

---

# Thermodynamics of Molecular Adsorption Processes on Mesoscopic Scales

---

- diploma thesis -

submitted by

**Monika Möddel**

May 13, 2008

supervisor : Prof. Dr. Wolfhard Janke  
Dr. Michael Bachmann

referee : Prof. Dr. Wolfhard Janke  
Prof. Dr. Anders Irbäck

Institut für Theoretische Physik  
Fakultät für Physik und Geowissenschaften  
Universität Leipzig





# Contents

<b>Overview</b>	<b>1</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Motivation . . . . .	3
1.2 Coarse-graining . . . . .	7
<b>2 The Model</b>	<b>11</b>
2.1 Definition of the Polymer . . . . .	11
2.2 The Surface Potential . . . . .	13
2.3 Definition of Measured Observables . . . . .	15
2.4 Boundary Conditions . . . . .	18
<b>3 Monte Carlo Simulations</b>	<b>21</b>
3.1 Generation of Relevant Configurations . . . . .	22
3.1.1 Spherical update – Forwards (F) and Backwards (B) . . . . .	24
3.1.2 Semi-Local Update (L) . . . . .	25
3.1.3 Rigid-Body Rotation (R) . . . . .	25
3.1.4 Translation (T) . . . . .	26
3.1.5 Sweep Sequence . . . . .	26
3.2 Multicanonical Sampling . . . . .	27
3.2.1 Multicanonical Recursion . . . . .	28
3.3 Reweighting . . . . .	30
3.3.1 Reweighting in $T$ . . . . .	30
3.3.2 Reweighting in $\epsilon_s$ . . . . .	31
3.3.3 Reweighting in $s$ . . . . .	32
3.4 Replica Exchange Monte Carlo . . . . .	32
3.5 Energy Landscape Paving (ELP) . . . . .	34
3.6 Error estimation . . . . .	34
3.6.1 Autocorrelation Time . . . . .	35
3.6.2 Blocking Jackknife Technique . . . . .	36
<b>4 Results</b>	<b>39</b>
4.1 Adsorption Behaviour for Various Surface Attraction Strengths . . . . .	39
4.1.1 Observables . . . . .	39
4.1.2 The Pseudophase Diagram . . . . .	55
4.1.3 Data for $\epsilon_s = 2$ and $N = 20$ . . . . .	57
4.1.4 Density of States . . . . .	59
4.1.5 Comparison with Lattice Results . . . . .	62
4.1.6 Low-Energy Configurations . . . . .	65

4.2 Behaviour at Various Monomer-Monomer Attraction Strengths .	67
<b>5 Summary &amp; Outlook</b>	<b>71</b>
<b>Bibliography</b>	<b>79</b>

# List of Figures

1.1	Lattice model of the excluded volume chain. . . . .	4
1.2	Transition from the chemical polymer to the coarse-grained model . . . . .	8
2.1	A segment of the semiflexible polymer model. . . . .	11
2.2	Lennard-Jones Potential in different solutions. . . . .	12
2.3	Bending energy for one bonding angle . . . . .	12
2.4	Integration over all surface layers. . . . .	14
2.5	The surface potential . . . . .	15
2.6	Boundary conditions . . . . .	18
2.7	$\langle z_{\text{cm}} \rangle$ and $d\langle z_{\text{cm}} \rangle/dT$ of the 13mer in different boxes . . . . .	19
2.8	$\langle R_{\text{gyr}} \rangle$ of the 13mer in different boxes . . . . .	19
3.1	The spherical update . . . . .	24
3.2	The semi-local update . . . . .	25
3.3	Schematic representation of the block Jackknife error estimation . . . . .	37
4.1	$\langle E \rangle$ and $\langle C \rangle$ of the 20mer – 3D plot . . . . .	40
4.2	$\langle E \rangle$ and $\langle C \rangle$ of the 13mer . . . . .	40
4.3	$\langle E \rangle$ and $\langle C \rangle$ of the 20mer . . . . .	40
4.4	$\langle R_{\text{gyr}} \rangle$ and $d\langle R_{\text{gyr}} \rangle/dT$ of the 20mer – 3D . . . . .	42
4.5	$\langle R_{\text{gyr}} \rangle$ and $d\langle R_{\text{gyr}} \rangle/dT$ of the 13mer . . . . .	42
4.6	$\langle R_{\text{gyr}} \rangle$ and $d\langle R_{\text{gyr}} \rangle/dT$ of the 20mer . . . . .	42
4.7	$\langle R_{\parallel} \rangle$ and $d\langle R_{\text{gyr},\parallel} \rangle/dT$ of the 20mer – 3D . . . . .	44
4.8	$\langle R_{\parallel} \rangle$ and $d\langle R_{\text{gyr},\parallel} \rangle/dT$ of the 13mer . . . . .	44
4.9	$\langle R_{\parallel} \rangle$ and $d\langle R_{\text{gyr},\parallel} \rangle/dT$ of the 20mer . . . . .	44
4.10	Exemplified compact self-avoiding polymer on a 3D lattice . . . . .	45
4.11	$l = (1 - \epsilon_s/\epsilon_m)^{2/3} N^{1/3}$ vs. $\epsilon_s/\epsilon_m$ for four different chain length . . . . .	45
4.12	$\langle R_{\perp} \rangle$ and $d\langle R_{\text{gyr},\perp} \rangle/dT$ of the 20mer – 3D . . . . .	46
4.13	$\langle R_{\perp} \rangle$ and $d\langle R_{\text{gyr},\perp} \rangle/dT$ of the 13mer . . . . .	46
4.14	$\langle R_{\perp} \rangle$ and $d\langle R_{\text{gyr},\perp} \rangle/dT$ of the 20mer . . . . .	46
4.15	Compact low energy states of the 13mer for $s = 1$ and $s = 2$ . . . . .	48
4.16	$\langle z_{\text{cm}} \rangle$ and $d\langle z_{\text{cm}} \rangle/dT$ of the 20mer – 3D . . . . .	49
4.17	$\langle z_{\text{cm}} \rangle$ and $d\langle z_{\text{cm}} \rangle/dT$ of the 13mer . . . . .	49
4.18	$\langle z_{\text{cm}} \rangle$ and $d\langle z_{\text{cm}} \rangle/dT$ of the 20mer . . . . .	49
4.19	$\langle n_s \rangle$ and $d\langle n_s \rangle/dT$ of the 20mer – 3D . . . . .	52
4.20	$\langle n_s \rangle$ and $d\langle n_s \rangle/dT$ of the 13mer . . . . .	52
4.21	$\langle n_s \rangle$ and $d\langle n_s \rangle/dT$ of the 20mer . . . . .	52
4.22	$\langle n_s \rangle$ vs. $\epsilon_s$ for small $T$ . . . . .	53
4.23	$\langle n_m \rangle$ and $d\langle n_m \rangle/dT$ of the 20mer – 3D . . . . .	54

4.24	$\langle n_m \rangle$ and $d \langle n_m \rangle / dT$ of the 13mer . . . . .	54
4.25	$\langle n_m \rangle$ and $d \langle n_m \rangle / dT$ of the 20mer . . . . .	54
4.26	Phase diagram of the 13mer and 20mer . . . . .	55
4.27	Representative examples of conformation of the different regions .	56
4.28	Observables for $\epsilon_s = 2$ and $N = 20$ and their derivatives w.r.t. $T$	58
4.29	Density of states $\Omega(E)$ for three different $N$ and various $\epsilon_s$ . . . .	60
4.30	$p(E) = \Omega(E) \exp(-E/T)$ for three different $N$ and various $\epsilon_s$ . .	61
4.31	$s$ - $T$ pseudo-phase diagram of a 179mer on the lattice . . . . .	62
4.32	Examples of conformation on the lattice . . . . .	63
4.33	Layering regime on the lattice . . . . .	65
4.34	$T = 0$ energy composition for the 13mer . . . . .	66
4.35	$\Omega(E)$ of the 13mer for various $s$ and $\epsilon_s = 1$ . . . . .	67
4.36	$\Omega(E)$ of the 13mer for various $s$ and $\epsilon_s = 1$ . . . . .	67
4.37	$\langle R_{\text{gyr}} \rangle$ vs. $s$ and $T$ . . . . .	68

# Overview

This diploma thesis deals with the canonical equilibrium behaviour of a semiflexible homopolymer near an attractive substrate within the frame of a mesoscopic coarse-grained model.

Its outline is as follows:

In chapter 1, the problem of polymers near surfaces is motivated. Some established knowledge about the bulk behaviour of self-avoiding self-interacting polymers in solution, in particular the  $\theta$ -transition, is shortly reviewed. And some reasons to use a mesoscopic coarse-grained model rather than an all-atom approach are given.

Chapter 2 introduces the studied model including intrinsic polymer interactions, surface interaction as well as the applied boundary conditions. Also the measured observables are motivated and introduced.

Chapter 3 deals with the Monte Carlo techniques applied in this thesis. After a short overview of the Monte Carlo method in general and its applications in statistical physics in particular, the concept of Markov processes is explained and the different updates used to generate new configurations are explained in detail. The main technique employed here is a generalised ensemble method: The multicanonical method. It is introduced as well as the multicanonical recursion that is used to generate the necessary weights. The concept of reweighting generated conformations in order to estimate canonical expectation values for different external parameters will also be outlined and explicitly described for  $T$ ,  $s$  and  $\epsilon_s$ . Additionally the idea of “Replica Exchange Monte Carlo”, that was used for some additional studies, and of the specially biased algorithm “Energy Landscape Paving” to find global energy minima is given. The chapter is concluded with some words on the applied Jackknife analysis for error estimation.

Chapter 4 finally presents the results obtained. Here, the main focus is on the behaviour of a polymer near an attractive substrate (section 4.1). The canonical expectation values over a wide range of surface attraction and temperature are presented together with the density of states and low energy conformations. All gained informations are summarised in a pseudophase diagram and compared with lattice results. Similarities and differences are discussed. In section 4.2, some additional results on the behaviour with varying solvent quality is presented.

Last but not least, the summary chapter 5 compiles the main facts in a short form.





# 1 Introduction

In this chapter the problem of polymers near or at substrates is motivated and some reasons to use a mesoscopic coarse-grained modelling approach are given.

## 1.1 Motivation

Polymers near surfaces is a fascinating field for both physicists and chemists. It provides a rewarding playground for basic and applied research. With the advent of new sophisticated experimental techniques with its enormous potential in polymer and surface research the interest in the hybrid interface of organic and inorganic matter has increased. Such techniques at the nanometer scale are, e.g., atomic force microscopy (AFM), where it is possible to measure the contour length and the end-to-end distance of individual polymers [1] or to quantitatively investigate the peptide adhesion on semiconductor surfaces [2]. Or optical tweezers that are an experimental tool with an extraordinary resolution in positioning ( $\pm 1\text{nm}$ ) a micron-sized colloid and in the measurement of forces ( $\pm 50\text{fN}$ ) acting on it [3]. It can be used to probe the behaviour of single polymer molecules such as DNA, titin or myosin.

Applications for adsorption phenomena in polymeric solutions can be found in such different fields as lubrication, adhesion and surface protection, steric stabilisation of colloid particles [4] as well as biological processes of membrane-polymer interaction. To understand the latter is important for the reconstruction of cell processes. An understanding of interfaces is also a prerequisite for making micro- or nanostructures because their behaviour is dominated by surface effects rather than gravitation or inertia.

Despite much effort, the problem of a dilute polymer solution of variable solvent quality exposed to an adsorbing substrate is not fully understood. In the case of a long flexible single chain in a good solvent, and no other interactions present than the excluded volume effect, polymers are modelled by self-avoiding walks (SAWs) on a regular lattice. De Gennes showed, that their scaling properties in the limit of an infinite number of steps  $N$  may be derived from a formal  $n \rightarrow 0$  limit of the  $O(n)$  vector model at its critical point [5]. Also the Ising model ( $n = 1$ ), the XY model ( $n = 2$ ) and the Heisenberg model ( $n = 3$ ) belong to this family. These models that do not intrinsically contain any boundary conditions, as is, e.g., introduced by a substrate, have been investigated intensively by various different methods including mean field approaches, perturbation theory, transfer-matrix methods but also by exact enumerations and Monte Carlo simulations. So it is, e.g., well established that the radius of gyration of SAWs scales

with  $N$  as  $\langle R^2 \rangle \sim N^{2\nu}$ , where  $\nu \approx 0.5874(2)$  for  $N \rightarrow \infty$  or that in a solvent the chain exhibits a transition from a compact globule to an expanded state when the temperature is increased. This transition is called *collapse*, *coil-globule* or  *$\theta$ -transition* and is induced by an effective attractive monomer-monomer interaction that is mainly due to van der Waals forces.

An illustrative lattice argument on what happens at this transition can be found in the book of M. Doi [6]. First, for an ideal chain on a lattice that has no intrinsic interaction whatsoever and can be mapped onto a random walk, the radius of gyration has to be derived for comparison. It is defined as

$$\langle R_{\text{gyr}}^2 \rangle = \frac{1}{N} \sum_{n=1}^N \left\langle \left( \vec{R}_n - \vec{R}_{\text{cm}} \right)^2 \right\rangle \quad \text{with} \quad \vec{R}_{\text{cm}} = \frac{1}{N} \sum_{n=1}^N \vec{R}_n, \quad (1.1)$$

which is equivalent to

$$\begin{aligned} \langle R_{\text{gyr}}^2 \rangle &= \frac{1}{2N^2} \sum_{n=1}^N \sum_{m=1}^N \left\langle \left( \vec{R}_n - \vec{R}_m \right)^2 \right\rangle \\ &= \frac{1}{2N^2} \sum_{n=1}^N \sum_{m=1}^N \sum_{i=n+1}^m \sum_{k=n+1}^m \langle \vec{r}_i \cdot \vec{r}_k \rangle \\ &= \frac{1}{2N^2} \sum_{n=1}^N \sum_{m=1}^N \sum_{i=n+1}^m \langle \vec{r}_i^2 \rangle \\ &= \frac{1}{2N^2} \sum_{n=1}^N \sum_{m=1}^N |n-m| b^2 \\ &\stackrel{N \gg 1}{\approx} \frac{b^2}{2N^2} \int_0^N dn \int_0^N dm |n-m| = \frac{1}{6} N b^2. \end{aligned}$$

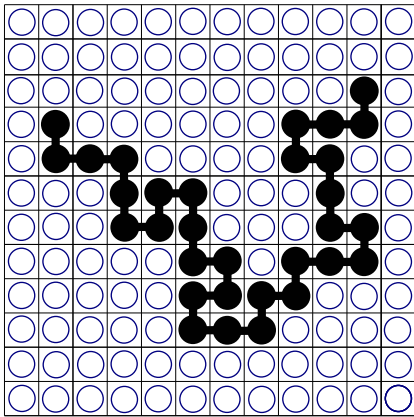


Figure 1.1: Lattice model of the excluded volume chain. The solid circles are the segments of the polymer and the hollow circles are the solvent molecules. Picture adapted from [6].

Here  $\vec{R}_n$  denote positions of monomers and  $\vec{r}_n = \vec{R}_n - \vec{R}_{n-1}$  are bond vectors such that  $\vec{R}_m - \vec{R}_n = \sum_{i=n+1}^m \vec{r}_i$  and it was used that there are no correlations between the directions of different bond vectors:  $\langle \vec{r}_i \cdot \vec{r}_k \rangle = \langle \vec{r}_i^2 \rangle \delta_{ik} = b^2 \delta_{ik}$ , where  $b$  is the bond length. Thus the typical polymer size of an ideal chain scales as  $\langle R_{\text{gyr}} \rangle \simeq N^\nu b$ , with  $\nu = 1/2$ .

For real polymers two effects have to be taken into account: the *excluded volume effect* and the *solvent effect*. The ideal chain model permits widely separated segment along the chain to occupy the same region in space. This physical impossibility is accounted for in an *excluded volume chain* (Fig. 1.1) that corresponds to a *self-avoiding walk* by imposing the condition that two segments

cannot occupy the same lattice site. It is intuitively clear, that this condition shifts the size distribution to higher values since it is mainly the dense conformations that are forbidden now. To quantify this effect, one considers  $W(R)dR$ , the total number of excluded chains with the  $N$ th step a distance between  $R$  and  $R + dR$  away from the origin.  $W(R)$  has to be proportional to the distribution function of  $R$  since all possible paths have the same weight. One estimates it by considering  $W_0(R)dR$  for an ideal chain and multiplies it with the probability that an ideal chain configuration is also allowed under excluded volume condition  $p(R)$  to get  $W(R) = p(R)W_0(R)$ . Since the distribution function for long ideal chains is known to be Gaussian and the overall number of ideal chains with  $N$  steps is  $z^N$ , where  $z$  is the coordination number, it holds

$$W_0(R) = z^N 4\pi R^2 P(\vec{R}, N) = z^N 4\pi R^2 \left( \frac{3}{2\pi N b^2} \right)^{3/2} \exp \left( -\frac{3\vec{R}^2}{2N b^2} \right). \quad (1.2)$$

$p(R)$  can be estimated under the assumption that the polymer segments are evenly distributed in a region of volume  $R^3$ . If the volume of one lattice element is denoted as  $\nu_c$  the number of lattice sites in  $R^3$  is  $R^3/\nu_c$ . Subsequently, if one places  $N$  segments on the sites the probability that one particular segment does not overlap with any other one is given by  $1 - \nu_c/R^3$ . The probability that none of the  $N(N-1)/2$  possible overlaps occurs is now

$$p(R) = (1 - \nu_c/R^3)^{N(N-1)/2} = \exp \left[ \frac{1}{2} N(N-1) \ln \left( 1 - \frac{\nu_c}{R^3} \right) \right] \stackrel{N \gg 1}{\approx} \exp \left( -\frac{N^2 \nu_c}{2R^3} \right), \quad (1.3)$$

which together with eq. (1.2) gives

$$W(R) = p(R)W_0(R) \propto R^2 \exp \left( -\frac{3R^2}{2N b^2} - \frac{N^2 \nu_c}{2R^3} \right). \quad (1.4)$$

Being interested in the behaviour of the radius of gyration and knowing that  $W(R)$  is proportional to the probability that the end-to-end distance of the excluded volume chain is  $R$ , one can use this expression to see how the maxima of  $W(R)$  and  $W_0(R)$  relate to each other.  $W_0(R)$  has its maximum at  $R_0^* = (2N b^2/3)^{1/2}$ . Denoting the maximum of  $W(R)$  by  $R^*$  and differentiating the logarithm of eq. (1.4) one gets  $(R^*/R_0^*)^5 - (R^*/R_0^*)^3 = (9\sqrt{6}\nu_c)/(16b^3)\sqrt{N}$ , which yields for  $N \gg 1$

$$R^* \simeq R_0^* \left( \frac{N^{1/2} \nu_c}{b^3} \right)^{1/5} \propto N^{3/5}. \quad (1.5)$$

This confirms the expectation that an excluded volume chain extends faster with  $N$  than an ideal chain. Extensive simulations found  $\langle R_{\text{gyr}} \rangle \simeq N^\nu b$ , with  $\nu \approx 0.588..$ , which is close to the estimated value.

The presence of a solvent has a considerable influence on the configuration of the polymer that has not been taken into account yet. If the polymer has a high

affinity with the solvent, the polymer is easily dissolved and in such a good solvent polymer conformations are extended random coils. On the other hand, in a bad solvent compact conformations are favoured. This solvent effect is modeled with a parameter  $s$  in this thesis, but is mainly considered to be constant. In the lattice model one can account for this effect by introducing a solvent that consists of single particles occupying all lattice sites that are not occupied by the polymers. Neighboring sites are assigned the energies  $-\epsilon_{pp}$  for two polymer segments,  $-\epsilon_{ss}$  for two solvent molecules and  $-\epsilon_{ps}$  for a solvent molecule-polymer segment interaction. Since these interactions are van der Waals energies,  $\epsilon_{pp}$ ,  $\epsilon_{ss}$  and  $\epsilon_{ps}$  are positive. This leads to an overall system energy for any configuration with  $N_{pp}$  polymer-polymer contacts,  $N_{ss}$  solvent-solvent contacts and  $N_{ps}$  polymer-solvent contacts of

$$E = -N_{pp}\epsilon_{pp} - N_{ss}\epsilon_{ss} - N_{ps}\epsilon_{ps}. \quad (1.6)$$

Since this changes the probability of an excluded volume chain in solution to have size  $R$  to

$$P(R) \propto W(R) \exp\left(\frac{\bar{E}(R)}{k_B T}\right), \quad (1.7)$$

one is interested in the average energy  $\bar{E}(R)$  of a polymer of size  $R$ . Assuming again that all polymer segments are uniformly distributed in a volume  $R^3$  and denoting the probability that a lattice site in this region is occupied by a polymer segment by  $\phi = N\nu_c/R^3$  one can estimate the average number of contacts to be  $\bar{N}_{pp} \simeq zN\phi/2$ ,  $\bar{N}_{ss} \simeq N_{ss}^0 - [zN\phi/2 + zN(1-\phi)]$  and  $\bar{N}_{ps} \simeq zN(1-\phi)$  respectively, where  $N_{ss}^0$  is the number of neighboring solvent pairs if there would be no polymer in the system. Substituting this into eq. (1.6) gives

$$\begin{aligned} \bar{E}(R) &\simeq -\frac{1}{2}zN\phi(\epsilon_{pp} + \epsilon_{ss} - 2\epsilon_{ps}) + \text{terms independent of } \phi \\ &= -\frac{zN^2\nu_c}{R^3}\Delta\epsilon + \text{terms independent of } R, \end{aligned} \quad (1.8)$$

where  $\Delta\epsilon = \frac{1}{2}(\epsilon_{pp} + \epsilon_{ss}) - \epsilon_{ps}$  is the decrease in energy when two polymer segments touch. Hence for  $\Delta\epsilon > 0$  compact conformations are energetically favourable and costly for  $\Delta\epsilon < 0$ . Plugging eq. (1.8) into eq. (1.7) finally yields

$$P(R) \propto R^2 \exp\left(-\frac{3R^2}{2N\nu_c} - \frac{N^2\nu_c}{2R^3}\left(1 - 2\frac{z\Delta\epsilon}{k_B T}\right)\right). \quad (1.9)$$

Using the same trick as before, one finds that eq. (1.5) is valid again if one substitutes  $\nu_c$  by  $\nu = \nu_c(1 - \frac{2z}{k_B T}\Delta\epsilon)$ , such that the scaling is the same.

Comparing eq. (1.9) with eq. (1.2), one can see now that the polymer behaves like an ideal polymer if  $\nu = 0$ . The temperature at which this is the case is called the  $\theta$ -temperature given by

$$\theta = \frac{2z\Delta\epsilon}{k_B}. \quad (1.10)$$

Especially for large  $N$ , close to the  $\theta$ -temperature only a small change in temperature leads to a big change in the size of the polymer, e.g., below the  $\theta$ -temperature, the size is much smaller than that of an ideal chain. In particular, an interacting self-avoiding polymer below the  $\theta$ -temperature scales as  $\langle R_{\text{gyr}} \rangle \sim N^{1/3}$ .

However, the behaviour of the system is strongly affected by the presence of an attractive surface. In its vicinity the monomer-monomer attraction responsible for the collapsed state below the  $\theta$ -transition and the surface-monomer attraction responsible for the adsorption will compete. This competition gives rise to a variety of interesting new conformations. The polymer will adsorb to the surface, if the temperature is lowered, but at high temperatures only a finite number of monomers lie on the surface, even if the polymer is grafted to it. This is due to the lower entropy of conformations spread out on the surface compared to those floating freely in solution.

Numerous detailed studies have been performed to elucidate the conformational behaviour close to and on a substrate for homo- as well as for heteropolymers. Compared to experiments computer simulations have the advantage that combinations of parameters can be varied at wish. Also, in many experiments it is not completely clear, if the system is entirely thermodynamically equilibrated or if, e.g., polymers that were initially in solution get irreversibly adsorbed at the substrate. The structure found can differ considerably [7]. In simulations these conditions can be adjusted at will.

Theoretical studies have, e.g., been performed analytically with scaling theory [8; 9], mean-field density functional theory [10] and series expansion [11; 12] and numerically with off-lattice models such as a bead-spring model of a single polymer chain grafted to a weakly attractive surface [13], multiscale modelling [14], Monte Carlo studies of self-avoiding walks [8; 15; 16; 17; 18; 19; 20], molecular dynamics combined with a stretching of an adsorbed homopolymer [21] or exact enumeration [22]. Also adsorption-desorption dynamics were investigated with molecular dynamics of coarse-grained models [23].

## 1.2 Coarse-graining

The noticeable frequency with which coarse-grained models, that include lattice models, are applied has good reasons. First, one has to understand the complexity of the problem posed by naturally occurring macromolecules with up to ten thousands of atoms. Although the physical interactions present are in principle known, the long-range overlap of many-body orbitals, the screening by the positively charged cores, the interaction with the solvent etc. make a precise prediction of the behaviour of the system based on ab initio quantum-mechanical calculations practically impossible. For the related problem of protein folding, classical models with many effective parameters (“force fields”) have been developed in the past decades to study folding dynamics or to predict native structures in computer simulations. So, e.g. the SMMP (Simple Molecular

Mechanics for Proteins) [24] implementation of the ECEPP/3 (Empirical Conformational Energies for Proteins and Polypeptides) force field has seven different parametrisations of hydrogen, depending on the chemical group it belongs to. In a different approach one starts with a simple ansatz for the interaction potential and calibrates against pertaining data to folding properties of whole chains. This is implemented in the package PROFASI (PROtein Folding and Aggregation SIMulator) [25].

But even those simplified models are still hard to manage even by sophisticated algorithms and powerful capability computers. And if one does, it turned out that the folding behaviour sensitively depends on the choice of the force field parameters, such that predictions of different models do not frequently coincide. Despite the exciting development and the successes in this field and its need if one wants to look closer at detailed structures, for our purposes it is sufficient to work on even coarser grains.

A linear polymer is a chain of molecular subunits called monomers. These can be identical (homopolymer) or vary along the polymer (heteropolymer). Here, we are only interested in general properties of semiflexible homopolymers and their adsorption properties on a substrate. General properties are those that are independent of the detailed chemical structure of the polymer. Hence also the model applied can be reduced to the basic properties of the system and the polymer is regarded as a chain of point-like effective monomers. This parametric reduction is called *coarse-graining* [26; 27]. In particular, in a coarse-grained model, the number of monomers and the degree of polymerisation are not necessarily the same thing. In fact, several chain segments are merged to form one effective monomer (Fig. 1.2). Hence the coarse-grained model has fewer degrees of freedom than those actually present in the system and the relevant length scales are increased. To work with a coarse-grained model has thus two advantages: Unnecessary details are disregarded and the computational analysis is much faster.

Suitable potentials are needed that give rise to the self-avoidance of the polymer, i.e. different parts of the chain should not be allowed to overlap, and reflect the interaction between different monomers. Also a bending-stiffness or torsional potential could be taken into account.

The solvent is implicitly modeled by the interaction between the monomers, which again decreases the degrees of freedom and increases the speed with which the simulation can be done. In real polymer solutions with a good solvent, the monomers and the solvent molecules attract each other. Consequently, solvent molecules accumulate between monomers and push the monomers apart. This can be modeled with a repulsive interaction between the monomers. In a bad solvent, it is just the other way around: repulsive interac-

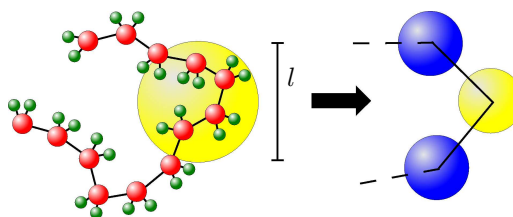


Figure 1.2: Transition from the chemical polymer to the coarse-grained model.

---

tion between monomers and solvent molecules but also entropic forces act like an effective attractive potential between the monomers and the polymer shrinks to form a globular conformation. There are several models that satisfy these conditions. The one applied here is described in more detail in the next chapter.





## 2 The Model

In this chapter, a complete description of the studied model including intrinsic polymer interactions, surface interaction as well as the applied boundary conditions will be given. Also the measured observables will be motivated and introduced.

### 2.1 Definition of the Polymer

The model applied here is a coarse-grained model of a semiflexible homopolymer. It is adapted from the AB model [29; 30], an off-lattice generalisation of the HP model [31], where the monomers are constrained to the grid points of a simple cubic lattice and the polymer is modeled by a self-avoiding walk. But while AB and HP model are heteropolymer models with hydrophobic (A, H) and polar (B, P) monomers, we restrict ourselves to homopolymers, such that only one kind of monomers is taken into account that builds up the whole polymer by mere repetition of always the same unit.

