10.5	0.0	0.0	10.5	10.5
$s_{\mathrm{l}}^{\mathrm{init}}$	−0.7	−0.7	−0.7	−0.7	−0.7	−0.7

B.7. Sequences

The sequence of yeast tRNA-phe reads [207]
GCGGAUUUAG CUCAGUUGGG AGAGCGCCAG ACUGAAGAUC UGGAGGUCCU
GUGUUCGAUC CACAGAAUUC GCACCA.

The sequence of the P5ab hairpin reads [125]
ACAGCCGUUC AGUACCAAGU CUCAGGGGAA ACUUUGAGAU GGGGUGCUGA CGGACA.

BIBLIOGRAPHY

[1] Thomas R. Einert, Paul Näger, Henri Orland, and Roland R. Netz. Impact of loop statistics on the thermodynamics of RNA folding. *Physical Review Letters*, 101:048103, 2008.

[2] Thomas R. Einert, Paul Näger, Henri Orland, and Roland R. Netz. Loop parameterization and RNA secondary structure folding. In Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty, Walter Nadler, and Olav Zimmermann, editors, *From Computational Biophysics to Systems Biology (CBSB08)*, volume 40 of *NIC Series*, pages 201–204, 2008.

[3] Thomas R. Einert, Douglas B. Staple, Hans-Jürgen Kreuzer, and Roland R. Netz. A three-state model with loop entropy for the over-stretching transition of DNA. *Biophysical Journal*, 99:578–587, 2010.

[4] Thomas R. Einert and Roland R. Netz. Theory for RNA folding, stretching, and melting including loops and salt. To be published, 2010.

[5] Thomas R. Einert, Henri Orland, and Roland R. Netz. Secondary structures of homopolymeric single-stranded nucleic acids including force and loop entropy: implications for hybridization. To be published, 2010.

[6] Thomas R. Einert, Charles E. Sing, Alfredo Alexander-Katz, and Roland R. Netz. Internal friction of homopolymeric systems studied by diffusion and non-equilibrium unfolding of globules. To be published, 2010.

[7] Charles E. Sing, Thomas R. Einert, Roland R. Netz, and Alfredo Alexander-Katz. Probing structural transitions in polymer globules with dynamic properties. To be published, 2010.

[8] Andreas Serr, Christian Sendner, Florian Müller, Thomas R. Einert, and Roland R. Netz. Single polymer adsorption in shear: flattening versus hydro-

dynamic lift and surface potential corrugation effects. *Europhysics Letters*, 92: 38002, 2010.

[9] Francis Crick. *What mad pursuit: a personal view of scientific discovery*. Alfred P. Sloan Foundation series. Basic Books, 1990.

[10] Johann Friedrich Miescher. Ueber die chemische Zusammensetzung der Eiterzellen. *Medicinisch-chemische Untersuchungen*, 4:441–460, 1871.

[11] Johann Friedrich Miescher. Die Spermatozoen einiger Wirbeltiere. Ein Beitrag zur Histochemie. *Verhandlungen der naturforschenden Gesellschaft in Basel*, VI: 138–208, 1874.

[12] Oswald T. Avery, Colin M. MacLeod, and Maclyn McCarty. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *The Journal of Experimental Medicine*, 79:137–158, 1944.

[13] A. D. Hershey and Martha Chase. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of General Physiology*, 36: 39–56, 1952.

[14] J. D. Watson and F. H. C. Crick. A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[15] Bruce Alberts. *Molecular Biology of the Cell*. Garland Science, 2002.

[16] Anthony J. F. Griffiths, Susan R. Wessler, Richard C. Lewontin, David T. Suzuki, and Jeffrey H. Miller. *Introduction to Genetic Analysis*. W. H. Freeman and Company, eighth edition, 2005.

[17] R. F. Gesteland, T. R. Cech, and J. F. Atkins, editors. *The RNA World*. Cold Spring Harbor Laboratory Press, Woodbury, second edition, 2005.

[18] Robert T. Batey, Sunny D. Gilbert, and Rebecca K. Montange. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, 432:411–415, 2004.

[19] Andrew Fire, SiQun Xu, Mary K. Montgomery, Steven A. Kostas, Samuel E. Driver, and Craig C. Mello. Potent and specific genetic interference by double-stranded RNA in caenorhabditis elegans. *Nature*, 391:806–811, 1998.

[20] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[21] Kathleen McAteer, Alejandro Aceves-Gaona, Ryszard Michalczyk, Garry W. Buchko, Nancy G. Isern, Louis A. “Pete” Silks Silks, John H. Miller, and

Michael A. Kennedy. Compensating bends in a 16-base-pair DNA oligomer containing a T3A3 segment: A NMR study of global DNA curvature. *Biopolymers*, 75:497–511, 2004.

[22] Nils E. Mikkelsen, Kenth Johansson, Anders Virtanen, and Leif A. Kirsebom. Aminoglycoside binding displaces a divalent metal ion in a tRNA-neomycin B complex. *Nature Structural Biology*, 8:510–514, 2001.