Our model is a coarse-grained model, that is not constrained to a lattice. The associated additional computational cost is accepted in order to get rid of undesired effects of the underlying lattice symmetries of lattice models. As on the lattice, the distance between adjacent monomers is fixed and set to unity in good relation to the monomer-monomer-potential, but the angles are now free to rotate. The energy function has three terms:

$$E_{\text{total}} = E_{\text{bend}} + E_{\text{LJ}} + E_{\text{sur}}, \quad (2.1)$$

the bending energy  $E_{\text{bend}}$ , a Lennard-Jones interaction energy  $E_{\text{LJ}}$  and a surface attraction energy  $E_{\text{sur}}$ . A torsional potential is not considered. A sketch of a

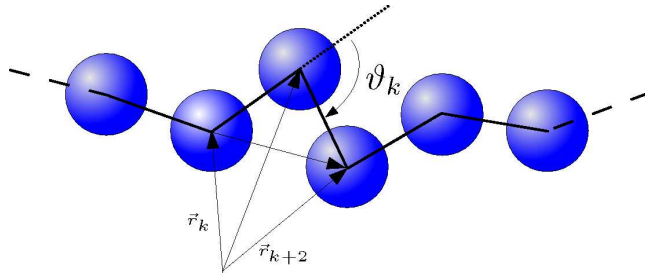


Figure 2.1: A segment of the semiflexible polymer model. The distance between two adjacent monomers is fixed and set to unity. The bonding angle at the  $(k+1)$ th monomer is denoted by  $\vartheta_k$  and the vector between the  $k$ th and  $(k+2)$ th monomer by  $\vec{r}_{k+2,k} \equiv \vec{r}_{k+2} - \vec{r}_k$ .

polymer segment of this model without the surface can be seen in Fig. 2.1. If the position vector of the  $k$ th monomer,  $k = 1, \dots, N$ , is denoted by  $\vec{r}_k$ , the condition of fixed monomer-monomer distance reads as

$$|\vec{r}_{k+1} - \vec{r}_k| = 1 \quad \forall \quad k = 1, \dots, N - 1. \quad (2.2)$$

A polymer with  $N$  monomers has  $N - 1$  bonds between neighboring monomers and  $N - 2$  bending angles  $\vartheta_k$ ,  $k = 1, \dots, N - 2$ , that are defined by

$$\cos(\vartheta_k) = (\vec{r}_{k+1} - \vec{r}_k) \cdot (\vec{r}_{k+2} - \vec{r}_{k+1}). \quad (2.3)$$

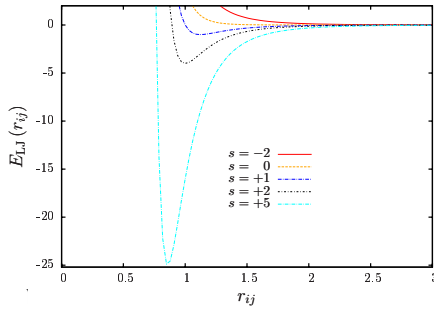


Figure 2.2: Lennard-Jones potential in different solutions. For  $s > 0$  the potential has a minimum at  $r_{\min} = \sqrt[6]{2/s}$  where the attraction between two monomers is  $E_{\text{LJ}}(r_{\min}) = -s^2$ . For  $s \leq 0$  the potential is purely repulsive.

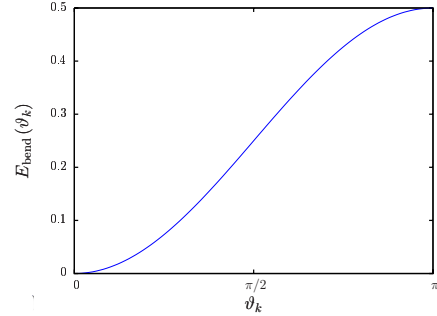


Figure 2.3: Bending energy for one bonding angle. It has its minimum at  $\vartheta_k = 0$  (straight bond) and its maximum at  $\vartheta_k = \pi$  (maximal bending).

With those definitions both intrinsic energy terms can be defined. The Lennard-Jones term that is always repulsive at short ranges is given by

$$E_{\text{LJ}} = 4 \sum_{i=1}^{N-2} \sum_{j=i+2}^N \left( \frac{1}{r_{ij}^{12}} - s \frac{1}{r_{ij}^6} \right) \quad (2.4)$$

with  $s$  being a constant inverse solubility parameter. For  $s \leq 0$  the monomers repel each other which leads to spread out conformations as can be found in good solvents (Fig. 2.2). On the other hand for  $s > 0$  the monomers attract each other which for high enough attraction leads to globular conformations like in a bad solvent. The minimum of the Lennard-Jones potential between two monomers  $r_{\min} = \sqrt[6]{2/s}$  changes with  $s$  and equals the distance of neighboring monomers along the chain at  $s = 2$ . The depth of the potential goes with  $s$  as  $E_{\text{LJ}}(r_{\min}) = -s^2$ . Hence, if one only wants to model attractive monomer-monomer interactions, it might be just as reasonable to fix  $s$  and vary the whole Lennard-Jones interaction  $E_{\text{LJ}}$  linearly. We mainly choose  $s = 1$  to model hydrophobic peptides, but also start to investigate the behaviour for  $s \in [-2, 5]$  for a short chain. The bending stiffness is defined as (Fig. 2.3)

$$E_{\text{bend}} = \frac{1}{4} \sum_{k=1}^{N-2} (1 - \cos(\vartheta_k)). \quad (2.5)$$

The angle  $\vartheta_k$  is in the interval  $[0, \pi)$  and the bending stiffness can be viewed as a penalty introduced for bonds that deviate from the straight conformation.

## 2.2 The Surface Potential

To define a suitable attractive surface potential, we first assume that the surface is made up of a single type of lattice planes, that are arranged in layers to form a crystal. The upper layer is in contact with the polymer and forms a regular lattice where next neighbours of an arbitrary atom with distance  $l_1$  and  $l_2$  lie in the direction of the unit vectors  $\vec{a}_1$  and  $\vec{a}_2$ , respectively. Thus, the surface lattice can be completely described by the lattice vectors  $\vec{a}_1$  and  $\vec{a}_2$ . Now, the interaction of a single monomer with the crystal  $V(\vec{s}, z)$  can be expressed in terms of a two-dimensional vector  $\vec{s}$ , that gives the position of the monomer in an  $xy$ -plane parallel to the surface, and the  $z$ -distance to the wall. Due to the periodic structure of the surface, this potential is a periodic function

$$V(\vec{s}, z) = V(\vec{s} + l_1\vec{a}_1 + l_2\vec{a}_2, z), \quad (2.6)$$

with integer  $l_1$  and  $l_2$ . The natural way of representing a periodic function such as eq. (2.6) is a Fourier series:

$$V(\vec{s}, z) = V_0(z) + \sum_{q \neq 0} V_q(z) e^{i\vec{q} \cdot \vec{s}}. \quad (2.7)$$

Here  $V_0(z)$  is a mean over the whole surface and the sum is over all two-dimensional reciprocal lattice vectors  $\vec{q} = 2\pi(n_1\vec{b}_1 + n_2\vec{b}_2)$  with natural numbers  $n_1, n_2$ .  $\vec{a}_1, \vec{a}_2, \vec{b}_1$  and  $\vec{b}_2$  are defined such that  $\vec{a}_1 \cdot \vec{b}_1 = 1 = \vec{a}_2 \cdot \vec{b}_2$  and  $\vec{a}_1 \cdot \vec{b}_2 = 0 = \vec{a}_2 \cdot \vec{b}_1$  holds. If the sum is neglected and one only works with  $V_0(z)$ , the wall is modeled as completely smooth and formless in lateral direction. This is what we are going to employ.

We assume that the interaction between polymer units and the substrate is of van der Waals type, modeled by the usual Lennard-Jones 12-6 expression

$$V_{LJ}(r) = 4\epsilon_s \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]. \quad (2.8)$$

In order to simplify the problem on mesoscopic scales, we integrate this potential over the plane parallel to the surface, since the potential only depends on the distance  $z$  to the surface. This is best done using cylindrical coordinates for

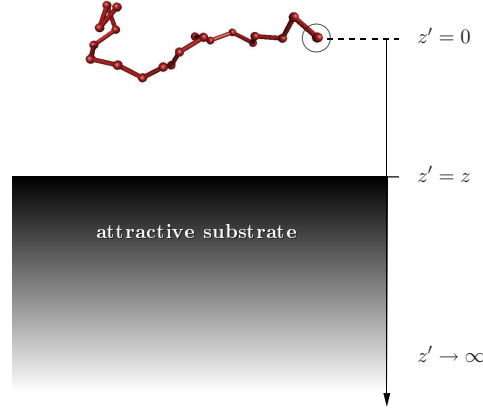


Figure 2.4: Graphical representation of the integration over all surface layers.

which  $d^3\vec{r} = d^2\vec{s}dz = \rho d\rho d\varphi dz$  and  $r = \sqrt{\rho^2 + z^2}$  holds:

$$\begin{aligned}
 \int_{-\infty}^{\infty} V_{\text{LJ}}(r) d^2\vec{s} &= 4\epsilon_s \int_0^{\infty} \rho d\rho \int_0^{2\pi} d\varphi \left[ \left( \frac{\sigma}{\sqrt{\rho^2 + z^2}} \right)^{12} - \left( \frac{\sigma}{\sqrt{\rho^2 + z^2}} \right)^6 \right] \\
 &= 8\pi\epsilon_s\sigma^2 \left[ \underbrace{\sigma^{10} \int_0^{\infty} d\rho \left( \frac{\rho}{\sqrt{\rho^2 + z^2}} \right)^{12}}_{1/(10z^{10})} - \underbrace{\sigma^4 \int_0^{\infty} d\rho \left( \frac{\rho}{\sqrt{\rho^2 + z^2}} \right)^6}_{1/(4z^4)} \right] \\
 &= 2\pi\epsilon_s\sigma^2 \left[ \frac{2}{5} \left( \frac{\sigma}{z} \right)^{10} - \left( \frac{\sigma}{z} \right)^4 \right]. \tag{2.9}
 \end{aligned}$$

Multiplying this with the density  $\rho_{\text{surf}}$  of the atoms in the area gives the interaction energy of a monomer with an area a distance  $z$  away:

$$V_{\text{LJ,plane}}(z) = 2\pi\epsilon_s\rho_{\text{surf}}\sigma^2 \left[ \frac{2}{5} \left( \frac{\sigma}{z} \right)^{10} - \left( \frac{\sigma}{z} \right)^4 \right]. \tag{2.10}$$

To model a compact wall, that consists not only of a single layer of atoms, but of plenty of them reaching from distance  $z$  to  $\infty$  (Fig. 2.4), an additional integration over the  $z$ -component has to be performed:

$$\begin{aligned}
 V_{\text{surf}}(z) &= \int_z^{\infty} V_{\text{LJ,plane}}(z') dz' \\
 &= 2\pi\epsilon_s\rho_{\text{surf}}\sigma^2 \int_z^{\infty} \left[ \frac{2}{5} \left( \frac{\sigma}{z'} \right)^{10} - \left( \frac{\sigma}{z'} \right)^4 \right] dz' \\
 &= 2\pi\epsilon_s\rho_{\text{surf}}\sigma^2 \left[ \frac{2}{45} \sigma \left( \frac{\sigma}{z} \right)^9 - \frac{1}{3} \sigma \left( \frac{\sigma}{z} \right)^3 \right] \\
 &= \underbrace{\frac{2\pi}{3} \rho_{\text{surf}} \sigma^3}_{=1} \epsilon_s \left[ \frac{2}{15} \left( \frac{\sigma}{z} \right)^9 - \left( \frac{\sigma}{z} \right)^3 \right] \\
 &\stackrel{\sigma=1}{=} \epsilon_s \left[ \frac{2}{15} \left( \frac{1}{z} \right)^9 - \left( \frac{1}{z} \right)^3 \right]. \tag{2.11}
 \end{aligned}$$

This finally is the potential we use in our simulations of hybrid systems. It is similar to the one derived in Ref. [32]. The underbraced constant factor as well as  $\sigma$  is set to unity in all the simulations. Different surfaces are modeled by varying  $\epsilon_s$ . The functional dependence is represented in Fig. 2.5.

All energy contributions of our homopolymer near an attractive substrate sum up to give the total energy:

$$E = 4 \sum_{i=1}^{N-2} \sum_{j=i+1}^N \left( \frac{1}{r_{ij}^{12}} - s \frac{1}{r_{ij}^6} \right) + \frac{1}{4} \sum_{k=1}^{N-2} (1 - \cos \vartheta_k) + \epsilon_s \sum_{i=1}^N \left[ \frac{2}{15} \left( \frac{1}{z_i} \right)^9 - \left( \frac{1}{z_i} \right)^3 \right]. \quad (2.12)$$

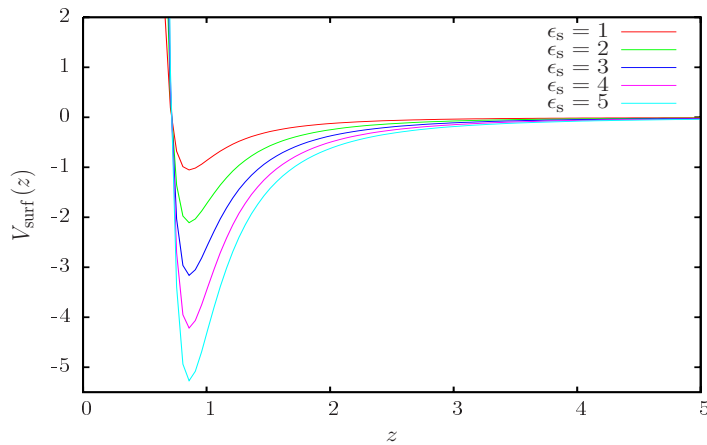


Figure 2.5: The surface potential (2.11) plotted for various values of  $\epsilon_s$ . The minimum is attained at  $z = (2/5)^{1/6} \approx 0.858$ .

## 2.3 Definition of Measured Observables

In order to extract as many information as possible about the canonical equilibrium behaviour out of the simulations, suitable measurable quantities have to be defined. Additionally to these quantities, it is very instructive to also consider their fluctuations.

The canonical expectation value of any quantity  $O$  is given by

$$\langle O \rangle_{can}(T) = \frac{\sum_{\mu \in \mathcal{M}} O_{\mu} e^{-\frac{E_{\mu}}{T}}}{\sum_{\mu \in \mathcal{M}} e^{-\frac{E_{\mu}}{T}}} \quad (2.13)$$

and its fluctuation is obtained from the temperature derivative of the expecta-

tion value:

$$\begin{aligned}
\frac{d\langle \mathbf{O} \rangle_{\text{can}}}{dT} &= \frac{\left( \frac{d}{dT} \sum_{\mu \in \mathcal{M}} O_{\mu} e^{-\frac{E_{\mu}}{T}} \right) \sum_{\mu \in \mathcal{M}} e^{-\frac{E_{\mu}}{T}}}{\left( \sum_{\mu \in \mathcal{M}} e^{-\frac{E_{\mu}}{T}} \right)^2} - \frac{\left( \frac{d}{dT} \sum_{\mu \in \mathcal{M}} e^{-\frac{E_{\mu}}{T}} \right) \sum_{\mu \in \mathcal{M}} O_{\mu} e^{-\frac{E_{\mu}}{T}}}{\left( \sum_{\mu \in \mathcal{M}} e^{-\frac{E_{\mu}}{T}} \right)^2} \\
&= \frac{1}{T^2} \frac{\sum_{\mu \in \mathcal{M}} O_{\mu} E_{\mu} e^{-\frac{E_{\mu}}{T}}}{\sum_{\mu \in \mathcal{M}} e^{-\frac{E_{\mu}}{T}}} \frac{\sum_{\mu \in \mathcal{M}} e^{-\frac{E_{\mu}}{T}}}{\sum_{\mu \in \mathcal{M}} e^{-\frac{E_{\mu}}{T}}} - \frac{1}{T^2} \frac{\sum_{\mu \in \mathcal{M}} E_{\mu} e^{-\frac{E_{\mu}}{T}}}{\sum_{\mu \in \mathcal{M}} e^{-\frac{E_{\mu}}{T}}} \frac{\sum_{\mu \in \mathcal{M}} O_{\mu} e^{-\frac{E_{\mu}}{T}}}{\sum_{\mu \in \mathcal{M}} e^{-\frac{E_{\mu}}{T}}} \\
&= \frac{\langle \mathbf{O} \mathbf{E} \rangle - \langle \mathbf{O} \rangle \langle \mathbf{E} \rangle}{T^2}.
\end{aligned} \tag{2.14}$$

Considered are the following energetical and structural quantities and their fluctuations.

**Energy:** The total energy  $E$  given in eq. (2.12) is measured over the whole parameter regime of interest. The *heat capacity* is defined as fluctuation of the energy

$$C = \frac{d\langle E \rangle}{dT}, \tag{2.15}$$

which is according to eq. (2.14) equivalent with:

$$C = \frac{\langle E^2 \rangle - \langle E \rangle^2}{T^2}. \tag{2.16}$$

**Radius of Gyration:** The radius of gyration is a measure used to describe the extension of a polymer chain. It is defined as the mean distance of a monomer from the centre-of-mass of the polymer:

$$R_{\text{gyr}}^2 \equiv \frac{1}{N} \sum_{n=1}^N \left\langle \left( \vec{R}_n - \vec{R}_{\text{cm}} \right)^2 \right\rangle \quad \text{with} \quad \vec{R}_{\text{cm}} = \frac{1}{N} \sum_{n=1}^N \vec{R}_n. \tag{2.17}$$

This representation is completely equivalent to

$$R_{\text{gyr}}^2 \equiv \frac{1}{2N^2} \sum_{n=1}^N \sum_{m=1}^N \left\langle \left( \vec{R}_n - \vec{R}_m \right)^2 \right\rangle, \tag{2.18}$$

as can be proven quite easily. What makes the radius of gyration an interesting property is that it is related to the pair correlation function

$$g(\vec{r}) \frac{1}{N} \sum_{n=1}^N \sum_{m=1}^N \left\langle \delta \left( \vec{r} - \left( \vec{R}_m - \vec{R}_n \right) \right) \right\rangle \tag{2.19}$$

of polymer segments and can be measured experimentally with static light scattering as well as with small angle neutron- and X-ray scattering which allows to check theoretical predictions against experiments [6; 33; 34].

Since the substrate introduces a structural anisotropy into the system, it is not only worthwhile to look at the overall compactness of the polymer expressed

by  $\langle R_{\text{gyr}} \rangle$ , but also to study the expected different behaviour of its components parallel and perpendicular to the surface:

$$R_{\text{gyr},\parallel}^2 = \frac{1}{2N^2} \sum_{n=1}^N \sum_{m=1}^N \langle (x_n - x_m)^2 + (y_n - y_m)^2 \rangle \quad (2.20)$$

and

$$R_{\text{gyr},\perp}^2 = \frac{1}{2N^2} \sum_{n=1}^N \sum_{m=1}^N \langle (z_n - z_m)^2 \rangle. \quad (2.21)$$

Here  $\vec{R}_n = (x_n, y_n, z_n)$  and  $\vec{R}_m = (x_m, y_m, z_m)$  such that  $R_{\text{gyr}}^2 = R_{\text{gyr},\parallel}^2 + R_{\text{gyr},\perp}^2$ . What we determined is  $\langle R_{\text{gyr}} \rangle$ ,  $\langle R_{\text{gyr},\parallel} \rangle$  and  $\langle R_{\text{gyr},\perp} \rangle$ , for which  $\langle R_{\text{gyr}} \rangle^2 \neq \langle R_{\text{gyr},\parallel} \rangle^2 + \langle R_{\text{gyr},\perp} \rangle^2$ . Additionally, the thermal fluctuations  $d\langle R_{\text{gyr}} \rangle/dT$ ,  $d\langle R_{\text{gyr},\parallel} \rangle/dT$  and  $d\langle R_{\text{gyr},\perp} \rangle/dT$  were measured using eq. (2.14).

We also estimated the end-to-end distance  $R_{ee} = R_N - R_1$  but gained no additional information from it.

**Distance of the Centre-of-Mass of the Polymer to the Surface:** This quantity is useful since it provides clear evidence if the polymer is on average freely moving in the box or very close to the surface. Apart from that it is very easily implemented, since one only needs to average over the  $z$ -components that one needs in the simulation anyway:

$$z_{\text{cm}} = \frac{1}{N} \sum_{i=1}^N z_i. \quad (2.22)$$

Again,  $\langle z_{\text{cm}} \rangle$  and its thermal fluctuations  $d\langle R_{\text{cm}} \rangle/dT$  were measured both.

**Mean Number of Surface Contacts:** Not only the distance to the surface, but in particular also the number of monomers docked to the surface, gives a useful contribution to identify pseudo-phases. So all monomers attached corresponds to a single layer structure, none attached to a free polymer and all the exciting things happen in between. The surface potential is a continuous potential and in order to differentiate monomers docked to the substrate from the ones not docked, it is necessary to introduce a cutoff. After regarding eq. (2.11) we decided somewhat arbitrarily but reasonably to define a monomer with  $z < z_c \equiv 1.2$  as a docked monomer. The corresponding measured quantity is the average ratio  $\langle n_s \rangle$  of monomers docked to the surface to the total number of monomers. This can be expressed as:

$$n_s = \frac{N_s}{N} \quad \text{with} \quad N_s = \sum_{i=1}^N \Theta(z_c - z_i), \quad (2.23)$$

where  $\Theta(z)$  is the Heaviside step function. Again also its thermal fluctuation  $d\langle n_s \rangle/dT$  is measured. Note that  $\langle n_s \rangle$  also reflects the energy contribution from the surface attraction  $E_{\text{sur}}$ .

**Mean Number of Intrinsic Contacts:** The mean number of intrinsic contacts is a measure of the overall compactness of the polymer just as the radius of gyration is and reflects the contribution of the intrinsic Lennard-Jones energy. Again, there is no obvious way how to define a monomer-monomer contact and we decided to introduce a cutoff as follows:

$$n_m = \frac{N_m}{N} \quad \text{with} \quad N_m = \sum_{i=1}^{N-2} \sum_{j=i+2}^N \Theta(E_c - E_{\text{LJ}}(r_{ij})). \quad (2.24)$$

Here  $E_{\text{LJ}}(r_{ij}) = 4 \left( r_{ij}^{-12} - s r_{ij}^{-6} \right)$  as in Fig. 2.2 and  $E_c \equiv -0.2$ .  $\Theta(E)$  is again the Heaviside step function. This definition works for varying  $s$  although for small  $s$   $n_s$  is always zero.  $\langle n_m \rangle$  and  $d \langle n_m \rangle / dT$  are looked at.

## 2.4 Boundary Conditions

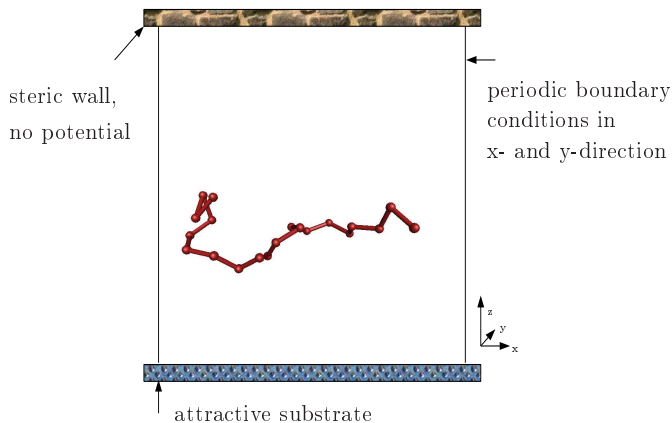


Figure 2.6: Schematic representation of the boundary conditions applied.

In the simulations, two different kinds of boundary conditions were used. Parallel to the wall – in  $z$ -direction – the boundary conditions are purely steric, i.e., an update is simply rejected if it suggests to cross the wall at  $z = 0$  or  $z = L_{\text{box}}$ . At  $z = 0$  the attractive surface potential is applied. This steric wall is necessary to prevent the peptide from escaping to large  $z$  away from the wall with the short range interaction that we want to investigate. Since the exact form of the density of states depends on the box height  $L_{\text{box}}$ , also all observables depend on the choice of  $L_{\text{box}}$ . As soon as the box size however exceeds the polymer size, the influence on the observables is reasonably small. For smaller boxes, interesting deviations in the thermodynamic behaviour can be observed [35].

Perpendicular to the wall in  $xy$ -direction, periodic boundary conditions with the minimum image convention are applied. This is strictly speaking not necessary and one might just as well apply no boundary conditions at all. They are applied here to allow for a possible addition of a second or more polymers into



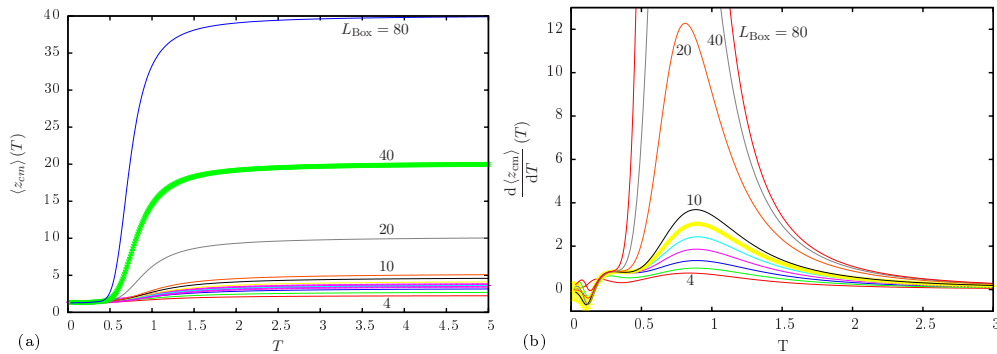


Figure 2.7: (a)  $\langle z_{\text{cm}} \rangle$  of the 13mer, (b)  $d\langle z_{\text{cm}} \rangle/dT$  of the 13mer in cubic boxes of different sizes. The errors are indicated for  $L_{\text{box}} = 40$  and 9, resp.

the systems that would rarely ever meet each other if there were no boundaries. On the other hand, the periodic boundary conditions for not too small boxes hardly have any influence on the results, since the monomer-monomer potential  $E_{\text{LJ}}$  is short-ranged.

To get an impression of how much the box really influences our results, we performed some test runs for cubic box sizes  $L_{\text{box}} = 4, 5, 6, 7, 8, 9, 10, 20, 40, 80$  and  $s = 1$  and  $\epsilon_s = 1$ . In Fig. 2.7, the average centre-of-mass distance of the polymer to the surface  $\langle z_{\text{cm}} \rangle$  and its thermal fluctuation are shown for various box sizes  $L_{\text{box}}$ .

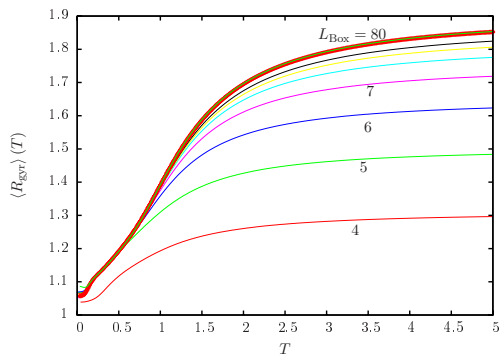


Figure 2.8:  $\langle R_{\text{gyr}} \rangle$  of the 13mer in cubic boxes of different sizes. The errors are indicated for  $L_{\text{box}} = 40$ .

This quantity clearly is the one that depends the most on the box size. For high temperatures, the polymer can move freely within the box such that  $\langle z_{\text{cm}} \rangle \approx L_{\text{box}}/2$ . For low  $T$  however the polymer prefers to stay close to the surface. Accordingly, also the fluctuation of  $\langle z_{\text{cm}} \rangle$  depends on the box size and as indicated in Figure 2.7(b) the peak height depends very strongly on it. Also the peak position shifts to lower temperatures if the box size increases. One has to keep this dependence in mind when deciding to concentrate on a fixed box size as we do.

But as soon as the box is clearly larger than the average size of the polymer, the influence of the box on the conformation of the polymer is not so striking any more. This is shown in Figure 2.8, where the radius of gyration is displayed for varying  $L_{\text{box}}$ . As soon as  $L_{\text{box}} \gtrsim N$ , the change with  $L_{\text{box}}$  gets smaller than the statistical error. We chose  $L_{\text{box}} = 20$  for  $N = 13$  and  $L_{\text{box}} = 40$  for  $N = 20$ . Although this reduction of the phase space will always effect the entropy, this choice allows us to still draw valuable conclusions out of the simulation.



## 3 Monte Carlo Simulations

The Monte Carlo method is a numerical method that computes its results by repeated random sampling of states. This is done by drawing pseudo random numbers that are generated by special algorithms, the so called random number generators [36; 37]. We only give a very short sketch of the idea here and refer for more informations to the standard literature [38; 39; 40; 41].

The term Monte Carlo was coined in the 1940s by physicists working on nuclear weapon projects in the Los Alamos National Laboratory and was clearly supposed to reflect some of the glamour of the eponymous city that is famous for its gambling – another area where random numbers play a role. The idea to use random numbers is, however, older and was, e.g., applied to estimate integrals or  $\pi$ .

For Monte Carlo integration one uses that the expectation value of the mean over a finite number of random variables  $f(x_i)$  with  $x_i$  drawn according to the probability density  $\rho(x)$  is identical with the expectation value of  $f(x)$  over the underlying distribution:

$$\langle \bar{f} \rangle = \langle f(x) \rangle, \quad \text{where } \bar{f} = \frac{1}{n} \sum_i^n f(x_i), \quad x_i \propto \rho(x). \quad (3.1)$$

Knowing that the original definition of the expectation value is an integral over the distribution

$$\langle f(x) \rangle = \int \rho(x) f(x) dx = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_i^n f(x_i), \quad x_i \propto \rho(x), \quad (3.2)$$

one can see that the mean  $\bar{f}$  in eq. (3.1), that is a random number fluctuating around the theoretically expected value, approximates the integral in eq. (3.2) with an error that diminishes proportional to  $1/\sqrt{n}$ . All true random sampling techniques behave like this as a universal consequence of the central limit theorem. What can be influenced by smart sampling techniques is the prefactor of  $n^{-1/2}$  that depends on the variance of the function being sampled. A very helpful observation is the fact that the  $n^{-1/2}$ -behaviour is independent of the dimension. Thus, especially for high-dimensional integrals, Monte Carlo simulations are *the* method of choice. Other numerical methods like quadrature rules based on interpolating functions [42] sometimes converge faster at low dimensions, but even for low-dimensional cases where Monte Carlo is not the most efficient method it may be an interesting way to produce a crude estimate due to its simplicity. Thanks to its great generality the Monte Carlo technique can be

applied to various problems and especially for ones with a high phase space dimensionality like in our case, there is no better method known.