[23] Kevin Y. Sanbonmatsu, Simpson Joseph, and Chang-Shung Tung. Simulating movement of tRNA into the ribosome during decoding. *Proceedings of the National Academy of Sciences of the United States of America*, 102:15854–15859, 2005.

[24] Srinivasaraghavan Kannan and Martin Zacharias. Folding of a DNA hairpin loop structure in explicit solvent using replica-exchange molecular dynamics simulations. *Biophysical Journal*, 93:3218–3228, 2007.

[25] A. Alexander-Katz, M. F. Schneider, S. W. Schneider, A. Wixforth, and R. R. Netz. Shear-flow-induced unfolding of polymeric globules. *Physical Review Letters*, 97:138101, 2006.

[26] John F. Marko and Eric D. Siggia. Stretching DNA. *Macromolecules*, 28:8759–8770, 1995.

[27] Douglas Poland and Harold A. Scheraga. Phase transitions in one dimension and the helix – coil transition in polyamino acids. *Journal of Chemical Physics*, 45:1456–1463, 1966.

[28] Douglas Poland and Harold A. Scheraga. Occurrence of a phase transition in nucleic acid models. *Journal of Chemical Physics*, 45:1464–1469, 1966.

[29] Ignacio Tinoco, Jr and Carlos Bustamante. How RNA folds. *Journal of Molecular Biology*, 293:271–281, 1999.

[30] S. B. Smith, Y. J. Cui, and C. Bustamante. Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded dna molecules. *Science*, 271:795–799, 1996.

[31] Y. Kafri, D. Mukamel, and L. Peliti. Melting and unzipping of DNA. *European Physical Journal B*, 27:135–146, 2002.

[32] D. Bensimon, A. J. Simon, V. Croquette, and A. Bensimon. Stretching DNA with a receding meniscus: Experiments and models. *Physical Review Letters*, 74: 4754–4757, 1995.

[33] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay, and F. Caron. DNA: An extensible molecule. *Science*, 271:792794, 1996.

[34] Simona Cocco, Jie Yan, Jean-Francois Léger, Didier Chatenay, and John F. Marko. Overstretching and force-driven strand separation of double-helix DNA. *Physical Review E*, 70:011910, 2004.

[35] J. F. Léger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko. Structural transitions of a twisted and stretched DNA molecule. *Physical Review Letters*, 83:1066, 1999.

[36] Stephen Whitelam, Sander Pronk, and Phillip L. Geissler. There and (slowly) back again: Entropy-driven hysteresis in a model of DNA overstretching. *Biophysical Journal*, 94:2452–2469, 2008.

[37] M.W. Konrad and J.I. Bolonick. Molecular dynamics simulation of DNA stretching is consistent with the tension observed for extension and strand separation and predicts a novel ladder structure. *Journal of the American Chemical Society*, 118:10989–10994, 1996.

[38] Konstantin M. Kosikov, Andrey A. Gorin, Victor B. Zhurkin, and Wilma K. Olson. DNA stretching and compression: large-scale simulations of double helical structures. *Journal of Molecular Biology*, 289:1301–1326, 1999.

[39] H. Li and T. Gisler. Overstretching of a 30 bp DNA duplex studied with steered molecular dynamics simulation: Effects of structural defects on structure and force-extension relation. *European Physical Journal E*, 30:325–332, 2009.

[40] Anne Lebrun and Richard Lavery. Modelling extreme stretching of DNA. *Nucleic Acids Research*, 24:2260–2267, 1996.

[41] Mark C. Williams, Jay R. Wenner, Ioulia Rouzina, and Victor A. Bloomfield. Entropy and heat capacity of DNA melting from temperature dependence of single molecule stretching. *Biophysical Journal*, 80:1932–1939, 2001.

[42] Ioulia Rouzina and Victor A. Bloomfield. Force-induced melting of the DNA double helix 1. thermodynamic analysis. *Biophysical Journal*, 80:882–893, 2001.

[43] Leila Shokri, Micah J. McCauley, Ioulia Rouzina, and Mark C. Williams. DNA overstretching in the presence of glyoxal: Structural evidence of force-induced DNA melting. *Biophysical Journal*, 95:1248–1255, 2008.

[44] Joost van Mameren, Peter Gross, Geraldine Farge, Pleuni Hooijman, Mauro Modesti, Maria Falkenberg, Gijs J. L. Wuite, and Erwin J. G. Peterman. Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 106:18231–18236, 2009.

[45] Matthias Rief, Hauke Clausen-Schaumann, and Hermann E. Gaub. Sequence-dependent mechanics of single DNA molecules. *Nature Structural Biology*, 6: 346–349, 1999.

[46] Hauke Clausen-Schaumann, Matthias Rief, Carolin Tolksdorf, and Hermann E. Gaub. Mechanical stability of single dna molecules. *Biophysical Journal*, 78: 1997–2007, 2000.