In *statistical physics*, the canonical expectation value of an observable  $O$  is given by eq. (2.13). Exact enumeration – or integration for continuous systems – is impossible for realistic models due to the overwhelming large number of possible configurations that scales exponentially with system size. In order to nevertheless get an estimates of canonical expectation values, Monte Carlo simulations are needed. Here one takes a random subset  $\mathcal{R}$  from some probability distribution  $p_\mu$  that is specified cleverly beforehand instead of the whole configuration space and estimate the expectation value as follows:

$$\langle O(E) \rangle (T) = \frac{\sum_{\mu \in \mathcal{R}} O(E_\mu) e^{-E_\mu/T} / p_\mu}{\sum_{\mu \in \mathcal{R}} e^{-E_\mu/T} / p_\mu}. \quad (3.3)$$

This estimate is rather poor if all  $p_\mu$  are equal (simple sampling), because it is only possible to sample a very small fraction of the total number of states. On the other hand, if one picks those states that make an important contribution to eq. (2.13) and ignores others, which is called *important sampling*, one can do rather well. How we choose such important states for our model is explained in the next section.

### 3.1 Generation of Relevant Configurations

If one would randomly generate new configurations that satisfy the given constraints of fixed bond length, the danger of generating many thermodynamically irrelevant high-energy conformations with two or more monomers in unphysical close proximity is high. One would effectively suggest all configurations with equal probability and perform a simple sampling with the mentioned poor performance. Hence it became common practice to generate new conformations using a *Markov process*, i.e. given a system in state  $\mu$  a new state of that system  $\nu$  is generated that still resembles state  $\mu$  and so forth. The probability for the transition from  $\mu$  to  $\nu$  is called *transition probability*  $P(\mu \rightarrow \nu)$  and for a true Markov process  $P(\mu \rightarrow \nu)$  does not vary over time and should depend only on the properties of the states  $\mu$  and  $\nu$  and nothing else. Also the constraint  $\sum_\nu P(\mu \rightarrow \nu) = 1$  has to hold since the Markov process must generate some state  $\nu$  when handed a system  $\mu$ , including  $\mu$  itself.

Different updates generate some random new state  $\nu$  if given state  $\mu$  that still is very similar to the old one. Unlike e.g. for the Ising model, where it is possible to update (flip) a single spin which only affects its nearest neighbours, no local updates exist for our continuous model, i.e. every update affects more than just its nearest neighbours. This is due to the long-range interactions of the monomers such that monomers that are far away along the chain, are spatially in close proximity and interact. That is why individual updates are more time consuming for our model and the acceptance ratio (eq. (3.7)) decreases for dense conformations.

In order for the Markov chain to produce a subset of states  $\nu$  distributed with probability  $p_\nu$ , the following two important conditions have to be met:

1. *Ergodicity*: With the applied sequence of updates it should be possible to reach any configuration in the configuration space from any other during the simulation. If the simulation contains more than  $N_s$  updates, this reads as

$$P(\mu \rightarrow \nu) = P(\mu \rightarrow \lambda_1) \left[ \prod_{i=1}^{N_s-1} P(\lambda_i \rightarrow \lambda_{i+1}) \right] P(\lambda_{N_s} \rightarrow \nu). \quad (3.4)$$

2. *Detailed balance*: If a system is in equilibrium, the rate at which the system enters and leaves any state  $\mu$  must be equal:

$$\sum_{\nu} p_{\mu} P(\mu \rightarrow \nu) = \sum_{\nu} p_{\nu} P(\nu \rightarrow \mu), \quad (3.5)$$

which is equivalent (using  $\sum_{\nu} P(\mu \rightarrow \nu) = 1$ ) to  $p_{\mu} = \sum_{\nu} p_{\nu} P(\nu \rightarrow \mu)$ . But this condition alone does not guarantee that the Markov chain will have the desired probability distribution  $p_{\nu}$ , due to the possibility of so-called limit circles [41]. Thus one requests the stronger condition of *detailed balance*

$$p_{\mu} P(\mu \rightarrow \nu) = p_{\nu} P(\nu \rightarrow \mu) \quad (3.6)$$

that ensures a generation of states with  $p_{\nu}$ .

We now need to implement a Markov chain with transition probabilities that satisfy the conditions given above. To do so, a trick is used: The transition probability  $P(\mu \rightarrow \nu)$  is split up into two parts:

$$P(\mu \rightarrow \nu) = g(\mu \rightarrow \nu) A(\mu \rightarrow \nu). \quad (3.7)$$

The quantity  $g(\mu \rightarrow \nu)$  is the *selection probability* with which, given a state  $\mu$ , the update generates a new state  $\nu$ , and  $A(\mu \rightarrow \nu)$  is the *acceptance probability* which gives the fraction of times that the generated state is adopted. This leaves complete freedom to choose an update to generate new states, since the rest can be taken care of with the right acceptance probabilities.

We chose a combination of updates that is ergodic and for which  $g(\mu \rightarrow \nu) = g(\nu \rightarrow \mu)$ . Then the condition of detailed balance (eq. (3.6)) reduces to

$$\frac{p_{\nu}}{p_{\mu}} = \frac{P(\mu \rightarrow \nu)}{P(\nu \rightarrow \mu)} = \frac{g(\mu \rightarrow \nu) A(\mu \rightarrow \nu)}{g(\nu \rightarrow \mu) A(\nu \rightarrow \mu)} = \frac{A(\mu \rightarrow \nu)}{A(\nu \rightarrow \mu)}. \quad (3.8)$$

In the following, the updates used in the simulations are described in detail.

### 3.1.1 Spherical update – Forwards (F) and Backwards (B)

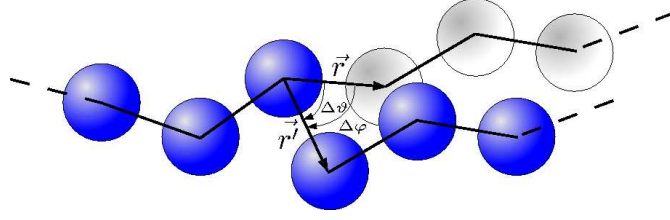


Figure 3.1: Graphical representation of the spherical update. One monomer is moved on the surface of a spherical sector around its preceding monomer. The change in the bonding angle is called  $\Delta\vartheta$  and the rotation angle is named  $\Delta\varphi$ . All following monomers are moved by the same difference

One possible update is to pick one bond at random, rotate it and attach the following bonds without rotation (Fig. 3.1). If the chosen bond is  $\vec{r} = \vec{r}_{k+1} - \vec{r}_k$ , the rotated  $(k+1)$ th monomer moves on a spherical surface around the  $k$ th one, since the bond length is fixed. This explains the name 'spherical update' [43].

There are several ways and means how to implement this update. The implementation used here is adopted from Ref. [44]. We replace the vector  $\vec{r} = \vec{r}_{k+1} - \vec{r}_k$  by its rotated version  $\vec{r}'$ :

$$\vec{r}' = \cos \Delta\vartheta \vec{e}_r + \sin \Delta\vartheta \sin \Delta\varphi \vec{e}_\varphi + \sin \Delta\vartheta \cos \Delta\varphi \vec{e}_\vartheta. \quad (3.9)$$

Since  $\vec{r} = |\vec{r}| \vec{e}_r = \vec{e}_r$  and  $\vec{r} \perp \vec{e}_\varphi \perp \vec{e}_\vartheta$  this is the rotation of  $\vec{r}' = \cos \Delta\vartheta \vec{e}_r + \sin \Delta\vartheta \vec{e}_\vartheta$ , some unit vector with  $\angle(\vec{r}, \vec{r}') = \Delta\vartheta$ , an arbitrary angle  $\Delta\varphi$  about  $\vec{r}$ . For reasons of efficiency, we would like to have a restriction on  $\Delta\vartheta$  that we choose to be  $\Delta\vartheta_{\max} = 5^\circ$ . Furthermore, to distribute the  $\vec{r}'$  evenly on the surface of the spherical cap ( $dA = \cos \vartheta d\vartheta d\varphi$ )  $\cos \Delta\vartheta$  has to be chosen evenly from the interval  $(\cos \Delta\vartheta_{\max}, 1]$  and  $\Delta\varphi$  from the interval  $[0, 2\pi)$ . Although the first suggested implementation proves that this equidistribution is not necessary to guarantee  $g(\mu \rightarrow \nu) = g(\nu \rightarrow \mu)$ , it is still a way to make sure, that it holds.

In practice, it is easiest to only use cartesian coordinates since they are already implemented. Our choices were

$$\vec{e}_r = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \vec{e}_\varphi = \frac{1}{\sqrt{x^2 + y^2}} \begin{pmatrix} -y \\ x \\ 0 \end{pmatrix}, \quad \vec{e}_\vartheta = \frac{1}{\sqrt{x^2 + y^2}} \begin{pmatrix} -xz \\ -yz \\ x^2 + y^2 \end{pmatrix} \quad (3.10)$$

with  $\vec{e}_\vartheta = \frac{\vec{e}_r \times \vec{e}_\varphi}{|\vec{e}_r \times \vec{e}_\varphi|}$  if  $x^2 + y^2 > 0.1$  and

$$\vec{e}_r = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \vec{e}_\varphi = \frac{1}{\sqrt{x^2 + z^2}} \begin{pmatrix} z \\ 0 \\ -x \end{pmatrix}, \quad \vec{e}_\vartheta = \frac{1}{\sqrt{x^2 + z^2}} \begin{pmatrix} -xy \\ x^2 + z^2 \\ -yz \end{pmatrix} \quad (3.11)$$

otherwise. The distinction is made to avoid problems for  $x \approx y \approx 0$ , where  $\vec{e}_\varphi$  and  $\vec{e}_\vartheta$  are not well defined for the first choice.

This update reaches all configurations with the first monomer fixed at its original position. If one uses this update not only in one direction, but also in backwards direction (the  $k$ th monomer rotates around the  $(k+1)$ th) all configurations can be obtained. These are two independent updates.

Especially in globular configurations, spherical updates might be energetically disadvantageous, since often large parts of the polymer are moved against each other. Here a semi-local update, that only moves one monomer at a time could help.

### 3.1.2 Semi-Local Update (L)

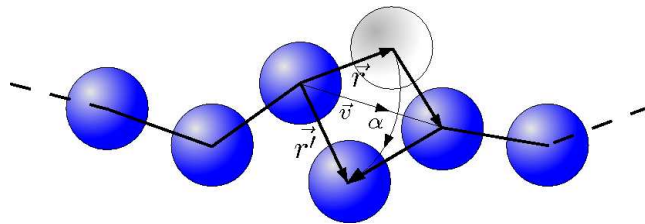


Figure 3.2: Graphical representation of the semi-local update. Here only one monomer is rotated by a random angle  $\alpha$  around the axis defined by the two neighboring monomers.

The semi-local update is inspired by the corner-flip update on a 2D lattice. But due to the continuous space, the rotation angle is not restricted to one angle ( $180^\circ$ ) anymore. One rather rotates the  $k$ th monomer, that is chosen randomly with  $1 < k < N$ , about the connection vector  $\vec{v}$  of its neighboring monomers

$$\vec{v} = \frac{\vec{r}_{k+1} - \vec{r}_{k-1}}{|\vec{r}_{k+1} - \vec{r}_{k-1}|} \quad (3.12)$$

by a random angle  $\alpha \in [0, 2\pi)$  (Fig. 3.2). This is done by applying the rotation matrix  $R(\vec{v}, \alpha)$  (eq. (3.14)) to the connection vector  $\vec{r} = \vec{r}_k - \vec{r}_{k-1}$ , such that the final position of the  $k$ th monomer is given by

$$\begin{aligned} \vec{r}_k^j &= R(\vec{v}, \alpha) \vec{r} + \vec{r}_{k-1} \\ &= \vec{r}^j + \vec{r}_{k-1}. \end{aligned} \quad (3.13)$$

This update alone is not ergodic, because the monomers at the edge are left unchanged. But at least  $g(\mu \rightarrow \nu) = g(\nu \rightarrow \mu)$  holds.

### 3.1.3 Rigid-Body Rotation (R)

Since the surface introduces a spatial anisotropy, the energy is not invariant under rotation. So as an additional update, a rotation of the whole polymer about its centre of mass is introduced. The centre of mass was defined within a second coordinate system without periodic boundary conditions to circumvent

the problems in defining a centre of mass in a periodic system. Again the rotation matrix  $R(\vec{v}, \alpha)$  is applied, that rotates the molecule about the axis of rotation  $\vec{v}$  by an angle  $\alpha$ :

$$R(\vec{v}, \alpha) = \begin{pmatrix} \cos \alpha + v_1^2 (1 - \cos \alpha) & v_1 v_2 (1 - \cos \alpha) - v_3 \sin \alpha & v_1 v_3 (1 - \cos \alpha) + v_2 \sin \alpha \\ v_2 v_1 (1 - \cos \alpha) + v_3 \sin \alpha & \cos \alpha + v_2^2 (1 - \cos \alpha) & v_2 v_3 (1 - \cos \alpha) - v_1 \sin \alpha \\ v_3 v_1 (1 - \cos \alpha) - v_2 \sin \alpha & v_3 v_2 (1 - \cos \alpha) + v_1 \sin \alpha & \cos \alpha + v_3^2 (1 - \cos \alpha) \end{pmatrix}. \quad (3.14)$$

We did not restrict this rotation to small angles, such that a good choice for  $\vec{v}$  is

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} \sin \vartheta \cos \varphi \\ \sin \vartheta \sin \varphi \\ \cos \vartheta \end{pmatrix} \quad (3.15)$$

with  $\cos \vartheta \in (1, -1]$ ,  $\varphi \in [0, 2\pi)$  and  $\alpha \in [0, 2\pi)$ .

After the rotation, the position  $\vec{r}_k$  of the  $k$ th monomer is replaced by

$$\vec{r}'_k = R(\vec{v}, \alpha) (\vec{r}_k - \vec{r}_{\text{cm}}) + \vec{r}_{\text{cm}}. \quad (3.16)$$

Just as the translation, this update does not influence the intrapolymeric interactions but only the interaction with the surface. It keeps the shape of the polymer fixed.

### 3.1.4 Translation (T)

The combination of updates introduced so far hardly changes the distance to the attractive surface. This causes the polymer to need many sweeps to finally feel the influence of the surface if the simulation is initialised with a random conformation in the middle of the box. On the other hand, once caught at the surface, it will take a long time to desorb.

Introducing a translation of the whole molecule eliminates the problem. We first implemented a translation by a length  $l$  chosen at random in the interval  $l \in [0, 1)$  in arbitrary direction but decided soon that with only a single polymer in the system and translational invariance in  $xy$ -direction, it is sufficient to only translate in  $z$ -direction.

### 3.1.5 Sweep Sequence

One sweep consists of an ergodic mixture of all of those updates and contains as many updates as the polymer contains monomers.

With the one-letter codes F, B, L, R and T given in the preceding subsection headings, a typical chain of updates is:

TFBRFBLFB TFBRFBLFB T...



## 3.2 Multicanonical Sampling

The technique we mainly used is multicanonical sampling [40; 45; 46; 47]. Its basic idea is to sample states with a flat energy histogram and to reweight the data to get the canonical expectation values for the relevant temperature range (see section 3.3). This approach has several advantages. Unlike with multihistogram techniques [41; 48], there is no need to create several histograms in order to accumulate enough statistics for each energy bin, but all the necessary statistics is generated and reweighted in a single simulation. So in a way multiple canonical simulations are substituted by a single long run which explains the name “multicanonical”.

Even more important is the ability of this technique to sample configurations with low probability. A canonical simulation samples states with the *Boltzmann distribution* of energies

$$p(E, \beta) \propto \Omega(E) e^{-\beta E} = e^{-\beta E + \ln \Omega(E)} = e^{-\beta F}. \quad (3.17)$$

Hence, states with rare realizations (low  $\Omega(E)$ ) or high energies (low  $e^{-\beta E}$ ) are suppressed. Because of the “rough” free energy landscape, the simulation is likely to get trapped in local free energy minima with an exponentially growing autocorrelation time  $\tau \propto e^{\beta \Delta F}$ , where  $\Delta F$  is the free energy barrier to overcome between two local free energy minima. Thus multicanonical simulations that ideally perform a random walk in energy space significantly reduce the autocorrelation time and the probability to not sample important states of the phase space that would not be reached by canonical simulations due to energy barriers, allowing to gain much more accurate results in a given CPU time.

In order to sample states with a flat histogram one needs multicanonical weights  $W_{\text{muca}}(E)$  satisfying

$$p_{\text{muca}}(E) = \Omega(E) W_{\text{muca}}(E) \approx \text{const.} \quad (3.18)$$

The reweighting back is then done via

$$\langle O \rangle(\beta) = \frac{\sum_i O(E_i) (W_{\text{muca}}(E_i))^{-1} e^{-\beta E_i}}{\sum_i (W_{\text{muca}}(E_i))^{-1} e^{-\beta E_i}}, \quad (3.19)$$

which in principle works for all kinds of weights and should give the same canonical expectations values, but the performance depends crucially on the choice of  $W(E)$ . So, e.g., for  $W(E) = e^{-\beta E}$ , the canonical ensemble is recovered.

Before the actual multicanonical simulation can be performed, the weights  $W_{\text{muca}}(E)$  have to be constructed. We use the *multicanonical recursion* [45; 47] to do so.

Altogether the method consists of the following three steps:

1. Determining the weights  $W_{\text{muca}}(E)$
2. Simulation run with fixed weights and high statistics
3. Reweighting to obtain the canonical expectation values

### 3.2.1 Multicanonical Recursion

The idea of the recursion is that the weights should be inversely proportional to the density of states, that is a priori unknown [45; 47]:

$$W_{\text{muca}}(E) \propto \frac{1}{\Omega(E)} = e^{-\ln \Omega(E)} \equiv e^{-S(E)}. \quad (3.20)$$

Using the dimensionless, microcanonical free energy

$$f(E) = \frac{F(E)}{T(E)} = \frac{U(E) - TS(E)}{T(E)} = \beta(E)E - S(E) \quad (3.21)$$

this can be rewritten as

$$W_{\text{muca}}(E) \propto e^{-\beta(E)E + f(E)}. \quad (3.22)$$

But  $f(E)$  and  $\beta(E)$  are not independent. A relation between them can be derived considering

$$\beta(E) = \frac{1}{T(E)} \stackrel{!}{=} \frac{\partial S}{\partial E} = \beta(E) + E \frac{\partial \beta(E)}{\partial E} - \frac{\partial f(E)}{\partial E} \approx \frac{S(E + \epsilon) - S(E)}{\epsilon} \quad (3.23)$$

according to the first law of thermodynamics  $dU = TdS - pdV$  and eq. (3.21). This requires

$$E \frac{\partial \beta(E)}{\partial E} - \frac{\partial f(E)}{\partial E} = 0 \quad (3.24)$$

to hold true, which is ensured by the relation

$$f(E) - f(E - \epsilon) = (\beta(E) - \beta(E - \epsilon))E, \quad (3.25)$$

where  $\epsilon$  is the smallest energy difference. This is a simplification, but since the recursion is implemented on a computer, one has to discretise anyway.

Using some initial values  $W_{\text{muca},0}(E)$  that corresponds to some  $f_0(E)$  and  $\beta_0(E)$ , the initial run can be performed to gain the histogram  $H_0(E)$ . We used  $W_{\text{muca},0}(E) = 0, \forall E$ , that gives the same weight to all energies<sup>1</sup>. The histogram in turn is used to determine  $f_1(E)$ ,  $\beta_1(E)$  and hence  $W_{\text{muca},1}(E)$ , a better estimate for the multicanonical weights. This is done recursively such that  $H_n(E)$  is used to find  $W_{\text{muca},n+1}(E)$  until the histogram eventually gets flat enough.

To get the most out of the simulations done so far, it makes sense to perform an error weighted average

$$\beta_{n+1}(E) = \kappa(E) \tilde{\beta}_n(E) + (1 - \kappa(E)) \beta_n(E). \quad (3.26)$$

The new estimate  $\tilde{\beta}_n(E)$  from the last simulation is determined considering

$$\tilde{W}_{\text{muca},n}(E) \propto \frac{1}{\Omega(E)} = e^{-\tilde{S}_n(E)} \propto \frac{W_{\text{muca},n}(E)}{H_n(E)} \quad (3.27)$$

<sup>1</sup>Note that this corresponds to  $f_0(E) = 0$  and  $\beta_0(E) = 0$  which is the same as canonical sampling at infinite temperature.

and eq. (3.23):

$$\begin{aligned}
\tilde{\beta}_n(E) &= \frac{\tilde{S}_n(E + \epsilon) - \tilde{S}_n(E)}{\epsilon} \\
&= \frac{\ln \tilde{W}_{\text{muca},n}(E) - \ln \tilde{W}_{\text{muca},n}(E + \epsilon)}{\epsilon} \\
&= \frac{\ln(H_n(E + \epsilon)) - \ln(H_n(E)) - (\ln W_{\text{muca},n}(E + \epsilon) - \ln W_{\text{muca},n}(E))}{\epsilon} \\
&= \frac{\ln(H_n(E + \epsilon)) - \ln(H_n(E))}{\epsilon} + \beta_n(E). \tag{3.28}
\end{aligned}$$

Taking the logarithm of empty histogram bins seems to be a problem here, but one can get around this considering the weights  $\kappa(E)$  in eq. (3.26) that disappear for empty histogram entries.

$\kappa(E)$  has to be inversely proportional to the variance of  $\tilde{\beta}_n(E)$ . According to eq. (3.28) this is

$$\sigma^2(\tilde{\beta}_n(E)) = \sigma^2(\beta_n(E)) + \frac{\sigma^2(\ln H_n(E + \epsilon))}{\epsilon^2} + \frac{\sigma^2(\ln H_n(E))}{\epsilon^2}. \tag{3.29}$$

$\sigma^2(\beta_n(E))$  vanishes since  $\beta_n(E)$  is kept fixed in each simulation.

For the remaining terms it can be used that

$$\begin{aligned}
\sigma^2(\ln H_n(E)) &= [\ln(H_n(E) + \Delta H_n(E)) - \ln(H_n(E))]^2 \\
&= \left[ \ln H_n(E) + \frac{\Delta H_n(E)}{H_n(E)} - \ln H_n(E) \right]^2 \\
&= \left[ \frac{\Delta H_n(E)}{H_n(E)} \right]^2 \propto \frac{1}{H_n(E)}, \tag{3.30}
\end{aligned}$$

where  $\Delta H_n(E)$  is the fluctuation in the  $n$ th histogram which is known to grow with the square root of the number of entries,  $\Delta H_n(E) \propto \sqrt{H_n(E)} \ll H_n(E)$ , which allows a Taylor-expansion.

This yields

$$\sigma^2(\tilde{\beta}_n(E)) \propto \frac{1}{H_n(E + \epsilon)} + \frac{1}{H_n(E)}. \tag{3.31}$$

Now,  $\kappa(E)$  is found by introducing

$$p(E) = \frac{H_n(E + \epsilon) H_n(E)}{H_n(E + \epsilon) + H_n(E)} \propto \frac{1}{\sigma^2(\tilde{\beta}_n(E))} \tag{3.32}$$

and normalising it to all simulations so far:

$$\kappa(E) = \frac{p(E)}{p(E) + p_n(E)}. \tag{3.33}$$

$p_n(E)$  is the sum of all previous  $p(E)$  and  $\kappa(E) = 0$  if  $p_n(E) = 0$  and/or  $p(E) = 0$ .

Matters can be simplified further if one rewrites the recursion in eq. (3.26) with eq. (3.25) to

$$\beta_{n+1}(E) = \beta_n(E) + \kappa(E) \frac{\ln H_n(E + \epsilon) - \ln H_n(E)}{\epsilon} \quad (3.34)$$

and use this together with the ratio of weights defined as

$$R(E) = \frac{W_{\text{muca}}(E)}{W_{\text{muca}}(E + \epsilon)} = \frac{e^{-\beta(E)E + f(E)}}{e^{-\beta(E+\epsilon)(E+\epsilon) + f(E+\epsilon)}} \quad (3.35)$$

to get an expression of the recursion in terms of the ratio of weights:

$$R_{n+1}(E) = R_n \left( \frac{H_n(E + \epsilon)}{H_n(E)} \right)^{\kappa(E)}. \quad (3.36)$$

This allows to calculate  $R_{n+1}(E)$  out of  $H_n(E)$ . We fixed  $W_{\text{muca},n+1}(E_{\text{max}}) = 1$  and got  $W_{\text{muca},n+1}(E)$  for all  $E$  from the ratios.

### 3.3 Reweighting

What we are interested in are canonical expectation values at different values of  $T$ ,  $\epsilon_s$  and  $s$ , respectively. To obtain them, one can save a considerable amount of CPU time by using each generated peptide configuration several times, i.e. calculate its contribution to each combination of  $T$ ,  $\epsilon_s$  and  $s$  of interest and average over those reweighted data. How this is done is described in more detail below.

The crucial point of this simple trick is to generate enough relevant configurations (configurations with a high probability  $p(E) = \Omega(E) \exp(-E/T)$ ) over the whole reweighted regime. This is necessary, since it is the relevant regime that mainly contributes to the average and low statistics here result in high statistical errors. For reweighting in  $T$  this is achieved with a the flat multicanonical energy histogram and how it is achieved for reweighting in  $\epsilon_s$  and  $s$  is explained in section 3.4.

#### 3.3.1 Reweighting in $T$

As long as enough relevant states are generated, it is always possible to obtain the canonical expectation values by averaging over the reweighted observables to the temperature of interest. More specifically, if a Markov chain is generated with probability distribution  $p_\mu = \Omega(E_\mu) W(E_\mu)$ , where e.g.  $W(E_\mu) = \exp(-E_\mu/T_0)$ ,  $k_B = 1$ , for a canonical simulation at temperature  $T_0$ , the canonical expectation value of an observable  $O$  at temperature  $T$  is given by eq. (3.3). Using this equation, all our canonical averages are obtained. In a multicanonical simulation  $W(E_\mu) = W_{\text{muca}}(E_\mu)$ , while for a canonical simulation at temperature  $T_0$  this expression simplifies further to

$$\langle O(E) \rangle (T) = \frac{\sum_\mu O(E_\mu) e^{-E_\mu/T} / e^{-E_\mu/T_0}}{\sum_\mu e^{-E_\mu/T} / e^{-E_\mu/T_0}} = \frac{\sum_\mu O(E_\mu) e^{-(1/T-1/T_0)E_\mu}}{\sum_\mu e^{-(1/T-1/T_0)E_\mu}}. \quad (3.37)$$

For systems with discrete energies, it is actually more common to store histograms  $H(E)$  and  $O(E)$  and obtain the canonical expectation values with a sum over all energies. But for models with a continuous energy spectrum, it is recommendable to sum over the time series since this works without any systematic discretisation errors. If the observable  $O$  does not explicitly depend on  $E$  the time series approach works as before with  $O(E_\mu)$  being the value of  $O$  at MC time step  $\mu$  and  $E_\mu$  being the corresponding energy. In the histogram approach,  $O(E)$  has to be replaced by the estimated multicanonical expectation value of  $O$  at fixed  $E$ .

Along the same lines of this well established reweighting procedure [47] for a reweighting in temperature, we derived a way to also reweight the canonical expectation values to different surface attraction strengths  $\epsilon_s$  and different intrapolymeric Lennard-Jones attraction strengths  $s$ .

### 3.3.2 Reweighting in $\epsilon_s$

Most of the results obtained in this thesis were obtained by multicanonical simulations with  $\epsilon_s$  and  $s$  set constant. This requires a long multicanonical run for every fixed combination of  $\epsilon_s$  and  $s$  and the canonical expectation values at various  $T$  are found by reweighting.