[47] Rupert Krautbauer, Hauke Clausen-Schaumann, and Hermann E. Gaub. Cisplatin changes the mechanics of single DNA molecules. *Angewandte Chemie - International Edition*, 39:3912–3915, 2000.

[48] Christoph G. Baumann, Steven B. Smith, Victor A. Bloomfield, and Carlos Bustamante. Ionic effects on the elasticity of single DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 94:6185–6190, 1997.

[49] Stefano Piana. Structure and energy of a DNA dodecamer under tensile load. *Nucleic Acids Research*, 33:7029–7038, 2005.

[50] Simona Cocco, John F. Marko, and Rémi Monasson. Theoretical models for single-molecule DNA and RNA experiments: from elasticity to unzipping. *Comptes Rendus Physique*, 3:569–584, 2002.

[51] Michel Peyrard, Santiago Cuesta-López, and Guillaume James. Modelling DNA at the mesoscale: a challenge for nonlinear science? *Nonlinearity*, 21:T91–T100, 2008.

[52] Roger M. Wartell and Albert S. Benight. Thermal denaturation of DNA molecules: A comparison of theory with experiment. *Physics Reports*, 126:67–107, 1985.

[53] Pierre Cizeau and Jean-Louis Viovy. Modeling extreme extension of DNA. *Biopolymers*, 42:383–385, 1997.

[54] Amir Ahsan, Joseph Rudnick, and Robijn Bruinsma. Elasticity theory of the B-DNA to S-DNA transition. *Biophysical Journal*, 74:132–137, 1998.

[55] C. Storm and P. C. Nelson. Theory of high-force DNA stretching and overstretching. *Physical Review E*, 67:051906, 2003.

[56] John F. Marko. DNA under high tension: Overstretching, undertwisting, and relaxation dynamics. *Physical Review E*, 57:2134, 1998.

[57] Daniel Jost and Ralf Everaers. A unified Poland-Scheraga model of oligo- and polynucleotide DNA melting: Salt effects and predictive power. *Biophysical Journal*, 96:1056–1067, 2009.

[58] Andreas Hanke, Martha G. Ochoa, and Ralf Metzler. Denaturation transition of stretched DNA. *Physical Review Letters*, 100:018106, 2008.

[59] Joseph Rudnick and Tatiana Kuriabova. Effect of external stress on the thermal melting of DNA. *Physical Review E*, 77:051903, 2008.

[60] Thomas Garel and Henri Orland. Generalized Poland-Scheraga model for DNA hybridization. *Biopolymers*, 75:453–467, 2004.

[61] R. D. Blake, J. W. Bizzaro, J.D. Blake, G. R. Day, S. G. Delcourt, J. Knowles, K. A. Marx, and J. SantaLucia, Jr. Statistical mechanical simulation of polymeric DNA melting with MELTSIM. *Bioinformatics*, 15:370–375, 1999.

[62] Enrico Carlon, Enzo Orlandini, and Attilio L. Stella. Roles of stiffness and excluded volume in DNA denaturation. *Physical Review Letters*, 88:198101, 2002.

[63] Ralf Everaers, Sanjay Kumar, and Christian Simm. Unified description of poly- and oligonucleotide DNA melting: Nearest-neighbor, Poland-Sheraga, and lattice models. *Physical Review E*, 75:041918, 2007.

[64] Sahand Jamal Rahi, Mark Peter Hertzberg, and Mehran Kardar. Melting of persistent double-stranded polymers. *Physical Review E*, 78:051910, 2008.

[65] J. Palmeri, M. Manghi, and N. Destainville. Thermal denaturation of fluctuating DNA driven by bending entropy. *Physical Review Letters*, 99:088103, 2007.

[66] Jae-Hyung Jeon, Wokyung Sung, and Francis H. Ree. A semiflexible chain model of local denaturation in double-stranded DNA. *Journal of Chemical Physics*, 124:164905, 2006.

[67] Michel Peyrard. Nonlinear dynamics and statistical physics of DNA. *Nonlinearity*, 17:R1–R40, 2004.

[68] Dinko Cule and Terence Hwa. Denaturation of heterogeneous DNA. *Physical Review Letters*, 79:2375, 1997.

[69] Dominik Ho, Julia L. Zimmermann, Florian A. Dehmelt, Uta Steinbach, Matthias Erdmann, Philip Severin, Katja Falter, and Hermann E. Gaub. Force-driven separation of short double-stranded DNA. *Biophysical Journal*, 97:3158–3167, 2009.

[70] A. Bar, Y. Kafri, and D. Mukamel. Loop dynamics in DNA denaturation. *Physical Review Letters*, 98:038103, 2007.

[71] Roland R. Netz. Strongly stretched semiflexible extensible polyelectrolytes and DNA. *Macromolecules*, 34:7522–7529, 2001.

[72] L. Livadaru, R. R. Netz, and H. J. Kreuzer. Stretching response of discrete semiflexible polymers. *Macromolecules*, 36:3732–3744, 2003.