But reweighting in  $\epsilon_s$  and  $s$  is also possible. To see this, take a closer look at the energy:

$$\begin{aligned}
 E &= \underbrace{\sum_{i=1}^{N-2} \sum_{j=i+1}^N \left( \frac{1}{r_{ij}^{12}} - s \frac{1}{r_{ij}^6} \right)}_{E_{\text{LJ},12} + s E_{\text{LJ},6}} + \underbrace{\frac{1}{4} \sum_{k=1}^{N-2} (1 - \cos \vartheta_k)}_{E_{\text{bend}}} + \epsilon_s \underbrace{\sum_{i=1}^N \left[ \frac{2}{15} \left( \frac{1}{z_i} \right)^9 - \left( \frac{1}{z_i} \right)^3 \right]}_{E_{\text{sur}}}. \quad (3.38)
 \end{aligned}$$

To calculate  $E_{\text{LJ},12}$ ,  $E_{\text{LJ},6}$ ,  $E_{\text{bend}}$  and  $E_{\text{sur}}$  all one needs is the configuration of the polymer. Once knowing those four quantities, the Boltzmann weight of the given configuration can be found for arbitrary  $T$ ,  $\epsilon_s$  and  $s$ . In order to simplify the interpretation of the data, only the influence of either  $\epsilon_s$  or  $s$  is analysed at a time while the other one is kept fixed. Choosing first  $s$  to be constant and renaming  $E_{\text{LJ},12} + s E_{\text{LJ},6} = E_{\text{LJ}}$ , the canonical weight reads

$$\begin{aligned}
 P_{\epsilon_s, T}(E_\mu) &\propto e^{-\frac{E_\mu, \epsilon_s}{T}} = e^{-\frac{E_{\text{LJ}, \mu} + E_{\text{bend}, \mu} + \epsilon_s E_{\text{sur}, \mu}}{T}} \\
 &= e^{-\frac{E_{\text{LJ}, \mu} + E_{\text{bend}, \mu} + \epsilon_s E_{\text{sur}, \mu}}{T}} e^{-\frac{E_{\text{LJ}, \mu} + E_{\text{bend}, \mu} + \epsilon_s E_{\text{sur}, \mu}}{T} + \frac{E_{\text{LJ}, \mu} + E_{\text{bend}, \mu} + \epsilon_{s_0} E_{\text{sur}, \mu}}{T}} \\
 &= e^{-\frac{(\epsilon_s - \epsilon_{s_0}) E_{\text{sur}, \mu}}{T}} P_{\epsilon_{s_0}, T}(E_\mu)
 \end{aligned} \quad (3.39)$$

So, if the simulation was done at  $\epsilon_{s_0}$ , the expectation value corresponding to

neighboring  $\epsilon_s$  can be found via

$$\begin{aligned}
\langle O \rangle_{\text{can}}(T, \epsilon_s) &= \frac{\sum_{\mu \in \mathcal{M}} O_\mu(E_{\mu, \epsilon_s}) P_{\epsilon_s, T}(E_\mu)}{\sum_{\mu \in \mathcal{M}} P_{\epsilon_s, T}(E_\mu)} \\
&= \frac{\sum_{\mu \in \mathcal{M}} O_\mu(E_{\mu, \epsilon_s}) P_{\epsilon_{s_0}, T}(E_\mu) e^{-\frac{(\epsilon_s - \epsilon_{s_0}) E_{\text{sur}, \mu}}{T}}}{\sum_{\mu \in \mathcal{M}} P_{\epsilon_{s_0}, T}(E_\mu) e^{-\frac{(\epsilon_s - \epsilon_{s_0}) E_{\text{sur}, \mu}}{T}}}.
\end{aligned} \tag{3.40}$$

### 3.3.3 Reweighting in $s$

Similarly, starting with eq. (2.12) and choosing  $\epsilon_s = 1$ , the canonical weight for arbitrary  $s$  is given by

$$\begin{aligned}
P_{s, T}(E_\mu) &\propto e^{-\frac{E_{\mu, s}}{T}} = e^{-\frac{E_{\text{LJ}, 12, \mu} + s E_{\text{LJ}, 6, \mu} + E_{\text{bend}, \mu} + E_{\text{sur}, \mu}}{T}} \\
&= e^{-\frac{E_{\text{LJ}, 12, \mu} + s_0 E_{\text{LJ}, 6, \mu} + E_{\text{bend}, \mu} + E_{\text{sur}, \mu}}{T}} \\
&\quad \times e^{-\frac{E_{\text{LJ}, 12, \mu} + s E_{\text{LJ}, 6, \mu} + E_{\text{bend}, \mu} + E_{\text{sur}, \mu}}{T} + \frac{E_{\text{LJ}, 12, \mu} + s_0 E_{\text{LJ}, 6, \mu} + E_{\text{bend}, \mu} + E_{\text{sur}, \mu}}{T}} \\
&= e^{-\frac{(s - s_0) E_{\text{LJ}, 6, \mu}}{T}} P_{s_0, T}(E_\mu),
\end{aligned} \tag{3.41}$$

which yields for the corresponding expectation values

$$\begin{aligned}
\langle O \rangle_{\text{can}}(T, s) &= \frac{\sum_{\mu \in \mathcal{M}} O_\mu(E_{\mu, s}) P_{s, T}(E_\mu)}{\sum_{\mu \in \mathcal{M}} P_{s, T}(E_\mu)} \\
&= \frac{\sum_{\mu \in \mathcal{M}} O_\mu(E_{\mu, s}) P_{s_0, T}(E_\mu) e^{-\frac{(s - \epsilon_{s_0}) E_{\text{LJ}, 6, \mu}}{T}}}{\sum_{\mu \in \mathcal{M}} P_{s_0, T}(E_\mu) e^{-\frac{(s - s_0) E_{\text{LJ}, 6, \mu}}{T}}}.
\end{aligned} \tag{3.42}$$

## 3.4 Replica Exchange Monte Carlo

When using the reweighting in  $\epsilon_s$  or  $s$ , one has the problem, that those are fixed parameters inserted into the simulation. So if we, e.g., perform a multicanonical simulation at  $s$  such that we obtain all expectation values for all  $T$  at that  $s$  by reweighting in  $T$  and want to also have all expectation values for all  $T$  at  $s + \delta s$ , one has to be careful<sup>2</sup>. The problem is, that there might be configurations that are important at  $s + \delta s$ , that are not so important at  $s$  and hence have not been sufficiently sampled. This is well known for the reweighting in  $T$  in canonical simulations where the reweighting range, in that one can still expect reliable results, is limited to the width of the input histogram [47]. Multicanonical simulation resolves that problem.

For  $s$  we use a technique called *replica exchange Monte Carlo* (REMC). REMC appears to have been discovered independently by various researchers [49; 50]

<sup>2</sup>We will only argue for  $s$  here, not for  $\epsilon_s$ , since the calculation is completely analogous.

and is also known as parallel tempering, multiple Markov chain Monte Carlo and exchange Monte Carlo search. It has been successfully applied to the off-lattice protein folding problem [51] and several other applications can be found in Ref. [52].

In short, we perform several simulations with several not too different  $s$  at the same system and reweight to the  $s$  in between. Every so often, one swaps the states of the system in two of the simulations with a certain probability which is chosen so that the states of each system still follow the distribution  $p_\mu$  one wants to have, in our case the flat distribution obtained by the multicanonical weights.

Consider two multicanonical simulations that run in parallel, one with  $s = s_1$  the other one with  $s = s_2$ . Since the multicanonical weights are different for different  $s$ , we denote the corresponding weights by  $W_{\text{muca},1}(E_1)$  and  $W_{\text{muca},2}(E_2)$  and the energies in both systems as well as its constituents get the same indices. On the majority of time steps, we simply do one step in the simulation of each system. But, every so often, we want to swap the states, i.e. the values of the coordinates in each of the two simulations are set to those in the other. If simulation 1 has configuration  $\mu$  ( $\mu_1$ ) and simulation 2 has configuration  $\nu$  ( $\nu_2$ ), and a swap is suggested, the acceptance probability is

$$A(\mu_1\nu_2 \rightarrow \nu_1\mu_2) = \begin{cases} p_{\nu_1\mu_2}/p_{\mu_1\nu_2} & \text{if } p_{\nu_1\mu_2}/p_{\mu_1\nu_2} < 1, \\ 1 & \text{otherwise.} \end{cases} \quad (3.43)$$

The proof that this satisfies ergodicity and detailed balance can, e.g., be found in Ref. [41]. In the case of two parallel canonical simulations at  $\beta_{\text{low}}$  and  $\beta_{\text{high}}$ ,  $p_{\nu_1\mu_2}/p_{\mu_1\nu_2}$  simplifies to  $p_{\nu_1\mu_2}/p_{\mu_1\nu_2} = \exp[-(\beta_{\text{low}} - \beta_{\text{high}})\Delta E]$ , with  $\Delta E = E_{\text{high}} - E_{\text{low}}$ . For our two multicanonical simulations at  $s_1$  and  $s_2$  with different weights, this gets a bit more complicated:

$$\begin{aligned} \frac{p_{\nu_1\mu_2}}{p_{\mu_1\nu_2}} &= \frac{p_{\nu_1}}{p_{\mu_1}} \cdot \frac{p_{\mu_2}}{p_{\nu_2}} = \frac{W_{\text{muca},1}(E_{\text{LJ},12,\nu} + s_1 E_{\text{LJ},6,\nu} + E_{\text{bend},\nu} + E_{\text{sur},\nu})}{W_{\text{muca},1}(E_{\text{LJ},12,\mu} + s_1 E_{\text{LJ},6,\mu} + E_{\text{bend},\mu} + E_{\text{sur},\mu})} \\ &\quad \times \frac{W_{\text{muca},2}(E_{\text{LJ},12,\mu} + s_2 E_{\text{LJ},6,\mu} + E_{\text{bend},\mu} + E_{\text{sur},\mu})}{W_{\text{muca},2}(E_{\text{LJ},12,\nu} + s_2 E_{\text{LJ},6,\nu} + E_{\text{bend},\nu} + E_{\text{sur},\nu})}, \end{aligned} \quad (3.44)$$

where the notations are analogous to eq. (3.38).

This method is used for the results presented in section 4.2. Two parallel simulations at similar  $s$  were performed and swapped from time to time. Expectation values in between are calculated using all generated configurations and presented. Probably, it would also be sufficient to calculate the expectation values from the data of two independent simulations that do not swap configurations. This should provide sufficiently many relevant configurations if the simulation is long enough. However, swapping from time to time helps a bit to prevent the simulation to get stuck and does at least not cost much if it is not done too often, e.g., 1000 independent sweeps followed by one swap move.

### 3.5 Energy Landscape Paving (ELP)

Although it is not the primary goal here to search for global energy minima, it is very instructive to investigate how the nature of the energy minima changes with  $\epsilon_s$ . In order to do so, an algorithm specialised on finding global energy minima was applied: The Energy Landscape Paving (ELP) [53]. It is easily implemented and has successfully proven to be applicable to find global energy minima in rough energy landscapes of AB heteropolymers [43]. Its central idea is to perform low-temperature canonical Monte Carlo simulations, but with an energy expression that is modified after each step in order to steer the search away from regions that have already been explored. Here, we use for the modified statistical weight of a state the simple version

$$w(\tilde{E}) = e^{-\tilde{E}/T} \quad \text{with} \quad \tilde{E} = E + H(E, t), \quad (3.45)$$

where  $H(E, t)$  is the histogram of energies at MC step  $t$ . In a regular low-temperature Metropolis simulation the probability to escape a local minimum depends only on the height of the surrounding energy barriers. ELP locally flattens the energy landscape by filling up such local energy minima. This again decreases the weight of states within those minima and consequently increases the probability to escape. So, initially when  $H(E, t) \approx 0$  ELP will favour low energies and avoids the sampling of unphysical high-energy conformations. After getting stuck in a local energy minimum, the energy landscape gets deformed and higher energies will be explored until eventually another local energy minimum is found or the higher energy histogram entries have similar frequencies and the original energy landscape gets approximately restored up to a constant irrelevant factor. Due to this bias, ELP violates detailed balance and is therefore inappropriate to unravel thermodynamic properties of the system. It also cannot distinguish between different structures of the same energy. However to find low-energy states, it is efficient and easy to implement.

### 3.6 Error estimation

Any data, experimentally determined or simulated, can only be trusted, if they come along with a reliable error estimation.

Markov chain MC update algorithms have autocorrelation times that enter directly into the statistical errors and can be dealt with using autocorrelation analysis as will be explained in the first subsection. This, however, becomes quite cumbersome for quantities that are not directly measured in the simulation but a nonlinear combination of those directly measured quantities, e.g., the heat capacity. In this case, error propagation has to be applied, if one wants to be very precise, but usually another simpler approach is preferred. This is the Jackknife method shortly explained in section 3.6.2. All our errors are Jackknife errors.



### 3.6.1 Autocorrelation Time

Successive states of Markov chain MC methods are correlated. This is immediately obvious considering any of the updates introduced above. The polymer after one update is still in a configuration very similar to the configuration before. Thus the variance of estimates produced from Markov chain MC simulations may be much higher than from the same amount of configurations that are sampled independently – without knowing how the configuration before looked like. To quantify this effect, the autocorrelation function is introduced:

$$A(k) = \frac{\langle O_i O_{i+k} \rangle - \langle O_i \rangle \langle O_{i+k} \rangle}{\langle O_i^2 \rangle - \langle O_i \rangle^2}. \quad (3.46)$$

The expectation value  $\langle O_i O_{i+k} \rangle$  is the correlation between observable  $O$  at time  $i$  and a later time  $i+k$ . If the value of  $O$  at both times is uncorrelated, the correlation factorises to the product of the individual expectation values  $\langle O_i O_{i+k} \rangle = \langle O_i \rangle \langle O_{i+k} \rangle$  and the autocorrelation vanishes. The denominator is just for normalisation, i.e.  $A(0) = 1$ . In equilibrium, time translation invariance holds. Hence  $\langle O_i \rangle = \langle O_{i+k} \rangle$  and in the numerator  $\langle O_i \rangle \langle O_{i+k} \rangle$  simplifies further to  $\langle O_i \rangle^2$ . The autocorrelation function is a measure for the similarity of an observable at time  $i$  with itself at time  $i+k$ . For small  $k$ ,  $A(k) \approx 1$  since each state still resembles the preceding one. On the other hand, for very large  $k$ , any configuration might have been reached and  $A(k) \approx 0$ . For not too small  $k$ , the autocorrelation decays exponentially,

$$A(k) \stackrel{k \rightarrow \infty}{\approx} a e^{-k/\tau_{\text{exp}}}, \quad (3.47)$$

which defines the *exponential* autocorrelation time  $\tau_{\text{exp}}$  and  $a$  is some constant. Even more useful is another definition of autocorrelation time: the *integrated* autocorrelation time

$$\tau'_{\text{int}} = \frac{1}{2} + \sum_{k=1}^N A(k) \left(1 - \frac{k}{N}\right) \stackrel{N \gg \tau_{\text{exp}}}{\approx} \frac{1}{2} + \sum_{k=1}^N A(k) = \tau_{\text{int}}. \quad (3.48)$$

Here it turns out that

$$\sigma_O^2 = \frac{\sigma_{O_i}^2}{N} 2\tau'_{\text{int}} \approx \frac{\sigma_{O_i}^2}{N} 2\tau_{\text{int}} \quad (3.49)$$

as is derived in [54]<sup>3</sup>.  $\sigma_{O_i}^2 = \langle O_i^2 \rangle - \langle O_i \rangle^2$  is the variance of the individual measurements and in the case of uncorrelated measurements  $\sigma_O^2 = \sigma_{O_i}^2/N$ . Hence the variance is increased by a factor of  $2\tau_{\text{int}}$  for correlated data. This effect can be quite significant. It is instructive to introduce a parameter

$$N_{\text{eff}} = \frac{N}{2\tau_{\text{int}}} \leq N \quad (3.50)$$

<sup>3</sup>In this definition of  $\tau_{\text{int}}$ ,  $\tau_{\text{int}} = 1/2$  for uncorrelated measurements. Some authors define  $\tau_{\text{int}} = 0$  for uncorrelated measurement, which is a bit more intuitive, but eq. (3.49) wouldn't look that nice.

that gives the *effective* statistics, i.e., the number of uncorrelated data one obtains from  $N$  measurements that are correlated with an integrated autocorrelation time  $\tau_{\text{int}}$ . Hence only every  $2\tau_{\text{int}}$  iterations, the produced data are uncorrelated again. This knowledge is an important input of the Jackknife method described below. What is left to explain is how to determine  $\tau_{\text{int}}$ . An estimator  $\tilde{A}(k)$  of  $A(k)$  is obtained by substituting the expectation values  $\langle O_i O_{i+k} \rangle$  by the mean values  $\overline{O_i O_{i+k}}$ . This, in practice leads to very noisy tails of  $\tilde{A}(k)$ , since there is less statistics for large time separations  $k$ . Summing over all available  $\tilde{A}(k)$  to obtain  $\tau_{\text{int}}$  would thus introduce a considerable error. One still gets a decent estimate of  $\tau_{\text{int}}$  by introducing a cut-off

$$\tilde{\tau}_{\text{int}}(k_{\text{max}}) = \frac{1}{2} + \sum_{k=1}^{k_{\text{max}}} \tilde{A}(k). \quad (3.51)$$

This approaches  $\tau_{\text{int}}$  in the limit of large  $k_{\text{max}}$ . But as soon as  $\tilde{A}(k)$  gets very small, it reaches a plateau and its statistical error increases rapidly. A compromise between those systematic and statistical errors is to determine an optimal  $k_{\text{max}}$  self-consistently by cutting off as soon as  $k_{\text{max}} \geq 6\tilde{\tau}_{\text{int}}(k_{\text{max}})$  as done here. Other choices such as cutting off as soon as  $\tilde{A}(k)$  first subtends the base line can also be applied.

It is important to notice, that the autocorrelation time can be significantly reduced by choosing efficient Monte Carlo algorithms. So, for instance, a multi-canonical simulation, that ideally executes a random walk in energy space, has a lower autocorrelation than the Metropolis algorithm, that more or less stays in the same energy regime. Hence, it produces effectively more statistics.

Also note, that while  $\tau_{\text{exp}}$  only depends on the algorithm but is theoretically independent of the observable under consideration,  $\tau_{\text{int}}$  depends on the observable. More mathematical details can be found in [41; 58]. We only performed a short autocorrelation analysis in order to have a rough estimate, how to choose the bins for the Jackknife analysis.

### 3.6.2 Blocking Jackknife Technique

To employ the whole autocorrelation analysis described above would result in a considerable effort to get accurate statistical errors. Thus, general but less accurate methods have been developed to estimate the error of a Monte Carlo run on a daily basis. These are the Binning and the Jackknife methods [54] that both divide the time series of  $N$  correlated measurements  $O_i$  in blocks. The Binning analysis considers a number  $N_B = N/k$  of equidistant blocks of length  $k$  (Fig. 3.3):

$$O_{B,n} \equiv \frac{1}{k} \sum_{i=1}^k O_{(n-1)k+i}, \quad n = 1, \dots, N_B. \quad (3.52)$$

The idea of this binning is to choose  $k$  larger than the autocorrelation time and thus to create with  $O_{B,n}$ ,  $n = 1, \dots, N_B$ , a new, shorter time series which is

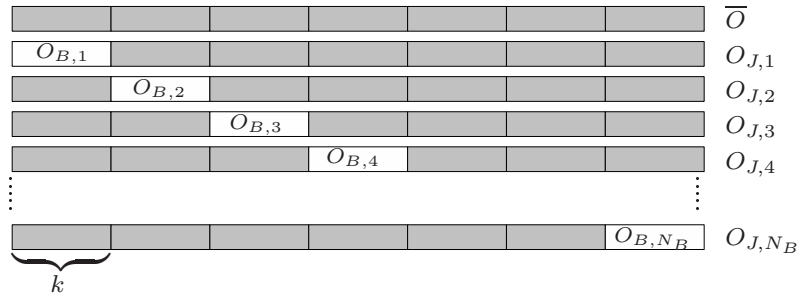


Figure 3.3: A schematic representation of the block Jackknife error estimation. The block on the top represents all  $N$  correlated measurements. Its block average is denoted by  $\bar{O}$ . In the following blocks, the grey blocks represent the Jackknife blocks with block average  $O_{J,n}$  and the white blocks the binning blocks with block average  $O_{B,n}$ ,  $n = 1, \dots, N_B$ , and block length  $k$ . It holds:  $N = N_B k$ .

almost uncorrelated and can thus be analysed by standard means. The Jackknife blocks  $O_{J,n}$  contain all data, but the ones of the binning blocks,

$$O_{J,n} \equiv \frac{N\bar{O} - kO_{B,n}}{N - k}, \quad n = 1, \dots, N_B. \quad (3.53)$$

$\bar{O}$  denotes the common mean value of the all  $N$  measurements. The advantage of those larger blocks is that the statistics of each block is better and fluctuations of individual blocks are reduced compared to the binning approach. For linear quantities like energy or radius of gyration there is no difference in the estimated error and both methods give an analytically equal result. Differences however occur, when nonlinear quantities like the specific heat are treated due to the bias of the estimator that reduces for larger sets of data as  $1/N$ . After correction for the trivial correlation of the jackknife bins – every value of the time series is used  $N - 1$  times – the Jackknife error of  $\bar{O}$  is given by

$$\epsilon_{\bar{O}}^2 \equiv \sigma_{\bar{O}}^2 = \frac{N_B - 1}{N_B} \sum_{n=1}^{N_B} (O_{J,n} - \bar{O})^2. \quad (3.54)$$

It is beneficial to choose  $k$  rather large.  $k \gg \tau_{\text{int}}$  has to hold and it is an empirical rule that  $k \approx 6\tau_{\text{int}}$  gives good results. Choosing  $k$  even larger does not alter  $\epsilon_{\bar{O}}^2$  much, but choosing it too small underestimates the error.



## 4 Results

In this chapter, various informations about the canonical equilibrium behaviour of short polymers near an attractive substrate are presented. The main focus is on the behaviour at varying surface attraction strength and temperature.

The overall goal aimed at is to summarise all those informations and construct a pseudophase diagram. This pseudophase diagram should contain information about the quality and position of conformational phases present and indicate transitions between them. We want to stress that all phases and transitions mentioned here are no phases in the strict thermodynamic sense, since we are dealing with finite chain lengths. But even for those short chains that are considered here, we obtain a good picture about the behaviour of the polymer at the surface and most of the phases are believed to still exist for longer chains.

### 4.1 Adsorption Behaviour for Various Surface Attraction Strengths

In this section, I present the results gained for various observables of two exemplified short peptide sequences with 13 and 20 monomers, respectively. Multicanonical simulations at 51 different surface attraction strengths  $\epsilon_s$ , ranging from  $\epsilon_s = 0, \dots, 5$ , were performed and reweighted to  $T = 0, \dots, 5$ . Since the main structural activity takes places below  $T = 3$ , usually only the lower temperature regime is displayed. Every simulation consists of  $10^8$  sweeps and was at least performed with 2 different random number seeds to assure that trapping that sometimes occurred, does not entail any systematic deviation. The interpolymeric interaction is regarded constant here with  $s = 1$ .

All implemented updates and boundary conditions were tested against several independent works on short hydrophobic-polar heteropolymers of the AB model [44; 55; 56; 57].

#### 4.1.1 Observables

##### Energy and Specific Heat

The total energy of the single polymer has already been defined in eq. (2.12). With  $s = 1$ , this simplifies to

$$E = 4 \sum_{i=1}^{N-2} \sum_{j=i+1}^N \left( \frac{1}{r_{ij}^{12}} - \frac{1}{r_{ij}^6} \right) + \frac{1}{4} \sum_{k=1}^{N-2} (1 - \cos \vartheta_k) + \epsilon_s \sum_{i=1}^N \left[ \frac{2}{15} \left( \frac{1}{z_i} \right)^9 - \left( \frac{1}{z_i} \right)^3 \right]. \quad (4.1)$$

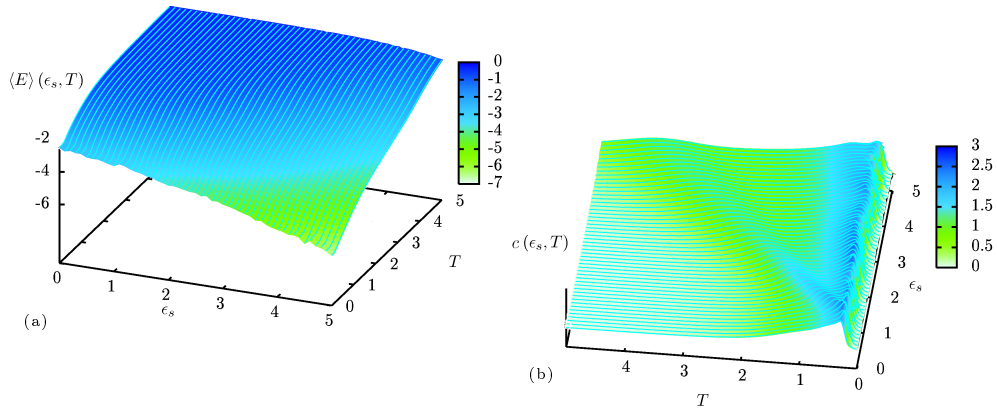


Figure 4.1: (a) Energy of the 20mer. (b) Heat capacity of the 20mer. The small lines represent the simulated data, the colour code is interpolated.

In Fig. 4.1, a 3D plot of the energy and specific heat of the 20mer vs.  $\epsilon_s$  and  $T$  is shown. The projections onto the surface-attraction-temperature plane, as obtained for the 13- and 20mer for both quantities, are plotted in Figs. 4.2 and 4.3. For both investigated polymer lengths the energy varies smoothly with changing  $\epsilon_s$  and  $T$ . The global minimum is reached at maximal surface attraction and minimal temperature. This is not unexpected since at low temperatures, energy dominates over entropy and hence lower energy conformations are more frequently assumed. Also for ever higher surface attractions, any given conformation close to or at the surface corresponds to lower energies.

Although the total energy varies smoothly with  $\epsilon_s$  and  $T$ , this does not hold true for the individual compositions, because many internal monomer-monomer

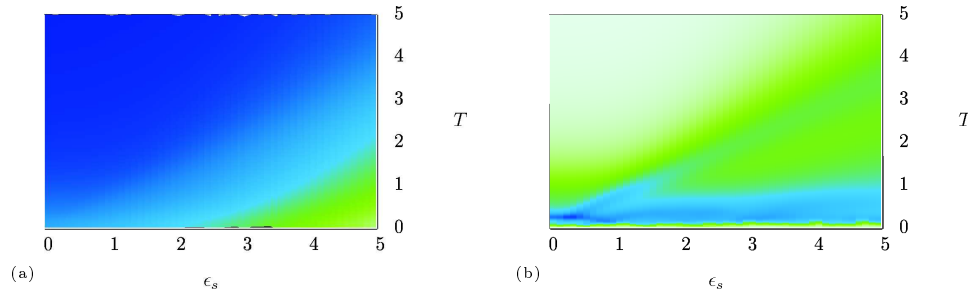


Figure 4.2: (a) Energy of the 13mer. (b) Heat capacity of the 13mer. The colour code is as in Fig. 4.1.

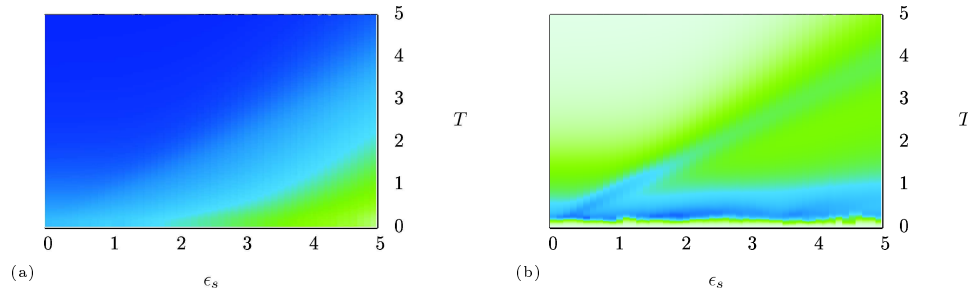


Figure 4.3: (a) Energy of the 20mer. (b) Heat capacity of the 20mer. The colour code is as in Fig. 4.1.

contacts correspond to a high intrinsic energy while many monomer-surface contacts correspond to a high surface contribution to the energy (see section 4.1.6). Those quantities indeed perform considerable fluctuations and jumps as will be discussed below.

The heat capacity turns out to be insufficient to characterise all phase transitions, since the chains are very short. Only two transitions can be identified as ridges in the profile: The first one is the *adsorption transition* between desorbed and adsorbed configurations. Where exactly in the projection plane desorbed and adsorbed configurations dominate is among others displayed below in the pseudophase diagram in Fig. 4.26. The adsorption transition exhibits indications to be first-order like, as it is characteristic for such short chains. For an sc lattice model it was shown in [15], that for a finite chain length (179mer) the free energy minima of adsorbed and desorbed conformations are separated by a gap when plotted versus monomer-monomer and monomer-surface contacts, i.e. none of the possible conformations in between are stable in equilibrium which reflects the first-order like behaviour for finite chains. But The other transition that can clearly be identified by the strong ridges in the specific heat landscape is a *freezing transition* at low temperatures. At roughly  $T \approx 0.25$  the heat capacity exhibits a strong peak and rapidly goes to zero at lower temperatures independent of the surface attraction strength. This and the crystalline structures found at those low temperatures indicate a freezing transition. Although the freezing temperature seems to be rather constant, the type of crystalline structure adapted by the peptide depends strongly on  $\epsilon_s$ . But to identify the shape of the peptide one has to take a closer look at conformational quantities like the radius of gyration.

### Radius of Gyration

The radius of gyration provides an excellent measure of the overall compactness of the polymer. Figures 4.5, 4.6 and 4.4 reveal that the most compact conformations dominate at low temperatures and low surface attractions. If the surface attraction is weak the polymer behaves like a free polymer and the transition from globular (DG) to random coil (DE) configurations (Fig. 4.26) corresponds to the well known  $\theta$ -transition [6; 5], where the repulsive excluded volume effect balances the attractive forces between the segments and the polymer behaves like an ideal chain (see chapter 1). For the 13mer we found  $\theta_{13} \simeq 0.94$  and for the 20-polymer  $\theta_{20} \simeq 1.28$  from the peaks in  $d\langle R_{\text{gyr}} \rangle / dT$  at  $\epsilon_s = 0$ . The higher value for the larger polymer is due to finite size effects. The overall polymer-polymer interaction per monomer (that corresponds to  $\epsilon_{pp}$ ) is higher for longer chains, because less monomers are at the outer part in the energetically favourable compact conformations which leads to a higher  $\epsilon_{pp}$ . This again increases the  $\theta$ -temperature.