[73] Theo Odijk. Stiff chains and filaments under tension. *Macromolecules*, 28: 7016–7018, 1995.

[74] Pierre-Gilles de Gennes. *Scaling Concepts in Polymer Physics*. Cornell University Press, 1979.

[75] Bertrand Duplantier. Polymer network of fixed topology: Renormalization, exact critical exponent γ in two dimensions, and $d = 4-\epsilon$. *Physical Review Letters*, 57:941–944, 1986.

[76] Arthur Erdélyi. *Higher Transcendental Functions*, volume 1. McGraw-Hill, 1953.

[77] George B. Arfken and Hans J. Weber. *Mathematical Methods for Physicists*. Academic Press, fifth edition, 2001.

[78] Philippe Flajolet and Andrew Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3:216–240, 1990.

[79] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, first edition, 2009.

[80] Jay R. Wenner, Mark C. Williams, Ioulia Rouzina, and Victor A. Bloomfield. Salt dependence of the elasticity and overstretching transition of single DNA molecules. *Biophysical Journal*, 82:3160–3169, 2002.

[81] Thorsten Hugel, Matthias Rief, Markus Seitz, Hermann E. Gaub, and Roland R. Netz. Highly stretched single polymers: Atomic-force-microscope experiments versus ab-initio theory. *Physical Review Letters*, 94:048301, 2005.

[82] M.-N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette. Stretching single stranded DNA, a model polyelectrolyte. *Phys. Rev. Lett.*, 89: 248102, 2002.

[83] M.C. Murphy, Ivan Rasnik, Wei Cheng, Timothy M. Lohman, and Taekjip Ha. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophysical Journal*, 86:2530–2537, 2004.

[84] John SantaLucia, Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor-thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 95:1460–1465, 1998.

[85] Osamu Gotoh, Yuzuru Husimi, Sadato Yabuki, and Akiyoshi Wada. Hyperfine structure in melting profile of bacteriophage lambda DNA. *Biopolymers*, 15: 655–670, 1976.

[86] Ralf Blossey and Enrico Carlon. Reparametrizing the loop entropy weights: Effect on DNA melting curves. *Physical Review E*, 68:061911, 2003.

[87] Ignacio Tinoco, Olke C. Uhlenbeck, and Mark D. Levine. Estimation of secondary structure in ribonucleic acids. *Nature*, 230:362–367, 1971.

[88] A. V. Finkelstein and O. V. Galzitskaya. Physics of protein folding. *Physics of Life Reviews*, 1:23–56, 2004.

[89] Pierre-Gilles de Gennes. Statistics of branching and hairpin helices for the dAT copolymer. *Biopolymers*, 6:715–729, 1968.

[90] M. S. Waterman and T. F. Smith. RNA secondary structure: A complete mathematical analysis. *Mathematical Biosciences*, 42:257–266, 1978.

[91] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie*, 125:167–188, 1994.

[92] J. S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.

[93] Andrea Montanari and Marc Mézard. Hairpin formation and elongation of biomolecules. *Physical Review Letters*, 86:2178–2181, 2001.

[94] Ulrich Gerland, Ralf Bundschuh, and Terence Hwa. Force-induced denaturation of RNA. *Biophysical Journal*, 81:1324–1332, 2001.

[95] M. Müller, F. Krzakala, and M. Mézard. The secondary structure of RNA under tension. *European Physical Journal E*, 9:67–77, 2002.

[96] David K. Lubensky and David R. Nelson. Single molecule statistics and the polynucleotide unzipping transition. *Physical Review E*, 65:031917, 2002.

[97] Ralf Bundschuh and Ulrich Gerland. Coupled dynamics of RNA folding and nanopore translocation. *Physical Review Letters*, 95:208104, 2005.

[98] Zhi-Jie Tan and Shi-Jie Chen. Nucleic acid helix stability: Effects of salt concentration, cation valence and size, and chain length. *Biophysical Journal*, 90: 1175–1190, 2006.

[99] Yevgeni Sh. Mamasakhlisov, Shura Hayryan, V. F. Morozov, and Chin-Kun Hu. RNA folding in the presence of counterions. *Physical Review E*, 75:061907, 2007.

[100] M. Baiesi, E. Orlandini, and A. L. Stella. RNA denaturation: Excluded volume, pseudoknots, and transition scenarios. *Physical Review Letters*, 91:198102, 2003.

[101] Henri Orland and A. Zee. RNA folding and large N matrix theory. *Nuclear Physics B*, 620:456–476, 2002.

[102] Yariv Kafri, David Mukamel, and Luca Peliti. Why is the DNA denaturation transition first order? *Physical Review Letters*, 85:4988, 2000.

[103] Jan Liphardt, Sophie Dumont, Steven B. Smith, Jr. Tinoco, Ignacio, and Carlos Bustamante. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*, 296:1832–1835, 2002.