Also the freezing transition can be found at the same temperatures as it was already clearly visible in the heat capacity. The adsorption transition hardly effects the overall size of the polymer and can thus hardly be seen in the radius

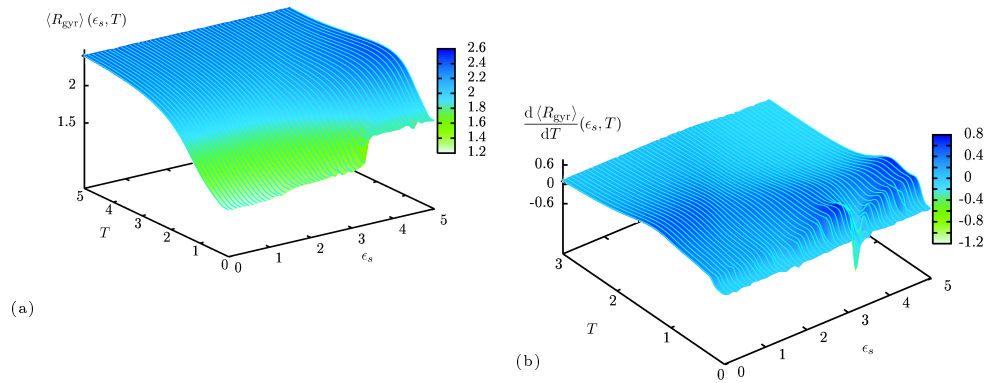


Figure 4.4: (a) Radius of gyration of the 20mer. (b)  $d\langle R_{\text{gyr}} \rangle/dT$  of the 20mer. The small lines represent the simulated data. The colour code is interpolated.

of gyration.

To extract further information out of the radius of gyration, it is more illustrative to take a closer look at its parallel (Figs. 4.7 and 4.8) and perpendicular (Figs. 4.12 and 4.13) component to the surface, respectively. These quantities are expected to behave different since the surface introduces a spatial anisotropy. So for instance, for  $\epsilon_s \geq 2.8$  for the 13mer and  $\epsilon_s \geq 3.4$  for the 20mer and low temperatures  $\langle R_{\perp} \rangle$  vanishes while  $\langle R_{\parallel} \rangle$  attains low values at lower  $\epsilon_s$ . A vanishing  $\langle R_{\perp} \rangle$  corresponds to a configuration where the polymer is spread out flat on the surface without any extension into the third dimension. These configurations

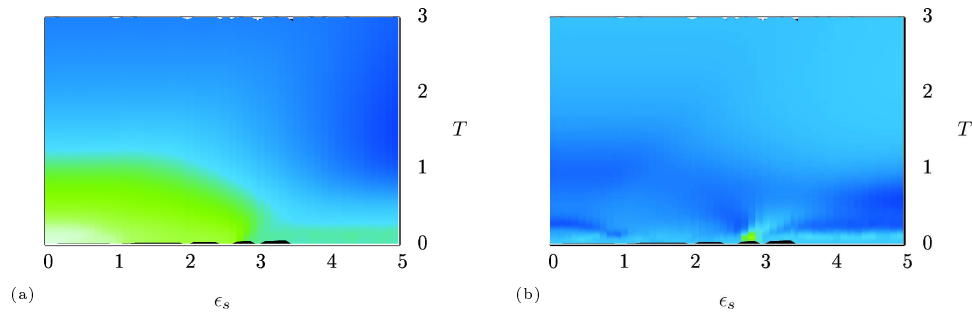


Figure 4.5: (a) Radius of gyration of the 13mer. (b)  $d\langle R_{\text{gyr}} \rangle/dT$  of the 13mer. The colour code is as in Fig. 4.4.

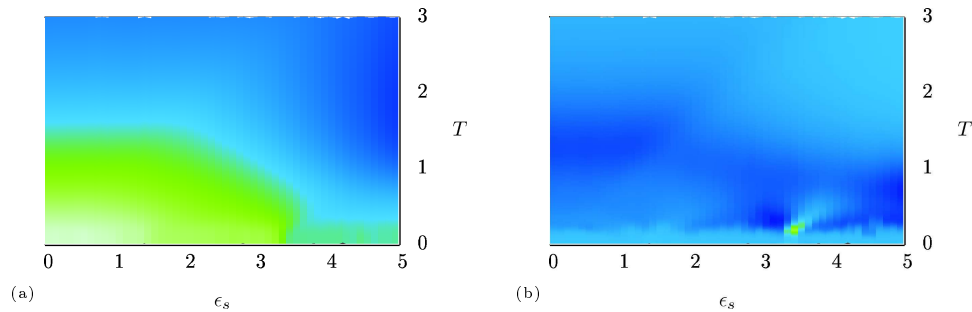


Figure 4.6: (a) Radius of gyration of the 20mer. (b)  $d\langle R_{\text{gyr}} \rangle/dT$  of the 20mer. The colour code as in Fig. 4.4.



are called adsorbed compact (AC1) and adsorbed expanded (AE1) (Fig. 4.26). The ‘1’ is added to distinguish those regions from regions that extend into the third dimension. AC1 and AE1 are separated by the freezing transition such that configurations in AC1 at lower temperatures are maximally compact while configurations in AE1 are less compact and more flexible but still mainly flat on the surface.

In order to confirm that conformations in AC1 are indeed maximally compact single layers, one can consider a simple argument. It is well known that the most compact shape in the 2D continuous space is the circle. Thus one can calculate  $\langle R_{\parallel} \rangle$  for a circle and compare it with the simulated value. Assuming  $N$  monomers to be distributed evenly in the circle,  $N \approx \pi r^2$ , where  $r$  is the radius of the circle in units of the mean distance of neighbouring monomers. The radius of gyration in the same units is thus given by

$$R_{\text{gyr}}^{\text{circ}^2} (\approx R_{\parallel}^2) = \frac{1}{\pi r^2} \int_{r' \leq r} d^2 r' r'^2 = \frac{1}{2} r^2 \approx \frac{N}{2\pi}. \quad (4.2)$$

We have two different types of mean distances between monomers in compact conformations. Neighboring monomers on the chain have distance one, while for all others the most favourable distance is  $r_{\text{min,LJ}} \approx 1.1225$ . So we expect for compact  $\langle R_{\parallel} \rangle$  on the surface to hold:  $\sqrt{13/2\pi} \approx 1.438 < \langle R_{\parallel,13} \rangle < 1.615 < r_{\text{min,LJ}} \sqrt{13/2\pi}$  and  $\sqrt{20/2\pi} \approx 1.784 < \langle R_{\parallel,20} \rangle < 2.026 < r_{\text{min,LJ}} \sqrt{20/2\pi}$ . The simulated data are  $R_{\parallel,13} = 1.45$  and  $R_{\parallel,20} = 1.81$  that nicely fit the estimate.

Trying the same thing in 3D however does not work that well. The most compact shape in 3D is the sphere, that we assume to be filled uniformly with  $N$  monomers,  $N = 4\pi r^3/3$ . Corresponding configurations are found as free compact chains (DC) as well as adsorbed compact configurations (AC) for weak surface attraction. Here the radius of gyration is given by

$$R_{\text{gyr}}^{\text{ball}^2} = \frac{1}{4\pi r^3/3} \int_{r' \leq r} d^3 r' r'^2 = \frac{3}{5} r^2 \approx \frac{3}{5} \left( \frac{3N}{4\pi} \right)^{1/3}. \quad (4.3)$$

This leads to the estimate  $1.130 < R_{\text{gyr},13} < 1.268$  and  $1.684 < R_{\text{gyr},20} < 1.464$ . But the simulated data are smaller:  $R_{\text{gyr},13} = 1.023$  and  $R_{\text{gyr},20} = 1.242$ . This slight deviation can however be explained by the fact, that the mass of the polymer is not uniformly distributed in the sphere as assumed in the calculation. For a compact packing of discrete monomer positions, it is far more realistic, that the outer thin shell of the sphere does not contain any monomers. Performing the integration not from  $r' = 0$  to  $r' = r$  but only to  $r' = r - \varepsilon$ , reduced the estimated radius of gyration significantly already for small  $\varepsilon$  due to the increased weight of the outer shells in higher dimensions. Taking into account this effect, the absolute values of  $\langle R_{\text{gyr}} \rangle$  obtained, seem to be very reasonable.

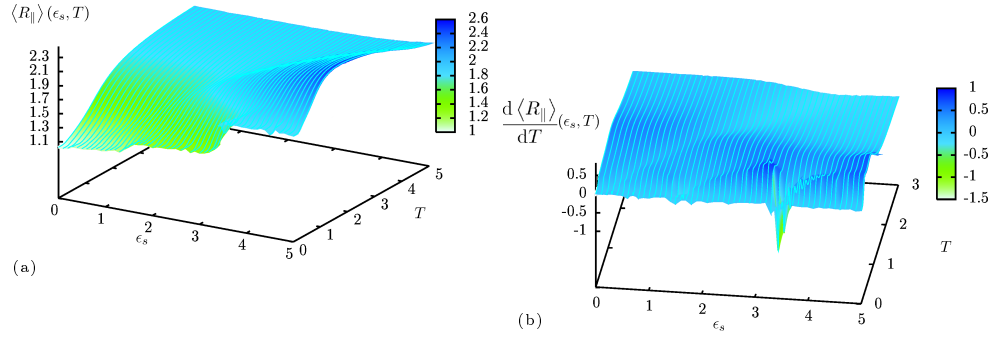


Figure 4.7: (a)  $\langle R_{\parallel} \rangle$  of the 20mer. (b)  $d\langle R_{\text{gyr},\parallel} \rangle/dT$  of the 20 polymer. The small lines represent the simulated data. The colour code is interpolated.

The most noticeable transition that can be seen in the  $R_{\text{gyr}}$ -components is the strong layering transition at  $\epsilon_s \approx 2.8$  for  $N = 13$  and  $\epsilon_s \approx 3.3$  for  $N = 20$  that separates regions of planar conformations (AE1, AC1) in the surface attraction-temperature plane from more compact 3D conformations (AG, AC2b) at low temperatures.

In a paper of J. Krawczyk *et al.* [16] a lattice argument can be found that nicely illustrates what happens here. The polymer is modeled as a self-avoiding walk on a 3D cubic lattice in a half-space interacting via a nearest-neighbour interaction of different monomers with  $\epsilon_m$  being the contributions per contact. In addition each polymer-surface contact lowers the energy by another  $\epsilon_s$  such that the overall energy of a configuration with  $n_s$  surface contacts and  $n_m$  intrinsic con-

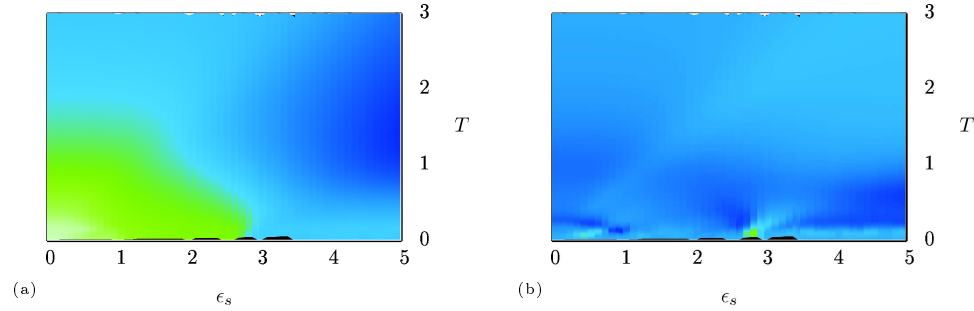


Figure 4.8: (a)  $\langle R_{\parallel} \rangle$  of the 13mer. (b)  $d\langle R_{\text{gyr},\parallel} \rangle/dT$  of the 13mer. The colour code is as in Fig. 4.7.

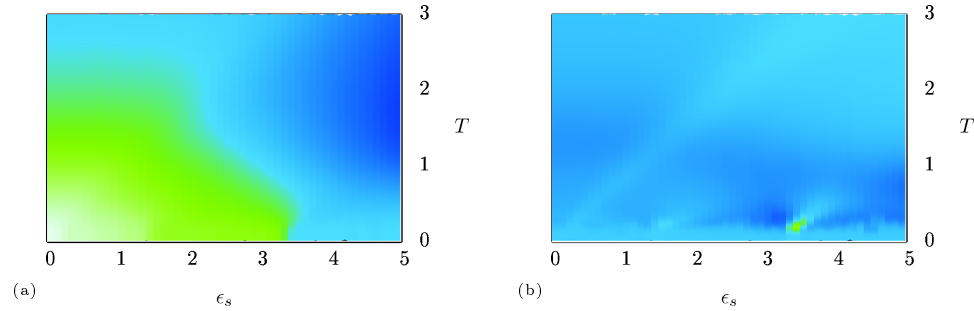


Figure 4.9: (a)  $\langle R_{\parallel} \rangle$  of the 20 polymer. (b)  $d\langle R_{\text{gyr},\parallel} \rangle/dT$  of the 20mer; colour code as in Fig. 4.7.

tacts is given by

$$E_N = -n_m \epsilon_m - n_s \epsilon_s. \quad (4.4)$$

Now one considers the zero-temperature situation, where for positive self-attraction and surface attraction, the polymer will take on some compact configuration touching the surface. For a cubic lattice model this is likely to be a rectangular parallelepiped with square cross-section parallel to the surface of site length  $w$  (Fig. 4.10). If the height to the surface is  $l$ , and there are  $N = lw^2$  monomers in the polymer, the total energy  $E_l$ , neglecting contributions from edges and corners) is

$$\begin{aligned} E_l(\epsilon_m, \epsilon_s) &\sim -(w^2(l-1) + 2(w(w-1))l)\epsilon_m - w^2\epsilon_s \\ &= -\left(\frac{N}{l}(l-1) + 2\left(\sqrt{\frac{N}{l}}\left(\sqrt{\frac{N}{l}} - 1\right)\right)l\right)\epsilon_m - \frac{N}{l}\epsilon_s \\ &= -N\epsilon_m + \frac{N}{l}\epsilon_m + 2\sqrt{Nl}\epsilon_m - N\epsilon_m - \frac{N}{l}\epsilon_s \\ &= -2\epsilon_m N + (\epsilon_m - \epsilon_s)\frac{N}{l} + 2\epsilon_m\sqrt{lN}. \end{aligned} \quad (4.5)$$

For fixed  $N$  the energy can be minimized with respect to  $l$  to see how the layer thickness depends on  $\epsilon_m$ ,  $\epsilon_s$  and  $N$ . One obtains for the minimum configuration

$$l^{3/2} = \left(1 - \frac{\epsilon_s}{\epsilon_m}\right) N^{1/2}. \quad (4.6)$$

A plot of  $l$  vs.  $\epsilon_s/\epsilon_m$  can be found in Fig. 4.11. Since there are only integer values of  $l$  possible – there is nothing like half a layer – a particular layer will be stable for a range of  $\epsilon_s/\epsilon_m$ . So, using this argument, for instance a single layer configuration is stable for  $\epsilon_s \geq \epsilon_m$  and for some values of  $\epsilon_s < \epsilon_m$  given by relation 4.6. If  $\epsilon_s$  is increased at fixed  $\epsilon_m$  the system's energy is minimised by

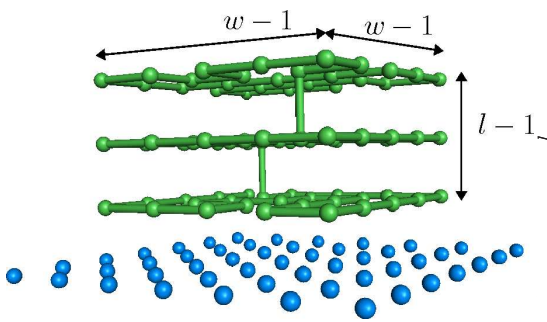


Figure 4.10: Exemplified compact conformation for a self-avoiding polymer on a 3D lattice with  $w = 5$  and  $l = 3$  to illustrate the argument of J. Krawczyk *et al.*.

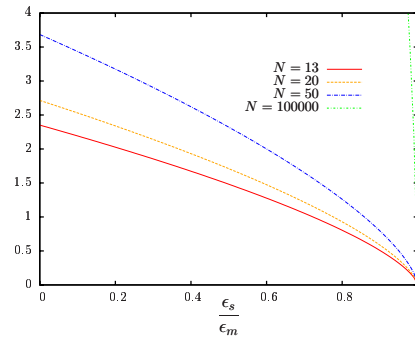


Figure 4.11: Plotted is  $l = (1 - \epsilon_s/\epsilon_m)^{2/3} N^{1/3}$  vs.  $\epsilon_s/\epsilon_m$  for four different chain lengths on the lattice. For  $N = 13$  only double- and single-layer structures are stable while for  $N = 20$  also triple-layer and  $N = 50$  even four layer structures seem to be energetically favourable for a low  $\epsilon_s/\epsilon_m$ -ratio.

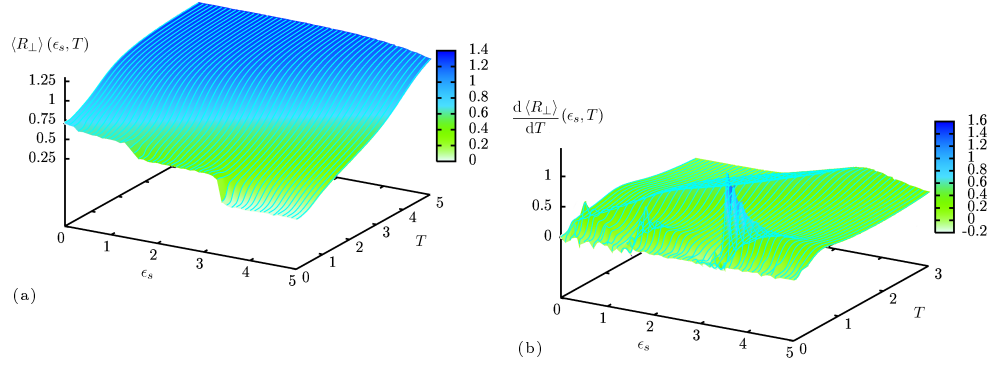


Figure 4.12: (a)  $\langle R_{\perp} \rangle$  of the 20mer. (b)  $d\langle R_{\text{gyr},\perp} \rangle/dT$  of the 20mer. The small lines represent the simulated data. The colour code is interpolated.

smaller values of  $l$ . This is exactly what is observed here with decreasing  $\langle R_{\perp} \rangle$  at low temperatures with increasing  $\epsilon_s$ . Since the total surface area is minimised if the layers are uniform layers with a roughly equal number of monomers in each layer, the polymer jumps from  $(l+1)$  layers to  $l$  layers at some value of  $\epsilon_s$  with a jump in the internal energy rather than continuously decreasing the occupation number of the top layer. This is also confirmed by the sharp layering transition from single- to double-layer structures for both chain length and ground state energy considerations and suggests that the transition is first-order like with a transition region that gets sharper for larger  $N$ , see Fig. 4.11.

However, one striking discrepancy between lattice theory and the data from our continuous model is observed: While the argument predicts that for  $\epsilon_s \geq \epsilon_m$  only single layer configurations are stable, our single layers are only stable for

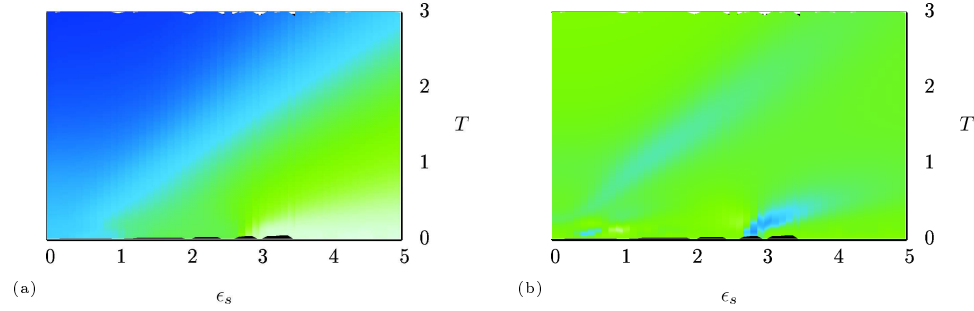


Figure 4.13: (a)  $\langle R_{\perp} \rangle$  of the 13mer. (b)  $d\langle R_{\text{gyr},\perp} \rangle/dT$  of the 13mer. The colour code is as in Fig. 4.12.

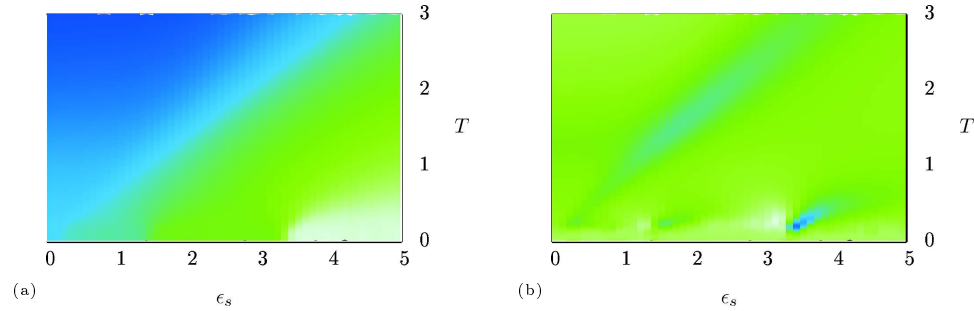


Figure 4.14: (a)  $\langle R_{\perp} \rangle$  of the 20mer, (b)  $d\langle R_{\text{gyr},\perp} \rangle/dT$  of the 20mer. The colour code is as in Fig. 4.12.

$\epsilon_s/\epsilon_m \gtrsim 3$ . We conjecture that this substantial difference is due to the different coordination number of both models. Inside the bulk of a compact polymer on an sc lattice, each monomer has maximally  $z = 6$  nearest neighbours while in the continuous model hexagonal close packing is observed at low temperatures with  $z = 12$  nearest neighbours. This gives more compact conformations with a higher  $l$  in our continuous model an additional stability. In fact, repeating the argument in eq. (4.5) for a hexagonal lattice, one obtains

$$\begin{aligned}
 E_l(\epsilon_m, \epsilon_s) &\sim -\left(4\frac{N}{l}(l-1) + 3\left(\sqrt{\frac{N}{l}}\left(\sqrt{\frac{N}{l}} - 1\right)\right)l\right)\epsilon_m - \frac{N}{l}\epsilon_s \\
 &= -4N\epsilon_m + 4\frac{N}{l}\epsilon_m + 3\sqrt{Nl}\epsilon_m - 3N\epsilon_m - \frac{N}{l}\epsilon_s \\
 &= -7\epsilon_m N + 4\frac{N}{l}\epsilon_m + 3\sqrt{Nl}\epsilon_m - \frac{N}{l}\epsilon_s.
 \end{aligned} \tag{4.7}$$

This yields for the equivalent of eq. (4.6) on a hexagonal lattice

$$l^{3/2} = \frac{2}{3}\left(4 - \frac{\epsilon_s}{\epsilon_m}\right)N^{1/2}, \tag{4.8}$$

which predicts a single-double layer transition for  $N = 13$  at  $\epsilon_s/\epsilon_m = 3.235$  and for  $N = 20$  at  $\epsilon_s/\epsilon_m = 3.384$ . Thus even if one approximates the more cylindrical shaped observed layered structures by a cubic, this argument gives a rather good estimate of the location of the layering transitions if a hexagonal lattice structure is assumed.

Although for the short chains considered in this section there are no triple layers observed, the components  $\langle R_{\parallel, \perp} \rangle$  indicate some activity at lower surface attraction. For  $N = 13$ ,  $\epsilon_s \approx 1.2$  is the lowest attraction strength at which there are still stable double layer configurations found below the freezing transition. For  $N = 20$  this is at  $\epsilon_s \approx 1.4$ . What follows at lower  $\epsilon_s$  is a low-temperature phase of surface attached compact conformations that we call AC2a. AC2a conformations occur if the monomer-surface attraction is not strong enough to induce a layering in the compact attached structure. One has to be careful in what exactly these conformations look like. On the one hand, the surface attraction is strong enough to attach the polymer, on the other hand the compact conformation of a free polymer below the  $\theta$ -transition shouldn't be distorted too much - no layering. We found two distinct structures that fit this scheme: 1. completely undistorted compact conformations located at the surface and 2. roughly semi-spherically shaped structures docked to the surface. Both are clearly observed over a substantial range of surface attractions  $\epsilon_s$ . Comparing where which kind of those two structures is observed at both chain length differences are found. For  $N = 13$  both,  $d\langle R_{\text{gyr}, \perp} \rangle/dT$  and  $d\langle R_{\text{gyr}, \parallel} \rangle/dT$  indicate a transition at  $\epsilon_s \approx 0.45$ . This transition is the wetting transition [13; 10], that has already attracted some interest in literature. We will meet this transition again, when discussing the effect of solvent variation (section 4.2). Here compact polymers that only move close to the surface at low  $T$  and lower  $\epsilon_s$  attach to form semi-spherically shaped structures at higher  $\epsilon_s$  that are stable

until at  $\epsilon_s \approx 1.2$  a seemingly continuous pseudophase transition to double layer structures takes place.

Without the additional information from the extracted low energy states, it would hardly be possible to locate this transition, although the radius of gyration as well as its components indeed change their absolute values at this  $\epsilon_s$  a little bit.

For  $N = 20$  no analog to the wetting transition at  $\epsilon_s \approx 0.45$  for  $N = 13$  was observed. Already at an attraction strength of  $\epsilon_s \approx 0.2$  adsorbed conformations are found at low temperatures but only  $d\langle R_{\text{gyr},\perp} \rangle/dT$  shows some very small transition peak at the intersection of the freezing transition at  $T \approx 0.25$  and the adsorption transition line that will be discussed below. The AC2a pseudophase here seems to consist of a mixture of compact conformations docked to the surface and the mentioned semi-spherical conformations without any transition between them. This is also confirmed by regarding the low energy conformations found in this regime. At  $\epsilon_s \approx 1.4$  the continuous transition to double-layer structures (AC2b) takes place. The higher  $\epsilon_s$  the higher the amount of semi-spherical conformations found in the AC2a phase, but a clear cut from the compact adsorbed conformations does not exist.

This difference in the wetting transition for  $N = 13$  and  $N = 20$  might be due to the fact, that the most compact conformation for  $N = 13$  is an almost perfect icosahedron (Fig. 4.15). “Almost” because the Lennard-Jones energy minimum differs from the distance of neighboring monomers for  $s = 1$ . For  $s = 2$  the most compact structure is indeed a perfect icosahedron. This additionally stabilises the conformation and is already known in cluster physics, where 13 spheres can form one of Mackay’s icosahedrons [59] with their typical fivefold symmetry. Since also the globular structure of the 13mer has a higher symmetry than that of the 20mer, it is not unexpected that the wetting transition for the 13mer is sharper than for the 20mer. It might be worthwhile to study the wetting transition also for other chain length in order to be able to predict a trend for longer chains which is not possible only knowing the behaviour for those two investigated chain lengths.

Raising the temperature, polymers from the AC2a as well as from the AC2b regime form adsorbed and still rather compact configurations above the freezing

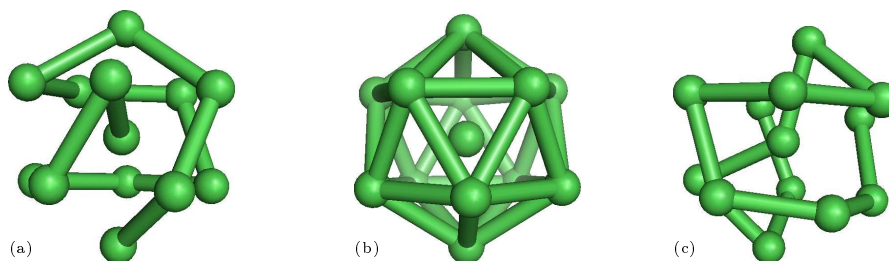


Figure 4.15: (a) Most compact conformation of the 13mer found for  $s = 2$ . (b) The same conformation as in (a), but all outer monomers are connected to show the icosahedral shape. (c) Most compact conformation of the 13mer found for  $s = 1$ .

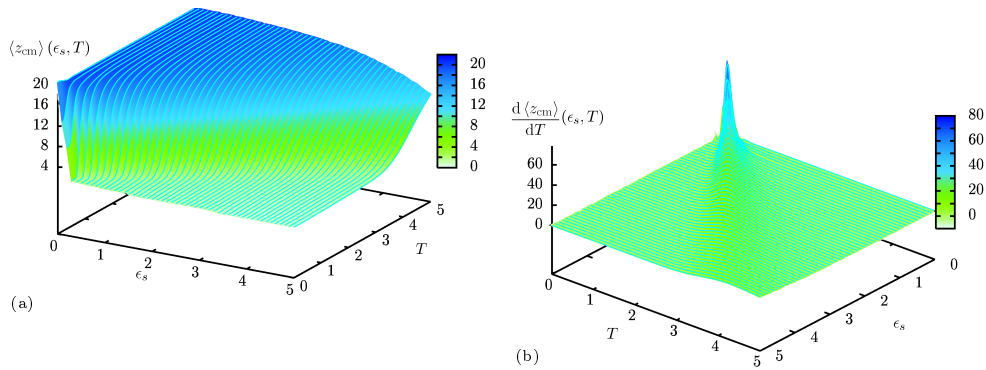


Figure 4.16: (a)  $\langle z_{\text{cm}} \rangle$  of the 20mer. (b)  $d\langle z_{\text{cm}} \rangle / dT$  of the 20mer. The small lines represent the simulated data. The colour code is interpolated.

temperature that are much like a drop on the surface. This pseudophase is called *Surface-Attached Globule* (AG) and sometimes also SAG. It has been first conjectured from short exact enumeration studies in 2D poor solvent [60], but was also found for instance in Ref. [15] and [16].

At even higher temperatures, two things can happen dependent on whether the monomer-monomer or the monomer-surface interaction is stronger. If the former is the case, the polymer first desorbs from the surface (from AG to DG) and expands at even higher temperatures (from DG to DE). In the latter case, the polymer expands while it is still on the surface (from AG to AE2) and desorbs at higher temperatures (from AE2 to DE). The point in the phase dia-

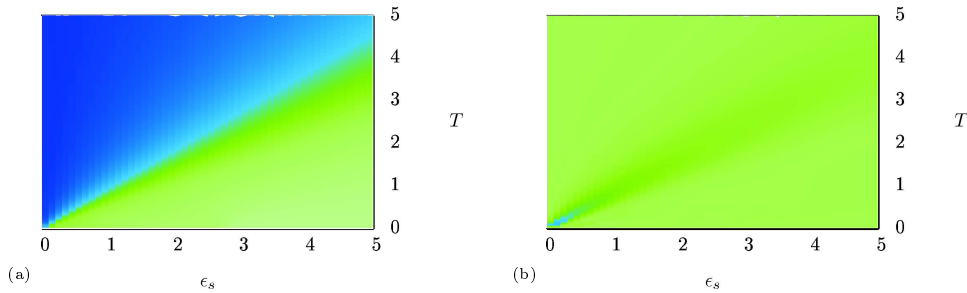


Figure 4.17: (a) Centre-of-mass distance  $\langle z_{\text{cm}} \rangle$  of the of 13mer to the surface. (b)  $d\langle z_{\text{cm}} \rangle / dT$  of the 13mer. The colour code is as in Fig. 4.16.

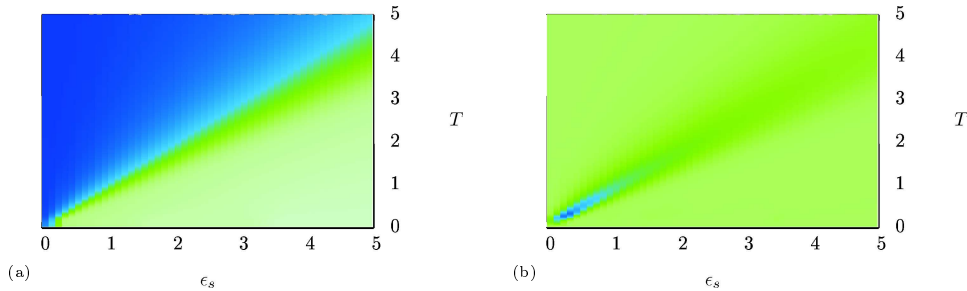


Figure 4.18: (a) Centre-of-mass distance  $\langle z_{\text{cm}} \rangle$  of the of 20mer to the surface. (b)  $d\langle z_{\text{cm}} \rangle / dT$  of the 20mer. The colour code is as in Fig. 4.16.

gram where the four phases AG, AE2, DG and DE coincide was found to be for  $N = 13$  at  $\epsilon_s \approx 1.57$  and  $T \approx 1.15$  and for  $N = 20$  at  $\epsilon_s \approx 1.9$  and  $T \approx 1.38$ . Due to the higher relative number of monomers in the bulk in compact longer chains, the  $\theta$ -temperature increases as explained above. Since at the adsorption transition there seems to be  $\epsilon_s \propto T$ , the four-phases-coexistence point is also shifted to higher  $\epsilon_s$  for longer chains.

The adsorption transition can be discussed best when looking at the mean number of surface contacts and the distance of the centre of mass of the polymer to the surface.

### Centre-of-Mass Distance of the Polymer to the Surface

The centre-of-mass distance of the polymer to the surface is the observable that displays the adsorption transition sharper than all the others, although one has to keep in mind the dependence on the size of the simulation box (section 2.4).

As can be seen in Figs. 4.16, 4.17 and 4.18, for large temperatures and low  $\epsilon_s$  the polymer can move freely within the simulation box without feeling the influence of the surface too much. Thus the average centre-of-mass distance  $\langle z_{\text{cm}} \rangle$  of the polymer above the surface is just half of the height of the simulation box. Since we chose the simulation box for  $N = 13$  to be  $L_{\text{box}} = 20$  and for  $N = 20$  to be  $L_{\text{box}} = 40$  this yields  $\langle z_{\text{cm}} \rangle = 10$  for  $N = 13$  and  $\langle z_{\text{cm}} \rangle = 20$  for  $N = 20$  if the attractive influence of the wall can be neglected. The steric influence is of course still there. On the other hand, for dominant surface attraction at high  $\epsilon_s$  and low temperatures, the polymer will preferably sit on the surface and the distance will be  $\langle z_{\text{cm}} \rangle \approx 0.858$ , corresponding to the minimum of the surface attraction potential, for single layer structures and a bit bigger for double-layer and globular structures.

When looking at  $\langle z_{\text{cm}} \rangle$  and  $d\langle z_{\text{cm}} \rangle/dT$  as presented in Figs. 4.16, 4.17 and 4.18, one notices first a quite sharp adsorption transition that divides the projection of  $\langle z_{\text{cm}} \rangle$  into an adsorbed (bright/green) regime and a desorbed (dark/blue) regime. This transition looks like a straight line in the phase diagram such that there seems to hold:  $\epsilon_s \propto T$  on the transition line. Intuitively, this makes sense since at higher  $T$  the stronger Brownian fluctuation is more likely to overcome the surface attraction. One can, however, also deduce it from the lattice model in Ref. [15] that we will check our results against at the end of the current chapter.

The model in question is a minimalistic simple-cubic (sc) excluded volume lattice model, where the polymer can move between two infinitely extended parallel walls, separated by a distance  $z_w$  expressed in lattice units. One wall is attractive to the monomers while the other one has a pure steric influence to prevent the polymer to escape into outer space. The energy term of such a polymer is given by

$$E(n_s, n_m) = -\epsilon_s n_s - \epsilon_m n_m, \quad (4.9)$$

where  $n_s$  is the number of nearest-neighbour monomer-substrate contacts,  $n_m$



the number of nearest-neighbour, but nonadjacent monomer-monomer contacts and  $\epsilon_s$  and  $\epsilon_m$  are the respective contact energy scales. The restricted partition sum for a macrostate with  $n_s$  surface contacts and  $n_m$  monomer-monomer contacts is then given by

$$Z_T(n_s, n_m) = \sum_{n'_s, n'_m} \delta_{n'_s, n_s} \delta_{n'_m, n_m} e^{-E(n'_s, n'_m)/k_B T} = g_{n_s, n_m} e^{-E(n_s, n_m)}, \quad (4.10)$$

where  $g_{n_s, n_m}$  is the contact density that only depends on the geometry of the system, in particular  $z_w$  and  $N$ . The partition function is hence  $Z = \sum_{n_s, n_m} Z_T(n_s, n_m)$  and more interestingly one can define the specific contact free energy as a function of the contact numbers  $n_s$  and  $n_m$ :

$$F_T(n_s, n_m) = -k_B T \ln \left( g_{n_s, n_m} e^{-E(n_s, n_m)/k_B T} \right) = E(n_s, n_m) - TS(n_s, n_m), \quad (4.11)$$

identifying  $k_B \ln g_{n_s, n_m} \equiv S(n_s, n_m)$  as a ‘‘microcontact’’ entropy. If one now minimises this with respect to  $n_s$ , one finds

$$\begin{aligned} \frac{dF(n_s, n_m)}{dn_s} &= \frac{d}{dn_s} (-k_B T \ln g_{n_s, n_m} + E(n_s, n_m)) \\ &= -k_B T \frac{d}{dn_s} \ln g_{n_s, n_m} - \epsilon_s = 0 \\ \Leftrightarrow \epsilon_s &= -k_B \left( \frac{d}{dn_s} \ln g_{n_s, n_m} \right) T \propto T \end{aligned} \quad (4.12)$$

because the factor in front to  $T$  does not depend on  $\epsilon_s$  and  $T$ . This argument should also be valid in our model since the surface energy is proportional to  $\epsilon_s$  and short ranged and the contribution of the other energy terms that do not depend on the number of surface contacts vanishes while differentiating. Our simulation yields  $\Delta T/\Delta\epsilon_s \approx 0.8806$  for  $N = 13$  and  $\Delta T/\Delta\epsilon_s \approx 0.9342$  for  $N = 20$  what should correspond to  $-1/(dg_{n_s, n_m}/dn_s)$  for the  $n_s$  at the adsorption transition that is naturally around one.

Looking at  $\langle z_{cm} \rangle$  and  $d\langle z_{cm} \rangle/dT$  in the adsorbed phases in more detail one observes the low temperature transitions between AC2a, AC2b and AC1 that were already discussed above. But since the data are so similar to what is found for the main number of surface contacts in that regime, we will skip a more detailed look here to go to the mean number of surface contacts immediately.

### Mean Number of Surface Contacts

The mean number of surface contacts provides a clear measure of the fraction of monomers adsorbed to the surface but gives no information on what happens

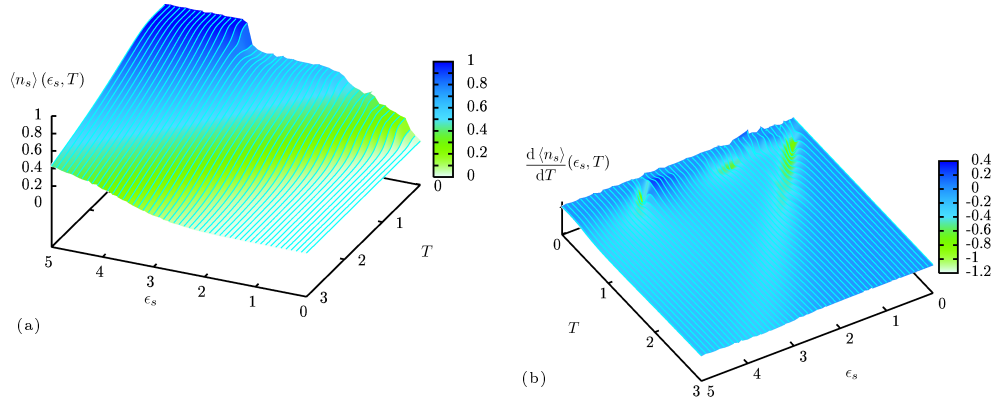


Figure 4.19: (a) Mean number  $\langle n_s \rangle$  of surface contacts per monomer of the 20mer. (b)  $d\langle n_s \rangle/dT$  of the 20mer. The small lines represent the simulated data, the colour code is interpolated.

in the desorbed states. Hence, it is a good observable to study the conformational behaviour in the adsorbed regime. For example, the regime, where single-layer structures dominate (AC1, AE1) can be identified easily, because  $\langle n_s \rangle \approx 1$  (dark/blue regime in Fig. 4.19, 4.20 and 4.21), which coincides nicely with the region, where  $\langle R_\perp \rangle \approx 0$ . One can also see that the average number of surface contacts are very similar in the phases AG, AC2a, AC2b, AE2, where most of the monomers are attached, with a monotonic decrease of  $\langle n_s \rangle$  with increasing  $T$ . The adsorption transition can be found at lower  $T$  as for  $\langle z_{cm} \rangle$  and is less sharp.

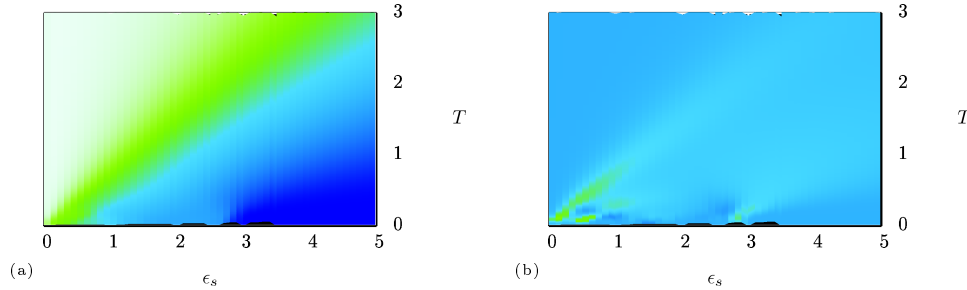


Figure 4.20: (a) Mean number  $\langle n_s \rangle$  of surface contacts per monomer of the 13mer. (b)  $d\langle n_s \rangle/dT$  of the 13mer. The colour code is as in Fig. 4.19.

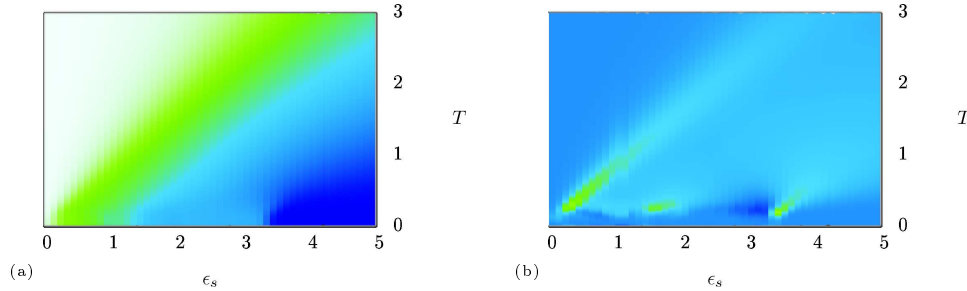


Figure 4.21: (a) Mean number  $\langle n_s \rangle$  of surface contacts per monomer of the 20mer. (b)  $d\langle n_s \rangle/dT$  of the 20mer. The colour code is as in Fig. 4.19.

At low temperatures  $d\langle n_s \rangle / dT$  indicates the following transitions:

transition	$N = 13$	$N = 20$
adsorption transition	$T \approx 0.2$	$T \approx 0.2$
transition between AC and AC2a	$T \approx 0.5$	–
transition between AC2a and AC2b	$T \approx 0.9$	$T \approx 1.7$
layering transition between AC2b and AC1	$T \approx 2.8$	$T \approx 3.4$

Unlike on the lattice, where one finds  $\langle n_s \rangle \approx 1$  for a single layer structure,  $\langle n_s \rangle \approx 1/2$  for a double layer structure,  $\langle n_s \rangle \approx 1/3$  for a triple layer structure etc. [16], our double layer structures have  $\langle n_s \rangle > 1/2$ . This indicates that while on the lattice, in order to obtain a compact configuration, all layers contain about the same amount of monomers, in our off-lattice model, there are always some more monomers in the layer on the surface. Since this only happens at the outer part of the layer, the difference is more pronounced the shorter the chain is.

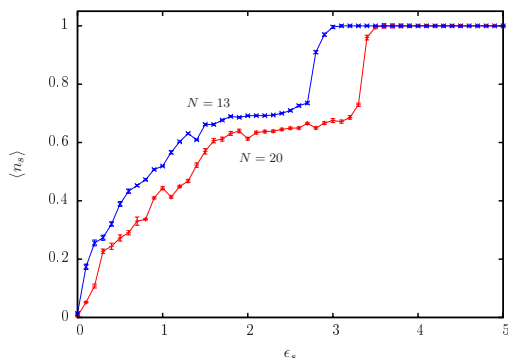


Figure 4.22:  $\langle n_s \rangle$  vs.  $\epsilon_s$  for small  $T$  for both investigated chain length  $N = 13$  and  $N = 20$ .

In Fig. 4.22,  $\langle n_s \rangle$  is shown as a function of  $\epsilon_s$  at small temperatures.  $\langle n_s \rangle$  is a good quantity to see layering. Starting at high  $\epsilon_s$ , first for both chain length  $\langle n_s \rangle \approx 1$  until at the layering transition,  $\langle n_s \rangle$  jumps to  $\langle n_s \rangle \approx 0.69$  for  $N = 13$  and to  $\langle n_s \rangle \approx 0.65$  for  $N = 20$ .

Further jumps corresponding to further layering transitions are not observed for those short chains. Instead what follows is a plateau regime where the relative amount of monomers that cover the surface is rather constant. When the double-layer structure gets unstable at lower  $\epsilon_s$ ,  $\langle n_s \rangle$  starts to decrease again. The conformations in AC and AC2a thus have no fixed  $\langle n_s \rangle$ , but it rather depends on  $\epsilon_s$ . At around  $\epsilon_s \approx 0.2$ , where the polymer desorbs,  $\langle n_s \rangle$  decrease rapidly until  $\langle n_s \rangle = 0$  at  $\epsilon_s = 0$ .

The observable left to discuss is the mean number of intrinsic contacts.

### Mean Number of Intrinsic Contacts

Since, if the mean number of intrinsic contacts  $\langle n_m \rangle$  is large, the polymer is very compact and its radius of gyration small, and if the mean number of intrinsic contacts is small, the polymer is very stretched out and the radius of gyration very high,  $\langle n_m \rangle$  and  $\langle R_{\text{gyr}} \rangle$  are complementary observables and yield similar informations.

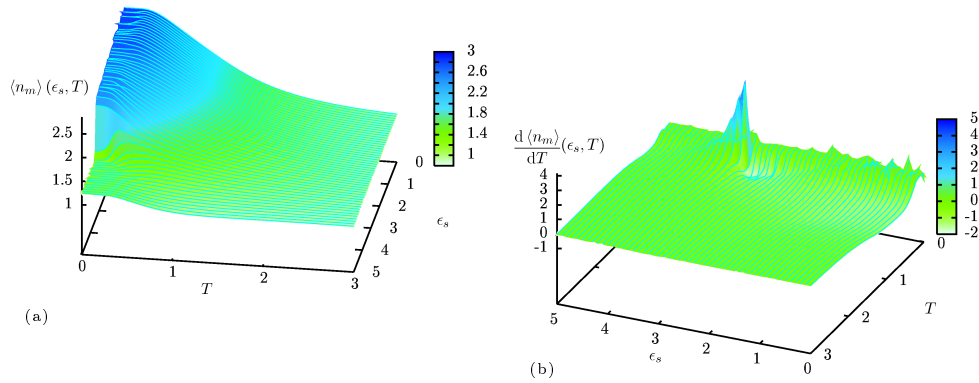


Figure 4.23: (a)  $\langle n_m \rangle$  of the 20mer, (b)  $d\langle n_m \rangle/dT$  of the 20mer. The small lines represent the simulated data. The colour code is interpolated.

One can see, that the projection of  $\langle n_m \rangle$  onto the  $\epsilon_s$ - $T$ -plane is divided into a compact regime comprising AC, AG, AC2a, AC2b, DC and DG and a regime of less compact conformations. This nicely confirms the results already obtained for  $\langle R_{\text{gyr}} \rangle$ . Apart from that, transitions from maximally compact conformations (DC, AC) to less compact ones (AC2a) and the  $\theta$ -transition of the free polymer are confirmed. And, once again, the layering transition from double- to single-layer configurations is strongly signalled.

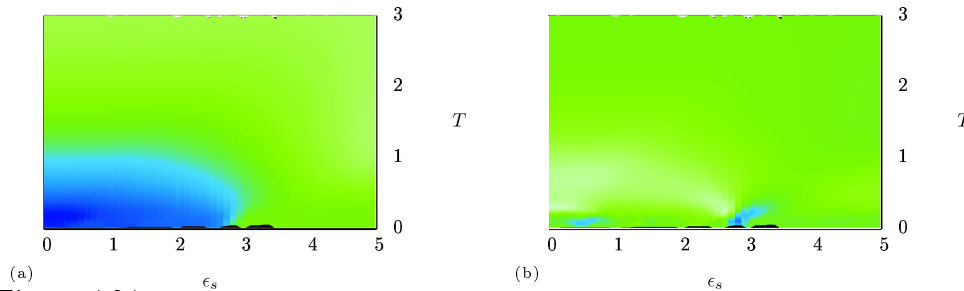


Figure 4.24: (a) Mean number  $\langle n_m \rangle$  of intrinsic contacts (without next neighbours along the chain) per monomer of the 13mer, (b)  $d\langle n_m \rangle/dT$  of the 13mer. The colour code is as in Fig. 4.23.

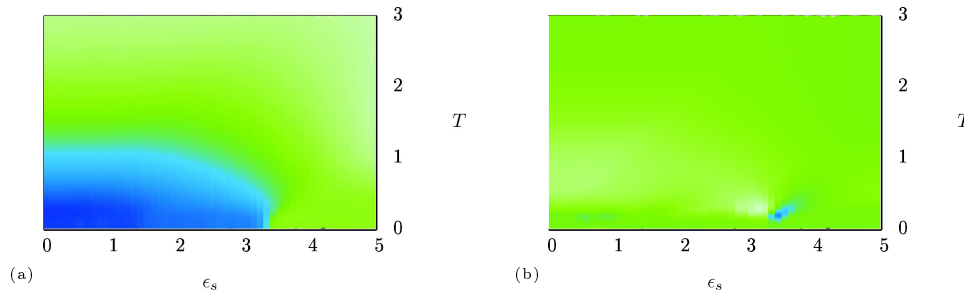


Figure 4.25: (a) Mean number  $\langle n_m \rangle$  of intrinsic contacts (without next neighbours along the chain) per monomer of the 20mer, (b)  $d\langle n_m \rangle/dT$  of the 20mer. The colour code is as in Fig. 4.23.

### 4.1.2 The Pseudophase Diagram

To summarise all the informations gained from the different observables, we drew the approximate boundaries of different regimes into the  $\epsilon_s$ - $T$ -plane and denoted the different pseudophases by the abbreviations already used (Fig. 4.26).

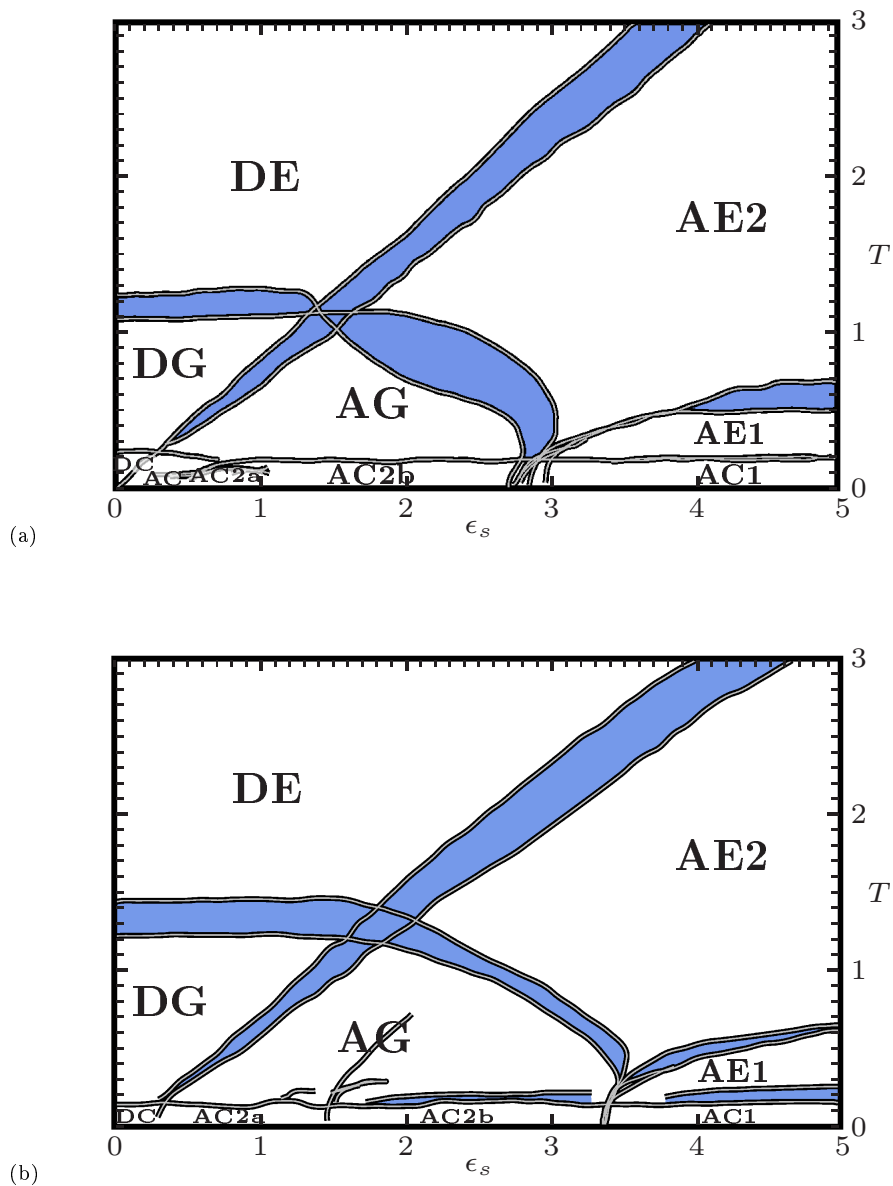


Figure 4.26: (a) Phase diagram of the 13mer, (b) Phase diagram of the 20mer. The coloured stripes indicates the regime where the phase transitions take place.

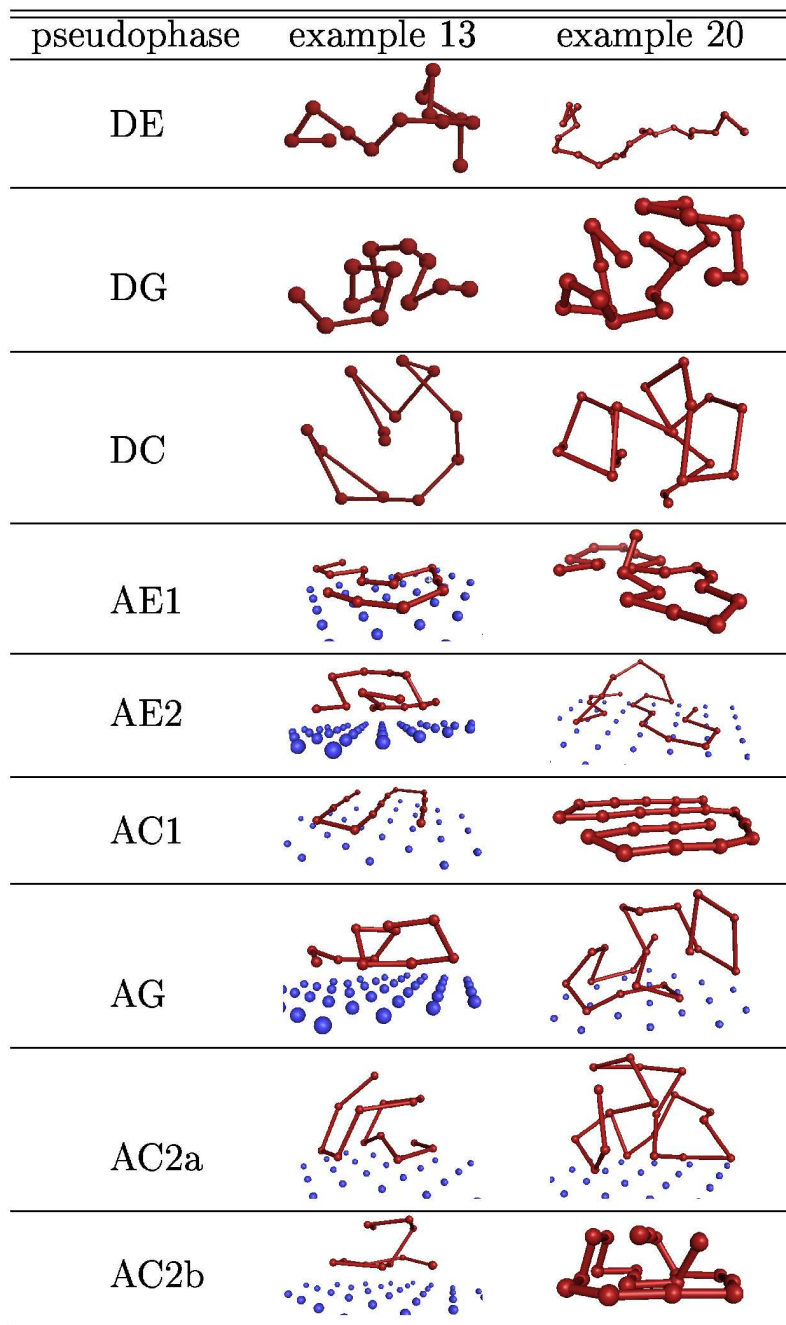


Figure 4.27: Representative examples of conformation of the different regions in the  $\epsilon_s$ - $T$ -plane. Conformations AC1a, AC1b, AC2a, AC2b and AG are located at the surface, the location of which is sometimes indicated with blue spheres. DE, DG and DC are desorbed.

The pseudophases found are (Fig. 4.27):

- **DE** (desorbed expanded): These are free desorbed random coil conformations above the  $\theta$ -transition.
- **DG** (desorbed globular): Globular free conformations below the  $\theta$ - and above the freezing-transition. Polymers in this pseudo-phase correspond to a liquid and are still rather flexible.
- **DC** (desorbed compact): Maximally compact, spherically shaped configurations below the freezing transition
- **AE1** (adsorbed expanded single layer): Expanded configurations above the freezing-transition that are flat on the surface but little compact.
- **AE2** (adsorbed expanded 3D conformations): Adsorbed expanded configurations above the freezing-transition with usually more than half of the monomers attached.
- **AC1** (adsorbed compact single layer): Adsorbed compact configurations below the freezing-transition that are flat on the substrate but compact like a circle.
- **AG** (adsorbed globular 3D conformations): Adsorbed conformations at the surface below the  $\theta$ -transition, above the freezing-transition and with extension into the 3rd dimension. Like a drop on the surface.
- **AC2a** (adsorbed compact 3D conformations): Compact configuration at the surface, that is semi-spherically shaped and below the freezing-transition.
- **AC2b** (adsorbed compact double layers): Adsorbed double layer configurations below the freezing-transition. The occupation of the layer at the surface is slightly higher than that of the other layer.
- **AC** (adsorbed spherical compact): Conformation as in DC, but the polymer often touches the surface in this regime.