[104] Berenike Maier, David Bensimon, and Vincent Croquette. Replication by a single DNA polymerase of a stretched single-stranded DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 97:12002–12007, 2000.

[105] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot. Molecular stick-slip motion revealed by opening DNA with piconewton forces. *Physical Review Letters*, 79: 4489–4492, 1997.

[106] Alessandro Mossa, Maria Manosas, Nuria Forns, Josep Maria Huguet, and Felix Ritort. Dynamic force spectroscopy of DNA hairpins. I. force kinetics and free energy landscapes. *Journal of Statistical Mechanics*, 2009:P02060, 2009.

[107] J. Christof M. Gebhardt, Thomas Bornschlögl, and Matthias Rief. Full distance-resolved folding energy landscape of one single protein molecule. *Proceedings of the National Academy of Sciences of the United States of America*, 107:2013–2018, 2010.

[108] E. G. Richards. 5S RNA. an analysis of possible base pairing schemes. *European Journal of Biochemistry*, 10:36–42, 1969.

[109] F. H. D. van Batenburg, A. P. Gultyaev, C. W. A. Pleij, J. Ng, and J. Oliehoek. PseudoBase: A database with RNA pseudoknots. *Nucleic Acids Research*, 28: 201–204, 2000.

[110] Paul G. Higgs. RNA secondary structure: Physical and computational aspects. *Quarterly Reviews of Biophysics*, 33:199–253, 2000.

[111] Samuel S. Cho, David L. Pincus, and D. Thirumalai. Assembly mechanisms of RNA pseudoknots are determined by the stabilities of constituent secondary structures. *Proceedings of the National Academy of Sciences of the United States of America*, 106:17349–17354, 2009.

[112] Tianbing Xia, John SantaLucia, Jr, Mark E. Burkard, Ryszard Kierzek, Susan J. Schroeder, Xiaoqi Jiao, Christopher Cox, and Douglas H. Turner. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, 37:14719–14735, 1998.

[113] R. Bundschuh and T. Hwa. Phases of the secondary structures of RNA sequences. *Europhysics Letters*, 59:903–909, 2002.

[114] Alexander Yu. Grosberg and Alexei R. Khokhlov. *Statistical Physics of Macromolecules*. American Institute of Physics, first edition, 2002.

[115] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of Mathematical Functions*. U.S. Department of Commerce, tenth edition, 2002.

[116] M. Müller. Statistical physics of RNA folding. *Physical Review E*, 67:021914, 2003.

[117] Ivo L. Hofacker. Vienna RNA secondary structure server. *Nucleic Acids Research*, 31:3429–3431, 2003.

[118] Nicholas R. Markham and Michael Zuker. Dinamelt web server for nucleic acid melting prediction. *Nucleic Acids Research*, 33:W577–581, 2005.

[119] Peter Schuster. Prediction of RNA secondary structures: From theory to models and real molecules. *Reports on Progress in Physics*, 69:1419–1477, 2006.

[120] Tim Liedl, Thomas L. Sobey, and Friedrich C. Simmel. Dna-based nanodevices. *Nano Today*, 2:36–41, 2007.

[121] Hendrik Dietz, Shawn M. Douglas, and William M. Shih. Folding DNA into twisted and curved nanoscale shapes. *Science*, 325:725–730, 2009.

[122] David H. Mathews, Jeffrey Sabina, Michael Zuker, and Douglas H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology*, 288:911–940, 1999.

[123] P. L. Privalov and V. V. Filimonov. Thermodynamic analysis of transfer RNA unfolding. *Journal of Molecular Biology*, 122:447–464, 1978.

[124] M. Vives, R. Tauler, and R. Gargallo. Study of the influence of metal ions on tRNAPhe thermal unfolding equilibria by UV spectroscopy and multivariate curve resolution. *Journal of Inorganic Biochemistry*, 89:115–122, 2002.

[125] Jan Liphardt, Bibiana Onoa, Steven B. Smith, Ignacio Tinoco, Jr, and Carlos Bustamante. Reversible unfolding of single RNA molecules by mechanical force. *Science*, 292:733–737, 2001.

[126] Jeffrey Vieregg, Wei Cheng, Carlos Bustamante, and Ignacio Tinoco. Measurement of the effect of monovalent cations on RNA hairpin stability. *Journal of the American Chemical Society*, 129:14966–14973, 2007.

[127] Josep M. Huguet, Cristiano V. Bizarro, Núria Forns, Steven B. Smith, Carlos Bustamante, and Felix Ritort. Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 107:15431–15436, 2010.

[128] Michael Zuker and Patrick Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9:133–148, 1981.

[129] Roumen A. Dimitrov and Michael Zuker. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophysical Journal*, 87:215–226, 2004.

[130] A. Imparato, A. Pelizzola, and M. Zamparo. Equilibrium properties and force-driven unfolding pathways of RNA molecules. *Physical Review Letters*, 103: 188102, 2009.

[131] Ulrich Gerland, Ralf Bundschuh, and Terence Hwa. Mechanically probing the folding pathway of single RNA molecules. *Biophysical Journal*, 84:2831–2840, 2003.

[132] C. Hyeon and D. Thirumalai. Mechanical unfolding of RNA hairpins. *Proceedings of the National Academy of Sciences of the United States of America*, 102: 6789–6794, 2005.

[133] Changbong Hyeon and D. Thirumalai. Forced-unfolding and force-quench refolding of RNA hairpins. *Biophysical Journal*, 90:3410–3427, 2006.

[134] Pascal Auffinger and Eric Westhof. Water and ion binding around rna and dna (c,g) oligomers. *Journal of Molecular Biology*, 300:1113–1131, 2000.

[135] Zhi-Jie Tan and Shi-Jie Chen. RNA helix stability in mixed Na+/Mg2+ solution. *Biophysical Journal*, 92:3615–3632, 2007.

[136] Zhi-Jie Tan and Shi-Jie Chen. Salt dependence of nucleic acid hairpin stability. *Biophysical Journal*, 95:738–752, 2008.

[137] R. Owczarzy, Y. You, B.G. Moreira, J.A. Manthey, L. Huang, M.A. Behlke, and J.A. Walder. Effects of sodium ions on DNA duplex oligomers: Improved predictions of melting temperatures. *Biochemistry*, 43:3537–3554, 2004.

[138] Gerald S. Manning. Limiting laws and counterion condensation in polyelectrolyte solutions I. colligative properties. *Journal of Chemical Physics*, 51:924–933, 1969.

[139] Martin J. Serra and Douglas H. Turner. Predicting thermodynamic properties of RNA. *Methods in Enzymology*, 259:242–261, 1995.

[140] K.-K. Kunze and R. R. Netz. Complexes of semiflexible polyelectrolytes and charged spheres as models for salt-modulated nucleosomal structures. *Physical Review E*, 66:011918, 2002.

[141] John Norman Murrell. *Properties of liquids and solutions*. John Wiley & Sons, Inc, 1994.

[142] R. R. Netz and H. Orland. Variational charge renormalization in charged systems. *European Physical Journal E*, 11:301–311, 2003.

[143] David E. Draper, Dan Grilley, and Ana Maria Soto. Ions and RNA folding. *Annual Review of Biophysics and Biomolecular Structure*, 34:221–243, 2005.

[144] Andreas R. Gruber, Ronny Lorenz, Stephan H. Bernhart, Richard Neubock, and Ivo L. Hofacker. The Vienna RNA websuite. *Nucleic Acids Research*, gkn188:1–5, 2008.

[145] Pan T.X. Li, Jeffrey Vieregg, and Ignacio Tinoco. How RNA unfolds and refolds. *Annual Review of Biochemistry*, 77:77–100, 2008.

[146] M. Manosas, D. Collin, and F. Ritort. Force-dependent fragility in RNA hairpins. *Physical Review Letters*, 96:218301, 2006.

[147] Yeonee Seol, Gary M. Skinner, Koen Visscher, Arnaud Buhot, and Avraham Halperin. Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Physical Review Letters*, 98:158103, 2007.

[148] Ignacio Tinoco, Jr. Force as a useful variable in reactions: Unfolding RNA. *Annual Review of Biophysics and Biomolecular Structure*, 33:363–385, 2004.

[149] Michael T. Woodside, Peter C. Anthony, William M. Behnke-Parks, Kevan Larizadeh, Daniel Herschlag, and Steven M. Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314:1001–1004, 2006.

[150] Michael T. Woodside, William M. Behnke-Parks, Kevan Larizadeh, Kevin Travers, Daniel Herschlag, and Steven M. Block. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proceedings of the National Academy of Sciences of the United States of America*, 103:6190–6195, 2006.

[151] Sanjay Kumar and Garima Mishra. Force-induced stretched state: Effects of temperature. *Physical Review E*, 78:011907, 2008.

[152] Jin-Der Wen, Maria Manosas, Pan T. X. Li, Steven B. Smith, Carlos Bustamante, Felix Ritort, and Ignacio Tinoco, Jr. Force Unfolding Kinetics of RNA Using Optical Tweezers. I. Effects of Experimental Variables on Measured Results. *Biophysical Journal*, 92:2996–3009, 2007.

[153] S. Cocco, J.F. Marko, and R. Monasson. Slow nucleic acid unzipping kinetics from sequence-defined barriers. *European Physical Journal E*, 10:153–161, 2003.

[154] Nadine Schwierz, Dominik Horinek, and Roland R. Netz. Reversed anionic Hofmeister series: The interplay of surface charge and surface polarity. *Langmuir*, 26:7370–7379, 2010.

[155] Jeffrey Chuang, Yacov Kantor, and Mehran Kardar. Anomalous dynamics of translocation. *Physical Review E*, 65:011802, 2001.

[156] K. Luo, T. Ala-Nissila, S.-C. Ying, and R. Metzler. Driven polymer translocation through nanopores: Slow-vs.-fast dynamics. *Europhysics Letters*, 88:68006, 2009.