#### 4.1.3 Data for $\epsilon_s = 2$ and $N = 20$

In order to give an impression of how the simulated data look like before merging them together in the shown 3D plots and to stress how careful one has to be in believing the sketched pseudophase transition positions, all observables for an arbitrary  $\epsilon_s$  ( $\epsilon_s = 2$ ) and for  $N = 20$  are presented here. A look on the phase diagram for  $N = 20$  tells us that we have to expect the freezing transition at  $T \approx 0.2$ , the expansion from AG to AE2 at  $T \approx 1.3$  and the desorption at  $T \approx 1.6$ . This seems less obvious when regarding Fig. 4.28. All temperature derivatives of  $\langle E \rangle$ ,  $\langle z_{\text{cm}} \rangle$ ,  $\langle R_{\text{gyr}} \rangle$ ,  $\langle R_{\text{gyr},\parallel} \rangle$ ,  $\langle R_{\text{gyr},\perp} \rangle$ ,  $\langle n_m \rangle$  and  $\langle n_s \rangle$  show some activity at the freezing transition around  $T \approx 0.2$ , but the exact

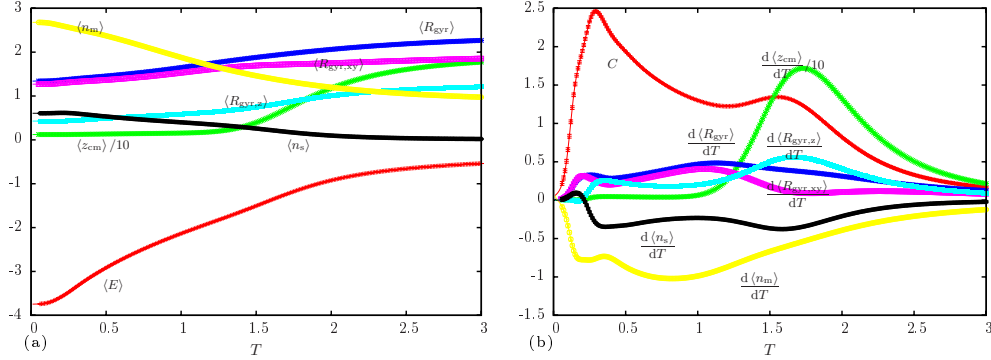


Figure 4.28: (a) Several observables for  $\epsilon_s = 2$  and  $N = 20$ , (b) The derivative with respect to  $T$  of the same observables.

position of the peak varies between  $T \approx 0.16$  for  $d\langle R_{\text{gyr},z} \rangle / dT$  and  $T \approx 0.35$  for  $d\langle z_{\text{cm}} \rangle / dT$ . This is clearly an effect of the finite size of the polymer. Only in the thermodynamic limit of very long chains the transitions are expected to take place at the same  $\epsilon_s$  and  $T$  for all observables. For finite chain length, the transition lines still change with  $N$  and are not well defined due to the broad peaks that are slightly different for different observables.

Below the freezing transition, all observables are quite constant with  $T$ . At such low temperatures, those macrostates are formed which are energetically favoured. Entropy is not yet relevant. Above the freezing transition however, entropy does play a role and the average energy increases, the conformation gets less compact – indicated by increasing  $\langle R_{\text{gyr}} \rangle$ ,  $\langle R_{\text{gyr},\parallel} \rangle$ ,  $\langle R_{\text{gyr},\perp} \rangle$  and decreasing  $\langle n_m \rangle$  – and the contact to the surface loosens – indicated by decreasing  $\langle n_s \rangle$  and increasing  $\langle z_{\text{cm}} \rangle$ . The expansion from AG to AE2 at  $T \approx 1.3$  is even less well defined. In fact, it is rather hard to locate this continuous transition. Not only expands the polymer monotonously with  $T$  and it is not clear, where to define the boundary between “compact” and “random coil” conformations. Also the peaks of the relevant observables  $d\langle R_{\text{gyr}} \rangle / dT$ ,  $d\langle R_{\text{gyr},\parallel} \rangle / dT$ ,  $d\langle R_{\text{gyr},\perp} \rangle / dT$  and  $d\langle n_m \rangle / dT$  are located between  $T \approx 0.81$  and  $T \approx 1.67$ . This makes it impossible to draw a clear transition line. The line in the pseudophase diagram is a compromise between all the observables and due to the need to indicate that in one regime compact conformations and in the other one expanded once dominate. A clear transition line should only exist in the thermodynamic limit just as for the  $\theta$ -transition of desorbed states. The desorption transition is easier to locate, but still the peaks of  $C$ ,  $d\langle z_{\text{cm}} \rangle / dT$  and  $d\langle n_s \rangle / dT$  that mainly indicate this transition, do not coincide.  $d\langle z_{\text{cm}} \rangle / dT$  already shows activity at  $T \approx 1.73$ , then  $\langle n_s \rangle$  at  $T \approx 1.6$  and  $C$  at  $T \approx 1.56$ . It makes however sense, that when the system is cooled, the polymer first has to move to the surface – activity in  $\langle z_{\text{cm}} \rangle$  at higher  $T$  – before monomer-surface contacts can have a considerable effect on the energy.

This hopefully made it clear, that the pseudophase diagrams in the previous subsection are supposed to give a good qualitative overview about the behaviour with varying  $\epsilon_s$  and  $T$ , but the positions of the phase boundaries should only be considered as rough guidelines.



#### 4.1.4 Density of States

It is rather illuminating to take a look at the density of states  $\Omega(E)$  that was estimated while generating the multicanonical weights. In Fig. 4.29  $\Omega(E)$  is plotted both logarithmically and linearly for three investigated chain lengths  $N = 13, 20, 50$ . For any fixed chain lengths  $\Omega(E)$  behaves very similarly for  $E \gtrsim -0.25$  for all  $\epsilon_s$ , while the behaviour of  $\Omega(E)$  changes for energies below this value. In the linear plot, hardly any differences apart from fluctuations can be seen, but the logarithmic plot reveals, that the number of low energy states with  $E < -0.25$  increases with increasing  $\epsilon_s$ . This is simply due to the fact, that configurations with a low degeneracy on the surface like single layer configurations are assigned a lower energy the lower the  $\epsilon_s$ . The fraction of total configurations in the simulation box, whose energy gets considerably influenced by the surface potential present, decreases with increasing simulation box size, which also has an influence on  $\Omega(E)$  and the canonical expectation values of the observables (see section 2.4).

It is also interesting to observe, how the overall shape of the density of states changes with increasing  $N$ . For small  $N$  there is a distinct maximum of  $\Omega(E)$  at  $E \approx -0.25$ , the energy with the most microscopic realizations. As  $N$  increases, the relative amount of high energy configurations increases until at some point,  $\Omega(E)$  is no longer a decreasing function of  $E$  for  $E > -0.25$ , and has no maximum any more. Configurations near the peak of  $\Omega(E)$  are random-coil configurations with a high degeneracy. Higher energy configurations are those that self-intersect which is punished by the high energy repulsive part of the Lennard-Jones energy. The longer the chain, the more likely it gets for the polymer to self-intersect which explains the higher amount of high-energy configurations for higher  $N$ .

The increasing amount of high energy configurations with increasing chain length has far less an effect on the canonical averages, especially at low temperatures, than the increasing number of low energy states with increasing  $\epsilon_s$ . Fig. 4.30, e.g., shows the probability  $p(E) = \Omega(E) \exp(-E/T)$  of configurations with energy  $E$  per monomer at temperatures close to the adsorption temperature for several  $\epsilon_s$ . Due to the exponential Boltzmann weight  $\exp(-E/T)$ , low energy configurations are getting more and more probable at lower temperatures. Eventually low energy configurations obtain a similar probability than configurations with  $E \gtrsim -0.25$  with a lower probability for energies in between.

Since this is clearly found in Fig. 4.30 for short chain lengths, and we know that lower energy configurations are attached to the surface for not too small  $\epsilon_s$ , while configurations with  $E \gtrsim -0.25$  are mainly free. Hence the adsorption transition appears to be first-order like for those short chains. But one can also see that with the disappearing peak of  $\Omega(E)$  also the double peak structure of  $p(E)$  gets less and less pronounced. Thus for longer chains, the transition might eventually become continuous.

For even lower temperatures, the adsorbed conformations will gain more and more weight compared to the free ones. Hence for temperatures lower than the

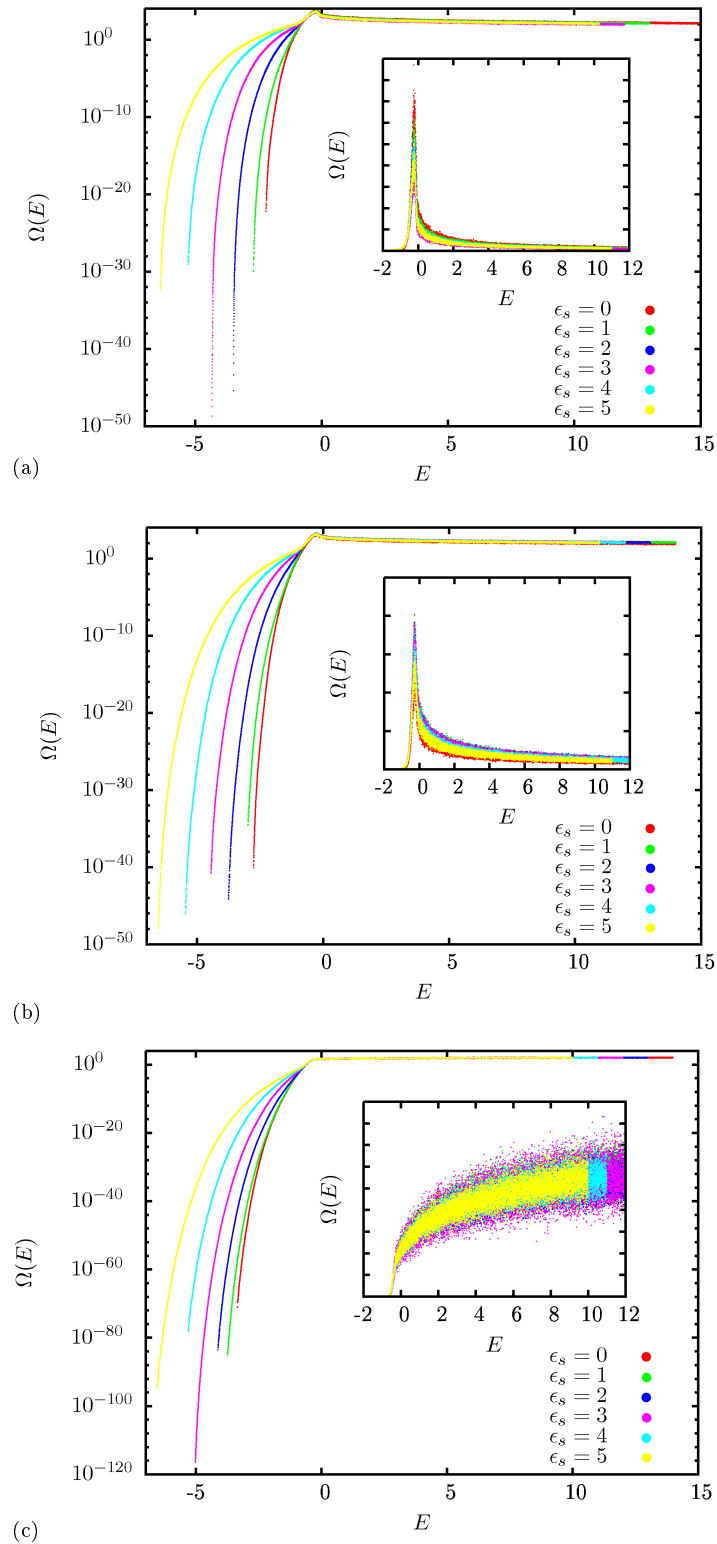


Figure 4.29: Density of states  $\Omega(E)$  for (a)  $N = 13$ , (b)  $N = 20$  and (c)  $N = 50$  and various  $\epsilon_s$ . The larger the  $\epsilon_s$  the lower the deepest possible energy of the polymer. Since only the relative  $\Omega(E)$  was approximated, the normalisation is some unknown number here. The inset gives the same data on a linear scale.

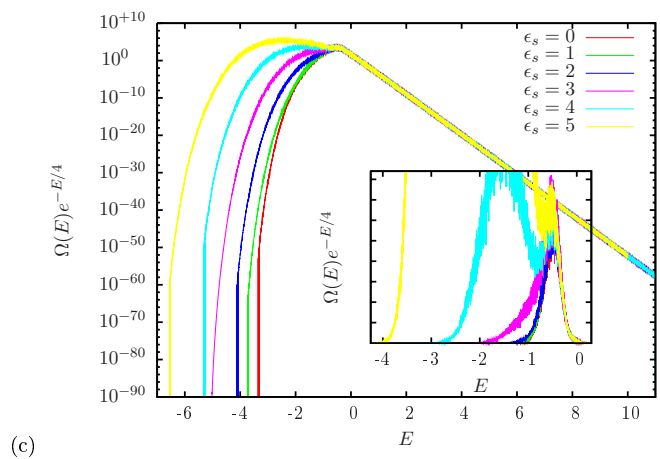
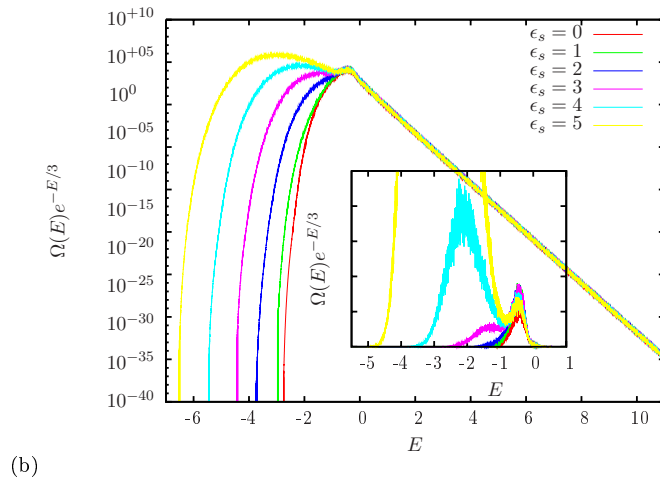
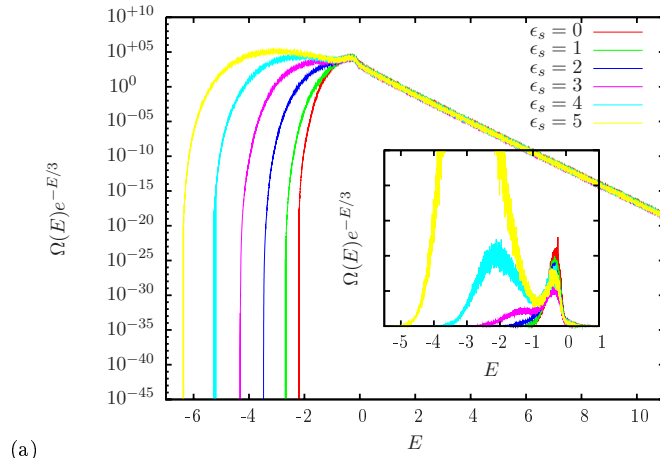


Figure 4.30: Probability  $p(E) = \Omega(E) \exp(-E_{\text{tot}}/T)$  of a polymer near a substrate to have an energy  $E$  in a canonical ensemble for different  $\epsilon_s$  for (a)  $N = 13$ , (b)  $N = 20$  and (c)  $N = 50$ .  $E$  on the abscissa is normalised per monomer, while  $E$  in the exponent is the absolute energy. The inset gives the same data on a linear scale.

plotted ones, only the rare low energy configurations on the substrate are of thermodynamic significance.

One more thing can be concluded from Fig. 4.30: The adsorption temperature increases with increasing  $N$ . This must be due to the decreasing relative amount of monomers on the surface that makes the free compact monomer more stable and the higher degeneracy of free conformations compared to desorbed one for longer chains.

#### 4.1.5 Comparison with Lattice Results

We would like to compare the results discussed with those obtained from a similar model on an sc lattice [15], that was already deployed to motivate the observation that  $\epsilon_s \propto T$  at the adsorption transition in section 4.1.1.

The polymer is modeled as a nongrafted self-avoiding polymer between two infinitely extended parallel planar walls, separated by a distance  $z_w$  expressed in lattice units. One wall is short-ranged attractive, while the other one has a purely steric influence to prevent the polymer from escaping. The energy of the system is given by

$$E(n_s, n_m) = -\epsilon_s^L n_s - \epsilon_m n_m, \quad (4.13)$$

where  $n_s$  is the number of nearest-neighbour monomer-substrate contacts,  $n_m$  the number of nearest-neighbour, but nonadjacent monomer-monomer contacts and  $\epsilon_s^L$  and  $\epsilon_m$  are the respective contact energy scales. We add the ‘‘L’’ to

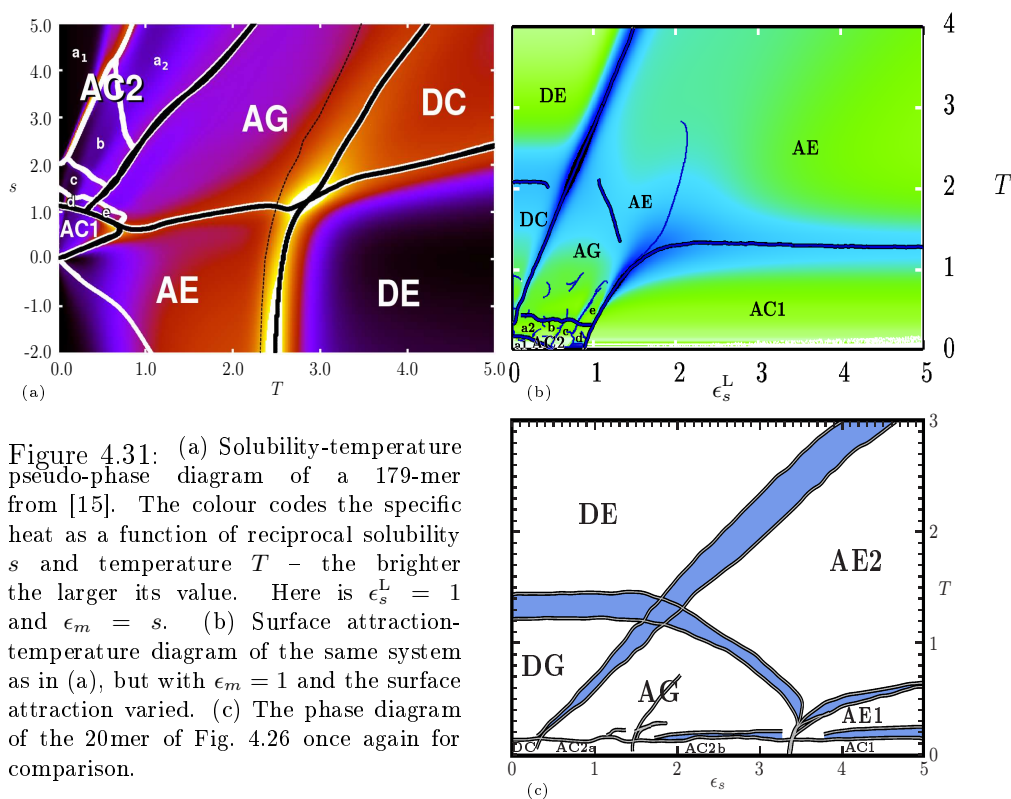


Figure 4.31: (a) Solubility-temperature pseudo-phase diagram of a 179-mer from [15]. The colour codes the specific heat as a function of reciprocal solubility  $s$  and temperature  $T$  – the brighter the larger its value. Here is  $\epsilon_s^L = 1$  and  $\epsilon_m = s$ . (b) Surface attraction-temperature diagram of the same system as in (a), but with  $\epsilon_m = 1$  and the surface attraction varied. (c) The phase diagram of the 20mer of Fig. 4.26 once again for comparison.

distinguish the equivalent surface contact energy scales on the lattice with those in our model. Thus, the probability for a conformation with  $n_s$  surface and  $n_m$  monomer-monomer contacts at temperature  $T$  and given  $\epsilon_s^L$  and  $\epsilon_m$  is

$$p_{T, \epsilon_s^L, \epsilon_m}(n_s, n_m) = \frac{1}{Z} g_{n_s, n_m} e^{(\epsilon_s^L n_s + \epsilon_m n_m)/k_B T}, \quad (4.14)$$

where  $Z$  is the partition sum  $Z = \sum_{n_s, n_m} g_{n_s, n_m} \exp((\epsilon_s^L n_s + \epsilon_m n_m)/k_B T)$ . This allows to calculate the canonical expectation values  $\langle O \rangle(T, \epsilon_s^L, \epsilon_m)$  for all  $T$ ,  $\epsilon_s^L$  and  $\epsilon_m$  for any function  $O(n_s, n_m)$ , if one knows  $g_{n_s, n_m}$  via:

$$\langle O \rangle(T, \epsilon_s^L, \epsilon_m) = \sum_{n_s, n_m} O(n_s, n_m) p_{T, \epsilon_s^L, \epsilon_m}(n_s, n_m). \quad (4.15)$$

The applied contact-density chain-growth algorithm in [15] which is an improved variant of the multicanonical chain-growth sampling method [61; 62] has the advantage that it directly samples the contact density  $g_{n_s, n_m}$  that only depends on the geometry of the system, in particular  $z_w$  and  $N$ . This allows to set the two independent energy scales  $\epsilon_s^L$  and  $\epsilon_m$  *after* the simulation to calculate the expectation values of interest.

In the paper, a simple rescaling was performed and a dimensionless reciprocal solubility was introduced by choosing

$$\epsilon_s^L = \epsilon_0, \quad \epsilon_m = s\epsilon_0 \quad \text{and} \quad \epsilon_0 = 1. \quad (4.16)$$

$s$  controls the quality of the implicit solvent surrounding the polymer, with a large  $s$  corresponding to a bad solvent and vice versa. Having an estimate of  $g_{n_s, n_m}$ , with eqs. (4.14), (4.15) and (4.16) the specific heat  $\langle C \rangle = (\langle E^2 \rangle - \langle E \rangle^2)/(k_B T^2)$ ,  $k_B = 1$ , was calculated for a range of  $s$  and  $T$ . The projection of  $\langle C \rangle(s, T)$  onto the solubility-temperature plane for a lattice homopolymer with 179 monomers in a cavity with  $z_w = 200$  is shown in Fig. 4.31(a). A legend to the various pseudophases can be found in Fig. 4.32.

Although the model is similar to our off-lattice model, with this parametrisation it is hard to directly compare the results, since one always has

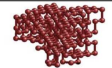
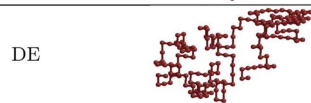
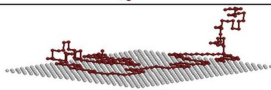
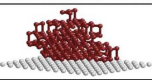




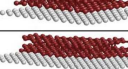
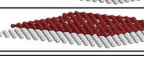
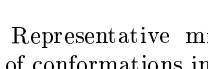
pseudophase	example	$n_s$	$n_m$
DC		0	219
DE		0	50
AE		135	33
AG		49	227
AC2a <sub>1</sub>		36	263
AC2a <sub>2</sub>		39	256
AC2b		46	257
AC2c		60	251
AC2d		90	231
AGe		103	207
AC1		179	153

Figure 4.32: Representative minimum free-energy examples of conformations in the different pseudo-phases of a 179mer [15].

to think, which  $s$ - $T$  combination in Fig. 4.31(a) corresponds to which  $\epsilon_s$ - $T$  combination in Fig. 4.31(c) – where we once again show the pseudophase diagram of the 20mer for convenience.

The rescaling of the lattice model that corresponds to the surface strength variation of our model is

$$\epsilon_s^L = \epsilon_s \epsilon_0 \quad \text{and} \quad \epsilon_m = \epsilon_0 \quad \text{and} \quad \epsilon_0 = 1. \quad (4.17)$$

Here  $\epsilon_s$  is allowed to adopt different values and is the quantity that corresponds to our  $\epsilon_s$  just as the  $s$  in eq. (4.16) roughly corresponds to our  $s$  – or more precisely to the square root of our  $s$ . The estimate for  $g_{n_s, n_m}$  can be reused to calculate  $\langle C \rangle(\epsilon_s, T)$  with this parametrisation over the same range of  $\epsilon_s$  and  $T$  than examined here. The result is depicted in Fig. 4.31(b).

In order to identify the pseudophases in Fig. 4.31(b) with the help of Fig. 4.31(a) a simple argument can be used: If the energy in the parametrisation (4.16) is denoted by  $E' = -n_s - sn_m$  and the temperature by  $T'$  and the energy in the parametrisation (4.17) by  $E = -\epsilon_s n_s - n_m$  and the temperature by  $T$ , in order to get the same Boltzmann weight in both parametrisations for a configuration with  $n_s$  surface and  $n_m$  monomer-monomer contacts, it has to hold:

$$\begin{aligned} \frac{E'}{T'} = \frac{E}{T} &\Leftrightarrow \frac{n_s + sn_m}{T'} = \frac{\epsilon_s n_s + n_m}{T} \\ \Leftrightarrow T = \frac{T'}{s} \quad \wedge \quad \epsilon_s = \frac{1}{s}. \end{aligned} \quad (4.18)$$

With these relations all pseudophases in Fig. 4.31(b) were identified. The transitions between them are located with a higher precision than in our results which is clearly due to the longer chain possible on the lattice.

It is quite reassuring to see, that there are certain similarities between Figs. 4.31(b) and (c). For instance, for the adsorption transition one finds in both models that  $\epsilon_s \propto T$ . The explanation for this noticeable transition was already given. Different however is not only the slope, derived in eq. (4.12) for the lattice model, that depends on the system's geometry and energy scales and we do not want to go into detail here. Also while for the off-lattice model, the extrapolation of the transition line seems to go through the origin  $\epsilon_s = 0$  and  $T = 0$ , there is an offset observed on the lattice such that the extrapolated transition line roughly crosses  $\epsilon_s = 0.4$  and  $T = 0$ . We speculate that this might be due to the intrinsic cubic structure of the lattice polymer that possesses flat surfaces at low temperatures even without a substrate. Unlike for off-lattice models, where a compact polymer attains a spherical shape, such a cubic conformation is likely to dock to a substrate without having to modify its conformation. Hence an important difference between lattice and off-lattice models is that while for not too high surface attraction strengths for off-lattice models, there is a competition between most compact spherical conformations that do not possess flat regions on the polymer surface and less compact conformations with flat regions that allow for more surface contacts but reduce the number of intrinsic contacts, such a competition is missing for sc lattice models.

This also explains, why a transition like the one observed for  $N = 13$  between AC and AC2a, the wetting transition, would never be possible on the lattice, since there simply is no spherically docked stable conformation AC. On the other hand, AC2 conformations at low  $T$  and for  $\epsilon_s$  between the adsorption and the single-double layering transitions can be observed in both models. And also for both models, the AG pseudophase was found when raising the temperature.

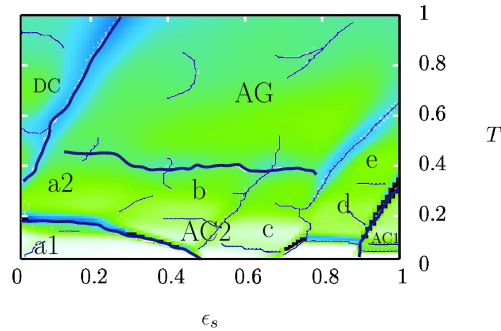


Figure 4.33: Low  $\epsilon_s$  and  $T$ -section of Fig. 4.31(b) to get a closer look at the AC2 subphases.

But while for the off-lattice model, apart from the wetting transition, there were only the transition from AC2a (semi-spherical shaped) into AC2b (double-layer structures) found, on the lattice a zoo of pseudo-transitions within AC2 can be seen (Fig. 4.33). It turns out, that these are the predicted higher-order layering transitions. With decreasing surface attraction, layer after layer is added until the number of layers is the same as in the most compact conformation. A lattice polymer has no other choice than to form those layers at low temperatures. The layering transition from AC1 to AC2 is very sharp for both models. Also the shape of the transition region from 2D adsorbed to 3D adsorbed conformations looks very similar. Interestingly, also the  $\epsilon_s/\epsilon_m$ -ratio predicted for this transition in Ref. [16] in eq. (4.6) agrees quite well with the observed one.

What could not be clearly identified on the lattice is a freezing transition that corresponds to the one between AC1 and AE1. There is some activity observed at very low  $T$  in AC1, but it is hard to draw precise conclusions from it.

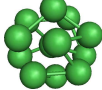
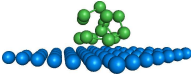
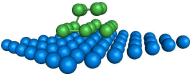
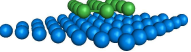
All the high-temperature pseudophases, DE, DC/DG, AG, AE, nicely correspond to each other in both models, which suggests that their overall positions are correctly identified.

#### 4.1.6 Low-Energy Configurations

Using the specialised algorithm Energy Landscape Paving (ELP) introduced in section 3.5, a search for low energy configurations at 100 different  $\epsilon_s$ ,  $\epsilon_s = 0, 0.05, 0.1, \dots, 5$ , was performed. Here, we did not only concentrate on the energetical favourable conformations that change with varying  $\epsilon_s$ , but also on the composition of the energy. It turned out that while ELP works very well for  $N = 13$ , the CPU time increased dramatically for  $N = 20$ , such that we only present rigorous results for  $N = 13$  here.

In Fig. 4.34, all energy contributions for  $N = 13$  are shown vs.  $\epsilon_s$ . One can nicely see that although the total energy  $E$  changes smoothly with  $\epsilon_s$ , this does not hold true for its subunits. Until  $\epsilon \approx 0.45$ , our ELP runs did not find

Table 4.1: Low energy configurations for the 13mer.