[157] Nikhil Gunari, Anna C. Balazs, and Gilbert C. Walker. Force-induced globule-coil transition in single polystyrene chains in water. *Journal of the American Chemical Society*, 129:10046–10047, 2007.

[158] Matthias Rief, Mathias Gautel, Filipp Oesterhelt, Julio M. Fernandez, and Hermann E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276:1109–1112, 1997.

[159] Douglas B. Staple, Stephen H. Payne, Andrew L. C. Reddin, and Hans Jürgen Kreuzer. Model for stretching and unfolding the giant multidomain muscle protein using single-molecule force spectroscopy. *Physical Review Letters*, 101: 248301, 2008.

[160] Maumita Mandal and Ronald R. Breaker. Gene regulation by riboswitches. *Nature Reviews Molecular Cell Biology*, 5:451–463, 2004.

[161] P. G. de Gennes. Reptation of a polymer chain in presence of fixed obstacles. *Journal of Chemical Physics*, 55:572, 1971.

[162] G. S. Jas, W. A. Eaton, and J. Hofrichter. Effect of viscosity on the kinetics of alpha-helix and beta-hairpin formation. *Journal of Physical Chemistry B*, 105: 261–272, 2001.

[163] Suzette A. Pabit, Heinrich Roder, and Stephen J. Hagen. Internal friction controls the speed of protein folding from a compact configuration. *Biochemistry*, 43:12532–12538, 2004.

[164] Frauke Graeter, Pascal Heider, Ronen Zangi, and B. J. Berne. Dissecting entropic coiling and poor solvent effects in protein collapse. *Journal of the American Chemical Society*, 130:11578–11579, 2008.

[165] D. E. Sagnella, J. E. Straub, and D. Thirumalai. Time scales and pathways for kinetic energy relaxation in solvated proteins: Application to carbonmonoxy myoglobin. *Journal of Chemical Physics*, 113:7702–7711, 2000.

[166] A. Y. Grosberg, S. K. Nechaev, and E. I. Shakhnovich. The role of topological constraints in the kinetics of collapse of macromolecules. *Journal de Physique*, 49:2095–2100, 1988.

[167] A. Grosberg, Y. Rabin, S. Havlin, and A. Neer. Crumpled globule model of the three-dimensional structure of DNA. *Europhysics Letters*, 23:373, 1993.

[168] Erez Lieberman-Aiden, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R. Lajoie, Peter J. Sabo, Michael O. Dorschner, Richard Sandstrom, Bradley Bernstein, M. A. Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A. Mirny, Eric S. Lander, and Job Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326:289–293, 2009.

[169] Michael G. Poirier, Ajay Nemani, Prateek Gupta, Sertac Eroglu, and John F. Marko. Probing chromosome structure with dynamic force relaxation. *Physical Review Letters*, 86:360–363, 2001.

[170] Thomas Frisch and Alberto Verga. Unwinding globules under tension and polymer collapse. *Physical Review E*, 65:041801, 2002.

[171] Natalia A. Denesyuk and John D. Weeks. Equilibrium and nonequilibrium effects in the collapse of a model polypeptide. *Physical Review Letters*, 102: 108101, 2009.

[172] C. F. Abrams, N.-K. Lee, and S. P. Obukhov. Collapse dynamics of a polymer chain: Theory and simulation. *Europhysics Letters*, 59:391–397, 2002.

[173] Franck Celestini, Thomas Frisch, and Xabier Oyharcabal. Stretching an adsorbed polymer globule. *Physical Review E*, 70:012801, 2004.

[174] J. Chuang, A. Y. Grosberg, and T. Tanaka. Topological repulsion between polymer globules. *Journal of Chemical Physics*, 112:6434–6442, 2000.

[175] V. Barsegov, G. Morrison, and D. Thirumalai. Role of internal chain dynamics on the rupture kinetic of adhesive contacts. *Physical Review Letters*, 100:248102, 2008.

[176] Bhavin S. Khatri, Masaru Kawakami, Katherine Byrne, D. Alastair Smith, and Tom C. B. McLeish. Entropy and barrier-controlled fluctuations determine conformational viscoelasticity of single biomolecules. *Biophysical Journal*, 92: 1825–1835, 2007.

[177] Yann von Hansen, Felix Sedlmeier, Michael Hinczewski, and Roland R. Netz. Friction contribution to water-bond breakage kinetics. To be published, 2010.

[178] A. E. Filippov, J. Klafter, and M. Urbakh. Friction through dynamical formation and rupture of molecular bonds. *Physical Review Letters*, 92:135503, 2004.

[179] Y. Murayama, H. Wada, and M. Sano. Dynamic force spectroscopy of a single condensed DNA . *Europhysics Letters*, 79:58001, 2007.

[180] Remo Gerber, Abdessamad Tahiri-Alaoui, P. J. Hore, and William James. Oligomerization of the human prion protein proceeds via a molten globule intermediate. *Journal of Biological Chemistry*, 282:6300–6307, 2007.

[181] R. Zwanzig. Diffusion in a rough potential. *Proceedings of the National Academy of Sciences of the United States of America*, 85:2029–2030, 1988.

[182] Changbong Hyeon and D. Thirumalai. Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? *Proceedings of the National Academy of Sciences of the United States of America*, 100: 10249–10253, 2003.

[183] Alfredo Alexander-Katz, Hirofumi Wada, and Roland R. Netz. Internal friction and non-equilibrium unfolding of homo-polymeric globules. *Physical Review Letters*, 103:028102, 2009.

[184] M. Hinczewski, Y. von Hansen, J. Dzubiella, and R. R. Netz. How the diffusivity profile reduces the arbitrariness of protein folding free energies. *Journal of Chemical Physics*, 132:245103, 2010.

[185] R. B. Best and G. Hummer. Coordinate-dependent diffusion in protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 107:1088–1093, 2010.

[186] D. K. Klimov and D. Thirumalai. Stretching single-domain proteins: Phase diagram and kinetics of force-induced unfolding. *Proceedings of the National Academy of Sciences of the United States of America*, 96:6166–6170, 1999.

[187] N. Yoshinaga, K. Yoshikawa, and T. Ohta. Different pathways in mechanical unfolding/folding cycle of a single semiflexible polymer. *European Physical Journal E*, 17:485–491, 2005.

[188] Greg Morrison, Changbong Hyeon, N. M. Toan, Bae-Yeun Ha, and D. Thirumalai. Stretching homopolymers. *Macromolecules*, 40:7343–7353, 2007.

[189] O. Braun and U. Seifert. Periodically driven stochastic un- and refolding transitions of biopolymers. *Europhysics Letters*, 68:746–752, 2004.

[190] R. Metzler, W. Reisner, R. Riehn, R. Austin, J. O. Tegenfeldt, and I. M. Sokolov. Diffusion mechanisms of localised knots along a polymer. *Europhysics Letters*, 76:696–702, 2006.

[191] Alexander Vologodskii. Brownian dynamics simulation of knot diffusion along a stretched DNA molecule. *Biophysical Journal*, 90:1594–1597, 2006.

[192] Lei Huang and Dmitrii E. Makarov. Langevin dynamics simulations of the diffusion of molecular knots in tensioned polymer chains. *Journal of Physical Chemistry A*, 111:10338–10344, 2007.

[193] Drew F. Parsons and David R. M. Williams. Globule transitions of a single homopolymer: A Wang-Landau Monte Carlo study. *Physical Review E*, 74:041804, 2006.

[194] Drew F. Parsons and David R. M. Williams. An off-lattice Wang-Landau study of the coil-globule and melting transitions of a flexible homopolymer. *Journal of Chemical Physics*, 124:221103, 2006.

[195] W. Paul, T. Strauch, F. Rampf, and K. Binder. Unexpectedly normal phase behavior of single homopolymer chains. *Physical Review E*, 75:060801, 2007.

[196] F. Rampf, W. Paul, and K. Binder. On the first-order collapse transition of a three-dimensional, flexible homopolymer chain model. *Europhysics Letters*, 70: 628, 2005.

[197] Yaoqi Zhou, Carol K. Hall, and Martin Karplus. First-order disorder-to-order transition in an isolated homopolymer model. *Physical Review Letters*, 77:2822–2825, 1996.

[198] Mark P. Taylor, Wolfgang Paul, and Kurt Binder. Phase transitions of a single polymer chain: A Wang?Landau simulation study. *Journal of Chemical Physics*, 131:114907, 2009.

[199] V. G. Rostiashvili, G. Migliorini, and T. A. Vilgis. Self-generated disorder and structural glass formation in homopolymer globules. *Physical Review E*, 64: 051112, 2001.

[200] Haojun Liang and Hanning Chen. First-order transition of a homopolymer chain with lennard-jones potential. *Journal of Chemical Physics*, 113:4469, 2000.

[201] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*. Clarendon Press, 1999.

[202] William T. Coffey. *The Langevin Equation*. World Scientific, 2005.

[203] Donald L. Ermak and J. A. McCammon. Brownian dynamics with hydrodynamic interactions. *Journal of Chemical Physics*, 69:1352, 1978.

[204] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 1989.

[205] Dennis Vitkup, Dagmar Ringe, Gregory A. Petsko, and Martin Karplus. Solvent mobility and the protein 'glass' transition. *Nature Structural Biology*, 7:34–38, 2000.

[206] M. Bon and H. Orland. Prediction of RNA secondary structures with pseudoknots. *Physica A*, 389:2987 – 2992, 2010. Statistical, Fluid and Biological Physics Problems - An MIT Symposium and Articles dedicated to A. Nihat Berker on his 60th Birthday.

[207] Huijing Shi and Peter B. Moore. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: A classic structure revisited. *RNA*, 6:1091–1105, 2000.