$\epsilon_s$ -range	0 – 0.45	0.5 – 1.15	1.2 – 2.8	2.85 – 5
configuration	DC/AC	AC2A	AC2b	AC1
				

low energy configurations that are docked to the surface, i.e.,  $E_{\text{sur}} \approx 0$  in this region and only  $E_{\text{LJ}}$  and  $E_{\text{bend}}$  contribute. This changes abruptly at  $\epsilon \approx 0.45$ , where  $E_{\text{sur}}$  increases to a finite value. At the same time the contribution due to the intrinsic Lennard-Jones energy decreases. This clearly indicates a wetting transition that is indeed observed when looking at the configurations displayed in Table 4.1.  $E_{\text{bend}}$  also shows some activity here and decreases. But for the overall behaviour the bending term hardly plays any role since it is much weaker than the competing surface and intrinsic attraction that determine the behaviour.

For  $\epsilon_s \gtrsim 0.5$  the surface contribution increases linearly with  $\epsilon_s$ , which is what one would expect.  $E_{\text{LJ}}$  stays rather constant until at  $\epsilon_s \approx 2.8$  it performs a jump down to the value attained in a compact single layer configuration. This is the double-single layer transition where  $E_{\text{LJ}}$  exactly coincides with  $E_{\text{sur}}$ . For even higher  $\epsilon_s$ ,  $E_{\text{LJ}}$  attains another plateau while  $E_{\text{sur}}$  again increases linearly but with a higher slope than in the double-layer region.

The transition from AC2a to AC2b is not visible in Fig. 4.34. This structural re-ordering from compact adsorbed semi-spherical conformation to the double-layer neither affects the number of intrinsic contacts nor the number of monomers docked to the surface considerably.

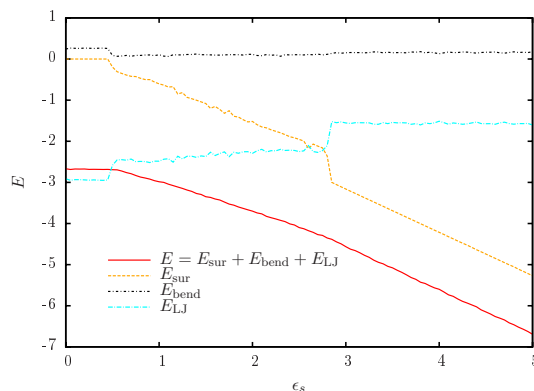


Figure 4.34: Different energy contributions of the 13mer normalised to the number of monomers.



## 4.2 Behaviour at Various Monomer-Monomer Attraction Strengths

Some additional work was performed on the thermodynamic behaviour of the polymer if the surface attraction strength is set constant ( $\epsilon_s = 1$ ), but the solvent quality  $s$  is varied. Only the case  $N = 13$  was considered. The hope is to obtain a pseudophase diagram like in Fig. 4.26(a) that can be compared with the one in Fig. 4.31(a). Here, we present first results of this still ongoing study.

In order to do so, the reweighting in  $s$  introduced in section 3.3 was applied. First, multicanonical weights for  $s = -2, -1.75, \dots, 5$  were generated. Afterwards, two parallel multicanonical simulations at neighboring  $s$  were performed for every pair of neighboring values of  $s$  as described in section 3.4 and the canonical expectation values for various  $T$  and  $s$  in between were calculated by reweighting the conformations. This approach allowed us to measure canonical expectation values over a wide range of  $s$  and having to generate relatively few multicanonical weights. In addition, every generated configuration can be used more often than in single multicanonical simulations where only a reweighting in  $T$  is performed.

We indeed obtained promising results. However, the longer reweighting procedure after every simulation step slowed the simulations down, such that no

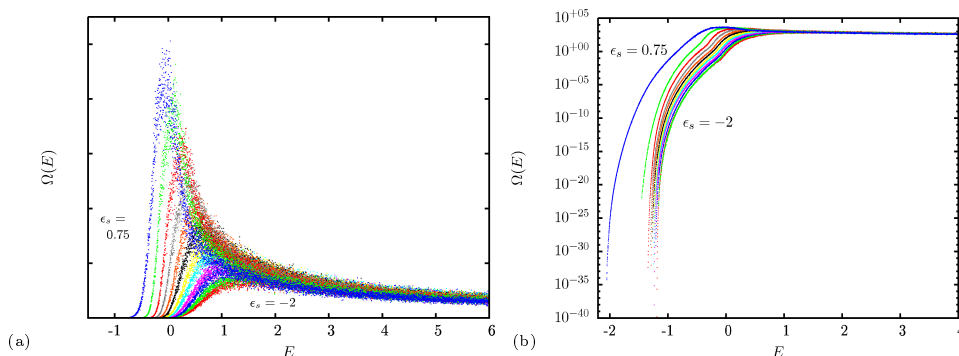


Figure 4.35:  $\Omega(E)$  for  $s = -2, -1.75, \dots, 0.75$  and  $\epsilon_s = 1$ . (a) Linear plot. (b) Logarithmic plot.

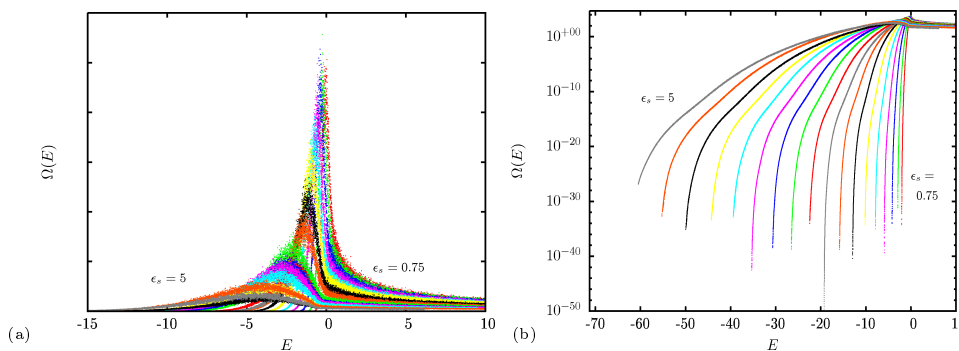
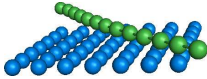
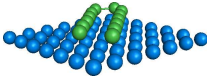
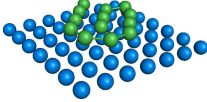
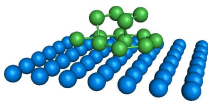
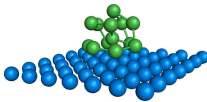
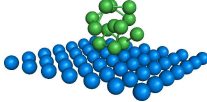
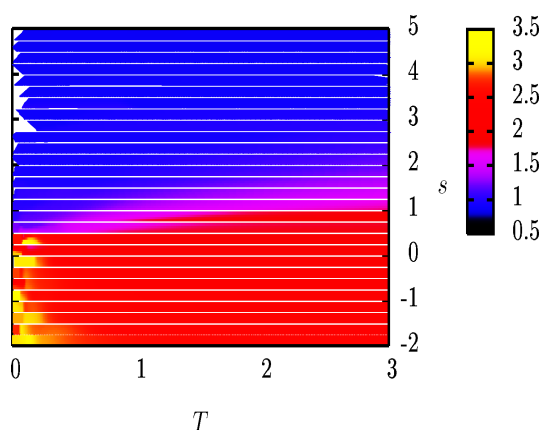


Figure 4.36:  $\Omega(E)$  for  $s = 0.75, 1, \dots, 5$  and  $\epsilon_s = 1$ . (a) Linear plot. (b) Logarithmic plot.

Table 4.2: Low energy configurations for the 13mer.

$s$	$-2 - 0$	0.25	0.5
configuration			
$s$	0.75	1	1.25 - 5
configuration			

final pseudophase diagram can be presented here. The statistics is not yet good enough. Thus, for such short chains, where the generation of new conformations is faster as for longer chains and the autocorrelation time is lower, one does not gain much with the reweighting in  $s$ . But it might become a useful technique for longer chains.

Figure 4.37:  $\langle R_{\text{gyr}} \rangle$  vs.  $s$  and  $T$  for the 13mer.

Some informations can however be extracted from the simulations performed. In Fig. 4.35 and Fig. 4.36 the density of states obtained from the multicanonical recursion are displayed. If one first looks at Fig. 4.35, one can see that for  $s = -2$  the density of states only attains a very weak maximum at  $E \approx 1.5$ . The lowest energy state found is at  $E \approx -1.188$ . If  $s$  increases until  $s = 0.25$  the maximum gets more pronounced and shifted to lower energies. The lowest energy found, however decreases only slightly. This changes for  $s = 0.5$ . Here the maximum increases further but the difference of the lowest energy found for  $s = 0.25$  and  $s = 0.5$  is considerably higher than it was for smaller  $s$ . For higher  $s$  the difference gets even more pronounced. To understand this, it helps to take a look at the low energy conformations found for different  $s$  that are presented in Table 4.2. For  $s \leq 0$ , the polymer prefers to be in a stretched conformation, like a rod. This is what one expects in this regime where individual monomers repel each other. From the canonical expectation values of, e.g., the radius of gyration, we know that the polymer stays rather stretched also for higher temperatures and there are indications for an adsorption transition somewhere between  $T = 0$  and  $T = 1$ . What is responsible for the much lower energies found for  $s = 0.25$  and  $s = 0.5$  is a collapse of the polymer. This is illustrated in Table 4.2, where low-energy con-

formations are displayed and around those values of  $s$ , one can see the polymer to collapse which corresponds to a wetting transition on the surface. As soon as the polymer is collapsed and monomer-monomer contacts are energetically favourable, an increasing  $s$  yields a much lower energy of compact conformations. As a consequence, the average energy goes down rapidly. This can also be seen in the density of states in Fig. 4.36. Here for increasing  $s$  the energy range of compact conformations gets much broader which also implies a reduction of the peak height.

Figure 4.37 shows the projection of the radius of gyration onto the  $s$ - $T$ -plane. Every colour stripe represents one run of two parallel multicanonical simulations. Although the fluctuations especially at low temperatures are still rather high, one can reliably read off that the  $s$ - $T$ -plane is split into a region of compact and a region of swollen conformations.

What we were not able to clearly identify is however the exact position of the adsorption transition. Especially for high  $s$ , the polymer only moves close to the surface without giving up its compact conformation. Hence an adsorption only means a small reduction in energy compared the contribution of the intrinsic contacts. On the other hand it reduces the entropy. Thus, an adsorption in this region is only expected to happen at very low temperatures. This is different on the lattice (see Fig. 4.31), where a docking to the substrate is much more favourable in bad solvents. For our model we found in several simulations containing about  $10^7 - 10^8$  sweeps no adsorption at all. In order to gain deeper insights into the adsorption behaviour of short polymers with this model, further and longer simulations are necessary. This is future work.



## 5 Summary & Outlook

The main objective of this work was to construct a pseudophase diagram of conformational thermodynamic phases of a single semi-flexible homopolymer near an attractive substrate in dependence of the external parameters surface attraction strength and temperature. The semiflexible homopolymer is modeled by a coarse-grained mesoscopic off-lattice model. In that model, a polymer is considered as a chain of monomers with covalent bonds of fixed length, a Lennard-Jones interaction potential for contacts between monomers being non-adjacent along the chain, a weak bending stiffness and a surface attraction term. The surface attraction is due to the van der Waals forces between the polymer and the substrate and is modeled by a Lennard-Jones-like potential.

The goal to construct a pseudophase diagram was indeed achieved for polymers of chain lengths  $N = 13$  and  $N = 20$ . For these chain lengths, the canonical expectation values of several observables were measured over a broad range of surface attraction strengths, and temperature using multicanonical simulations. Namely, the observables designed to unravel the equilibrium behaviour of such a system are the energy, the radius of gyration as well as its components parallel and perpendicular to the substrate, the distance of the centre-of-mass of the polymer to the surface, the mean number of monomer-surface contacts, the mean number of intrinsic monomer-monomer contacts and the thermal fluctuations of all those quantities.

The pseudophase diagram was constructed using all statistical informations about energetic and structural fluctuations. Although the computational expense to explore such a broad parameter range accurately restricted us to rather short chains, we were still able to identify several conformational pseudophases and pseudophase transitions. These are:

- Crystalline structures below the freezing transition. With increasing surface attraction strength we identified maximally compact conformations freely floating in solution (DC) or attached to the substrate (AC), semi-spherical compact conformations (AC2a) that are distorted by the surface but show no layers, double-layer conformation (AC2b) and single-layer conformations (AC1).
- Adsorbed conformations above the freezing transition. Here three conformational pseudophases were distinguished: globular, rather unstructured three-dimensional conformations (AG), expanded flat conformations on the substrate (AE1) and three-dimensional expanded adsorbed conformations (AE2).

- Desorbed conformations. Compact conformations (DC) are separated by the freezing transition from globular conformations (DG). At even higher temperatures above the well known  $\theta$ -transition random-coil conformations were found (DE).

The sharpest pseudophase transition identified is the layering transition between single- and double-layer-structures. The surface attraction strength at which this transition is observed coincides quite well with the one estimated for a hexagonal lattice – the lattice type adopted for single- and double-layer structures below the freezing transition. Higher-layer conformations were not found for these short chains. In a recent study on an sc lattice, for weak surface attraction and positive self-attraction, layering transitions were observed until a maximally compact cubic structure is reached. This is different for our off-lattice model. Here the maximally compact conformation is spherical and has no flat area on its surface that can easily form many monomer-substrate contacts without the need to deform. Hence the Lennard-Jones energy for monomer-monomer contacts competes with the monomer-substrate contact energy. This induces a wetting transition. For  $N = 13$  we clearly identified this wetting transition, while for  $N = 20$  the transition appears to be rather continuous. This difference might be due to the high symmetry of the most compact conformation of the 13mer that is an icosahedron. Hence it is not possible to predict the wetting behaviour of longer chains from just those two investigated chain lengths but more simulations at different chain lengths are needed.

At the adsorption transition the substrate adhesion strength was found to be proportional to temperature. This is expected since at a higher temperature the stronger Brownian fluctuation is more likely to overcome the surface attraction. But also a lattice argument that minimises the free energy is given and indeed the relation was also verified on the lattice.

For a single polymer in bulk solution, it is known that three phases exist, namely a swollen globule, a collapsed fluid globule and a solid crystalline state, that are separated by the  $\theta$ - and the freezing transition, respectively. We found those phases for low surface attraction and also identified their corresponding adsorbed pseudophases. So, the existence of an adsorbed globule (AG) was verified. The  $\theta$ -transition for adsorbed polymers between AG and AE2 shifts to lower temperatures compared to the free polymer until it disappears at the single-double layer transition. The freezing transition always stays at about the same low temperature. Only the exact structure of the crystalline state changes with  $\epsilon_s$ .

Due to the finite length of the investigated polymers, different observables indicate transitions at slightly different positions. This yields an uncertainty in the pseudophase transition regions indicated in the pseudophase diagrams, which should disappear in the thermodynamics limit. For  $N = 20$  it is not noticeably smaller than for  $N = 13$  which might be due to the higher possible symmetry of the 13mer, but this needs further investigation. From a short test of the dependence of the canonical expectations values on the size on the simulation box,

we also know that they always depend on the distance to the introduced steric wall. This dependence is however very weak for most of the observables if the size of box is chosen to be considerably larger than the polymer extension such that the conclusions drawn are independent on the exact box size chosen. Only the average distance of the polymer to the surface shows a stronger dependence. This however cannot be avoided since without it, the chance of the polymer to find the substrate and stay in its vicinity is negligible.

The density of states for  $N = 13$ ,  $N = 20$  and  $N = 50$  for different substrate adhesion strengths reveals a pronounced peak for  $N = 13$  and  $N = 20$ . The configurations that correspond to the peak are highly degenerate random-coil configurations. It vanishes for increasing chain length while the number of higher-energy states increases. Accordingly, the canonical probability  $p(E)$  at the adsorption transition exhibits a double-peaked structure for small  $N$  with a minimum between both peaks that seems to disappear for longer chains. Hence the transition seems to change from first-order like to continuous for increasing chain length.

Additional studies were performed for fixed positive surface attraction strength and varying solvent quality. The applied parallel multicanonical simulation with a reweighting in temperature and solvent quality works in principle, but the reweighting over a wide range of parameters after every sweep slowed the simulation considerably down such that the statistics is not yet good enough. However this approach might be very valuable if one only reweights to a small number of parameters at a time or if the chains are longer such that the generation of new configurations are more expensive compared to the reweighting. We were still able to draw some conclusions out of those simulations. So a wetting transition induced by increasing solvent quality and a collapse transition induced by decreasing solvent quality were identified and located. The question if the polymer adsorbs for very high intrinsic monomer-monomer attraction strengths and the determination of the adsorption temperature must be left open here for future investigations.





# Bibliography

- [1] F. Kuhner, M. Erdmann, and H. E. Gaub, *Scaling exponent and Kuhn length of pinned polymers by single molecule force spectroscopy*, Phys. Rev. Lett. **97**, 218301 (2006).
- [2] K. Goede, P. Busch, and M. Grundmann, *Binding specificity of a peptide on semiconductor surfaces*, Nano Lett. **4**, 2115 (2004).
- [3] K. Kegler, M. Salomo, and F. Kremer, *Forces of interaction between DNA-grafted colloids: An optical tweezer measurement*, Phys. Rev. Lett. **98**, 058304 (2007).
- [4] D. Napper, *Polymeric Stabilization of Colloidal Dispersions* (Academic Press Inc., London, 1983).
- [5] P. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca and London, 1979).
- [6] M. Doi, *Introduction to polymer physics* (Oxford University Press, New York, 1995).
- [7] H. Schneider, S. Granick, and S. Smith, *Kinetic traps in polymer adsorption. 1. Polystyrene displaced by polyisoprene at 12°C*, Macromolecules **27**, 4714 (1994).
- [8] E. Eisenriegler, K. Kremer, and K. Binder, *Adsorption of polymer chains at surfaces: Scaling and Monte Carlo analyses*, J. Chem. Phys. **77**, 6296 (1982).
- [9] Z. Usatenko, *Adsorption of long flexible polymer chains at planar surfaces: scaling analysis and critical exponents*, Journal of Statistical Mechanics: Theory and Experiment **2006**, P03009 (2006).
- [10] J. Forsman and C. E. Woodward, *Prewetting and layering in athermal polymer solutions*, Phys. Rev. Lett. **94**, 118301 (2005).
- [11] R. Rajesh, D. Dhar, D. Giri, S. Kumar, and Y. Singh, *Adsorption and collapse transitions in a linear polymer chain near an attractive wall*, Phys. Rev. E **65**, 056124 (2002).
- [12] E. Eisenriegler, *Polymers near surfaces: Conformation properties and relation to critical phenomena* (World Scientific, Singapore and New Jersey, 1993).

- 
- [13] S. Metzger, M. Müller, and K. Binder, *Surface excess in dilute polymer solutions and the adsorption transition versus wetting phenomena*, J. Chem. Phys. **118**, 8489 (2003).
- [14] P. Schravendijk, L. Ghiringhelli, L. Site, and N. vanderVegt, *Interaction of hydrated amino acids with metal surfaces: A multiscale modeling description*, J. Phys. Chem. C **111**, 2631 (2007).
- [15] M. Bachmann and W. Janke, *Substrate adhesion of a nongrafted flexible polymer in a cavity*, Phys. Rev. E **73**, 041802 (2006).
- [16] J. Krawczyk, A. L. Owczarek, T. Prellberg, and A. Rechnitzer, *Layering transitions for adsorbing polymers in poor solvents*, Europhys. Lett. **70**, 726 (2005).
- [17] N. Gupta and A. Irbäck, *Coupled folding-binding versus docking: A lattice model study*, J. Chem. Phys. **120**, 3983 (2004).
- [18] M. Bachmann and W. Janke, *Conformational transitions of nongrafted polymers near an absorbing substrate*, Phys. Rev. Lett. **95**, 058102 (2005).
- [19] T. Vrbová and S. G. Whittington, *Adsorption and collapse of self-avoiding walks in three dimensions: A Monte Carlo study*, J. Phys. A **31**, 3989 (1998).
- [20] R. Hegger and P. Grassberger, *Chain polymers near an adsorbing surface*, J. Phys. A **27**, 4069 (1994).
- [21] F. Celestini, T. Frisch, and X. Oyharcabal, *Stretching an adsorbed polymer globule*, Phys. Rev. E **70**, 012801 (2004).
- [22] Y. Singh, D. Giri, and S. Kumar, *Crossover of a polymer chain from bulk to surface states*, J. Phys. A **34**, L67 (2001).
- [23] N. Källrot and P. Linse, *Dynamic study of single-chain adsorption and desorption*, Macromolecules **40**, 4669 (2007).
- [24] F. Eisenmenger, U. Hansmann, S. Hayryan, and C.-K. Hu, *[SMMP] A modern package for protein simulations*, Comp. Phys. Comm. **138**, 192 (2001).
- [25] A. Irbäck and S. Mohanty, *PROFASI: a Monte Carlo simulation package for protein folding and aggregation*, J. Comp. Chem. **27**, 1548 (2006).
- [26] F. Müller-Plathe, *Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back*, ChemPhysChem **3**, 754 (2002).
- [27] M. Bachmann and W. Janke, *Thermodynamics of protein folding from coarse-grained models' perspectives* in: Lect. Notes Phys. **736**, ed. by W. Janke (Springer, Berlin, 2008), pp. 203-246.
- [28] R. Schiemann, M. Bachmann, and W. Janke, *Exact enumeration of three-dimensional lattice proteins*, Comp. Phys. Commun. **166**, 8 (2005).

- 
- [29] F. Stillinger, T. Head-Gordon, and C. Hirshfeld, *Toy model for protein folding*, Phys. Rev. E **48**, 1469 (1993).
- [30] F. Stillinger and T. Head-Gordon, *Collective aspects of protein folding illustrated by a toy model*, Phys. Rev. E **52**, 2872 (1995).
- [31] K. F. Lau and K. A. Dill, *A lattice statistical mechanics model of the conformational and sequence spaces of proteins*, Macromolecules **22**, 3986 (1989).
- [32] W. A. Steele, *The physical interaction of gases with crystalline solids*, Surf. Sc. **36**, 317 (1973).
- [33] S.-T. Sun, I. Nishio, G. Swislow, and T. Tanaka, *The coil-globule transition: Radius of gyration of polystyrene in cyclohexane*, J. Chem. Phys. **73**, 5971 (1980).
- [34] D. I. Svergun and M. H. J. Kocha, *Advances in structure analysis using small-angle scattering in solution*, Curr. Opin. Struct. Biol. **12**, 654 (2002).
- [35] P. Mishra and S. Kumar, *Effect of confinement on coil-globule transition*, J. Chem. Phys. **121**, 8642 (2004).
- [36] G. Marsaglia and A. Zaman, *Toward a universal random number generator*, Stat. & Prob. Lett. **9**, 35 (1990).
- [37] W. Janke, *Pseudo random numbers: Generation and quality checks*, invited lecture notes, in: Proceedings of the Euro Winter School Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms, edited by J. Grotendorst, D. Marx, and A. Muramatsu, John von Neumann Institute for Computing, Jülich, NIC Series, Vol. **10**, pp. 447-458 (2002).
- [38] D. Landau and K. Binder, *Monte Carlo simulations in statistical physics* (Cambridge University Press, Cambridge, 2000).
- [39] W. Krauth, *Statistical Mechanics: Algorithms and Computations* (Oxford University Press, New York, 2006).
- [40] B. Berg, *Markov Chain Monte Carlo Simulations and their Statistical Analysis* (World Scientific, Singapore, 2005).
- [41] M. Newman and G. Barkema, *Monte Carlo Methods in Statistical Physics* (Oxford University Press Inc., New York, 1998).
- [42] W. H. Press, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1993).
- [43] M. Bachmann, H. Arkin, and W. Janke, *Multicanonical study of coarse-grained off-lattice models for folding heteropolymers*, Phys. Rev. E **71**, 031906 (2005).
- [44] C. Junghans, *Aggregation of Mesoscopic Protein-Like Heteropolymers*, diploma thesis, University of Leipzig, 2006.

- [45] B. Berg and T. Neuhaus, *Multicanonical algorithms for first order phase transitions*, Phys. Lett. B **267**, 249 (1991).
- [46] W. Janke, *Multicanonical Monte Carlo simulations*, Physica A **254**, 164 (1998).
- [47] W. Janke, *Histograms and All That*, Computer Simulations of Surfaces and Interfaces, NATO Science Series, II. Mathematics, Physics and Chemistry - Vol. **114**, Proceedings of the NATO Advanced Study Institute, Albena, Bulgaria, 9 - 20 September 2002, edited by B. Dünweg, D.P. Landau, and A.I. Milchev (Kluwer, Dordrecht, 2003), pp. 137 - 157.
- [48] A. Ferrenberg and R. Swendsen, *Optimized Monte Carlo data analysis*, Phys. Rev. Lett. **63**, 1195 (1989).
- [49] E. Marinari and G. Parisi, *Simulated tempering: A new Monte Carlo scheme*, Europhys. Lett. **19**, 451 (1992).
- [50] K. Hukushima and K. Nemoto, *Exchange Monte Carlo method and application to spin glass simulations*, J. Phys. Soc. Japan **65**, 1604 (1996).
- [51] U. H. E. Hansmann, *Parallel tempering algorithm for conformational studies of biological molecules*, Chem. Phys. Lett. **281**, 140 (1997).
- [52] D. J. Earl and M. W. Deem, *Parallel tempering: Theory, applications, and new perspectives*, PhysChemChemPhys. **7**, 3910 (2005).
- [53] U. Hansmann and L. Wille, *Global optimization by energy landscape paving*, Phys. Rev. Lett. **88**, 4 (2002).
- [54] W. Janke, *Statistical Analysis of Simulation: Data Correlation and Error Estimation*, invited lecture notes, in: Proceedings of the Euro Winter School Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms, edited by J. Grotendorst, D. Marx, and A. Muramatsu, John von Neumann Institute for Computing, Jülich, NIC Series, Vol. **10**, pp. 423-445 (2002).
- [55] A. Kallias, *Thermodynamics and Folding Kinetics of Coarse-Grained Protein Models*, diploma thesis, University of Leipzig, 2005.
- [56] S. Schnabel, *Thermodynamische Eigenschaften und Faltungskanäle von coarse-grained Heteropolymeren*, diploma thesis, University of Leipzig, 2005.
- [57] J. Schluttig, *Molecular Mechanics of Coarse-Grained Protein Models*, diploma thesis, University of Leipzig, 2005.
- [58] A. Sokal, *Monte Carlo methods in statistical mechanics: Foundations and new algorithms* (Lecture Notes, Course de Trisieme Cycle de la Physique en Suisse Romande, Lausanne, 1989).

- 
- [59] A. L. Mackay, *A dense non-crystallographic packing of equal spheres*, Acta Cryst. **15**, 916 (1962).
- [60] Y. Singh, D. Giri, and S. Kumar, *Crossover of a polymer chain from bulk to surface states*, J. Phys. A **34**, L67 (2001).
- [61] M. Bachmann and W. Janke, *Multicanonical chain-growth algorithm*, Phys. Rev. Lett. **91**, 208105 (2003).
- [62] T. Prellberg and J. Krawczyk, *Flat histogram version of the pruned and enriched Rosenbluth method*, Phys. Rev. Lett. **92**, 120602 (2004).



# Danksagung

Eine Reihe von Menschen standen mir während meiner Diplomarbeitsphase hilfreich und motivierend zur Seite, bei denen ich mich nun bedanken möchte.

Zunächst danke ich meinen Betreuern Prof. Dr. Wolfhard Janke und Dr. Michael Bachmann. Dr. Michael Bachmann gelang es zusammen mit Prof. Dr. Klaus Kroy meine Begeisterung für die Statistische Physik und Polymerphysik zu wecken. Dank des interessanten Themas und der ausgezeichneten Betreuung in der Arbeitsgruppe von Prof. Dr. Wolfhard Janke hat sich diese keinesfalls abgekühlt und ich habe ein in vielfacher Hinsicht lehrreiches Jahr in einem angenehmen Umfeld hinter mir.

Zu dem angenehmen Umfeld trugen ganz erheblich Frank Beyer, Micha Wiedemann und Mathias Aust bei, mit denen ich viel Zeit im Raum 2L14 verbracht habe. Viele kleine technische Probleme wurden durch den sich entwickelnden regen Informationsaustausch schnell behoben und ich habe durch sie einiges Neues gelernt. Auch Thomas Vogel, Stefan Schnabel und Sandro Wenzel halfen mir bei Fragen gelegentlich weiter.

Die Simulationen liefen auf Rechnern des Instituts für Theoretische Physik der Universität Leipzig. Dr. Elmar Bittner und Andreas Nußbaumer verwalteten und pflegten das System während meiner Diplomarbeitszeit neben ihrer eigenen wissenschaftlichen Arbeit und kümmerten sich bei Wünschen oder Problemen sofort. Vielen Dank dafür.

Nicht zuletzt verdanke ich einer Reihe von Kommilitonen und Freunden Rat, Rückhalt und Freude während des Studiums.

Ein ganz besonderer Dank gilt natürlich meiner ganzen Familie, ohne deren moralische und finanzielle Unterstützung mein Studium so kaum möglich gewesen wäre.





## **Selbstständigkeitserklärung**

Hiermit erkläre ich, die vorliegende Arbeit selbstständig und ohne unerlaubte fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die in der Literatur angegebenen Quellen benutzt und sämtliche der Literatur entnommenen Textstellen und Bilder als solche kenntlich gemacht.

Ferner erkläre ich, dass die Arbeit noch nicht in einem anderen Studiengang als Prüfungsleistung verwendet wurde.

Leipzig, den 13.05.2008

## **Einverständniserklärung**

Hiermit erkläre ich mich einverstanden, dass meine Diplomarbeit nach positiver Begutachtung zur Benutzung in der Zweigstelle Physik der Universitätsbibliothek zur Verfügung gestellt wird.

Leipzig, den 13.05.2008