

# Theoretical Biophysics

## A Computational Approach

### Concepts, Models, Methods and Algorithms

#### RNA, Protein and DNA

---

Dieter W. Heermann

April 7, 2020

Heidelberg University

## 1. Introduction

- Some Measures
- Freely-Jointed Chain
- Freely-rotating Chain
- Gaussian Chain Model
- Worm-like Chain Model
- Self Avoiding Random Walk and Variants

## 2. Polymers at Membranes

## 3. Modelling of Biopolymers

- Intermolecular Interactions and Electrostatic Screening
- Lattice Polymer Models

## ■ United Atom Model

## ■ Proteins

- Protein Folding
- Folding as a Spin Glass Problem
- Lattice Protein Models

## ■ Conformations and Energy Landscapes

## ■ Helix-Coil Transition

## ■ DNA Melting

## ■ Chromatin

## ■ Chromosomes

## 4. Exercises

## 5. Bibliography

## 6. Index



# Introduction

---

There are four fundamental types of bio-macromolecules. Each type of macromolecule is a polymer composed of a different type of subunit

- Proteins which are composed of 20 amino acids
- Polysaccharides which are composed of monosaccharides
- Nucleic acids which are composed of 4 nucleotides
- Ribonucleic acids which are composed of ribonucleotides.

In passing we note that these macromolecules are polar, i.e. they have a head and a tail, because they are formed by head to tail condensation of polar monomers.

Many molecules essential to living systems, such as proteins and fats, are very large. They are polymers. These are very large molecules made up of smaller units, called monomers or repeating units, covalently bonded together. They are produced from a small set of about 50 monomers. In the biological setting, macromolecules are often created through a condensation or dehydration reaction, i.e. a loss of a water molecule or other small molecule as two monomers or molecules join.

Why should we study macromolecules? Because they provide structural integrity and shape in biological systems. Further the coupling of geometry and dynamics leads us to insights into the workings of biological systems such as ion pumps for example.

There are four macromolecules essential to living matter containing C, H, O, N and sometimes S

- Proteins
- Carbohydrates
- Nucleic Acids
- Lipids.

Bio-polymers consisting of regularly repeating units tend to form helices. Thus we are interested in the relationship between form and function and other physical properties of these macromolecules in this chapter.

DNA, RNA and Proteins can be modeled for computational purposes in a variety of ways [1]. Depending on the kind of question and the degree of abstraction, one has the basic choice between a model on a lattice or in continuous space. The bond fluctuation model [2] is one of the prominent representatives of a polymer model on the lattice. The main advantage of this type of models is the computational efficiency due to the restricted configuration space. With increasing computer power it was possible to stay closer to reality by simulating polymers by continuum models. Two widely used models of this class are the bead-spring [3] and the united-atom model [4].

The first measure describing a polymer is the contour length. Let  $N$  be the number of repeating units (monomers).  $N$  is the *degree of polymerization* or *chain length*. Each monomer unit has length  $b$ . Then the total *contour length* of the chain is  $L = Nb$ .

The *conformation* describes the geometric structure of a polymer. If two atoms are joined by a single bond then a rotation about that bond is possible. If the two atoms have other atoms or groups attached to them then configurations which vary in torsional angle are possible. This is shown in Figure 1. Here we have introduced the polar angle as the *bond angle*, i.e. the angle between two adjacent bonds.

$$H_{\text{bond angle}} = \frac{k_{\theta}}{2} \sum_{\text{angles}} (\cos \theta_{\text{angle}} - \cos \theta_0)^2 \quad . \quad (1)$$

Vibrations corresponding to bond-angle bending have frequencies of the order of  $10^{13}$  sec. Non-vibrational internal motions are geometrically distinguishable at time scales of around  $10^{11}$  sec [5].

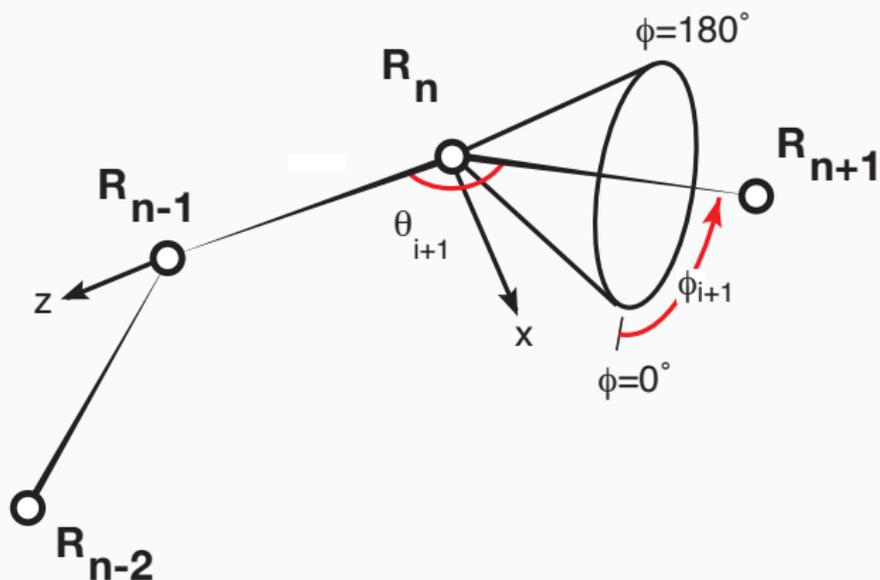


Figure 1: Angle definition

Since different conformations represent varying distances between the atoms or groups rotating about the bond, these distances determine the amount and type of interaction between adjacent atoms or groups. Thus different conformations represent different potential energies. There are several possible generalized conformations: Anti (t, Trans), Eclipsed (Cis), and Gauche (g, + or -). In Figure 2 is shown the possible potential energy with the corresponding labeling (the angle and labelling is also listed in table 1).

$$H_{\text{torsion}} = \sum_{\text{dihedral angle}} \left[ \frac{k_1}{2}(1 - \cos \phi) + \frac{k_2}{2}(1 - \cos 2\phi) + \frac{k_3}{2}(1 - \cos 3\phi) \right] \quad . \quad (2)$$

**Table 1:** Name convention for specific angles and their property

Name of conformation	Torsion angle	Symbol	Stability
Cis	$\pm 180^\circ$	c	unstable
Gauche	$\pm 120^\circ$	$g^+, g^-$	stable
Anti	$\pm 60^\circ$	$a^+, a^-$	unstable
Trans	$0^\circ$	t	stable

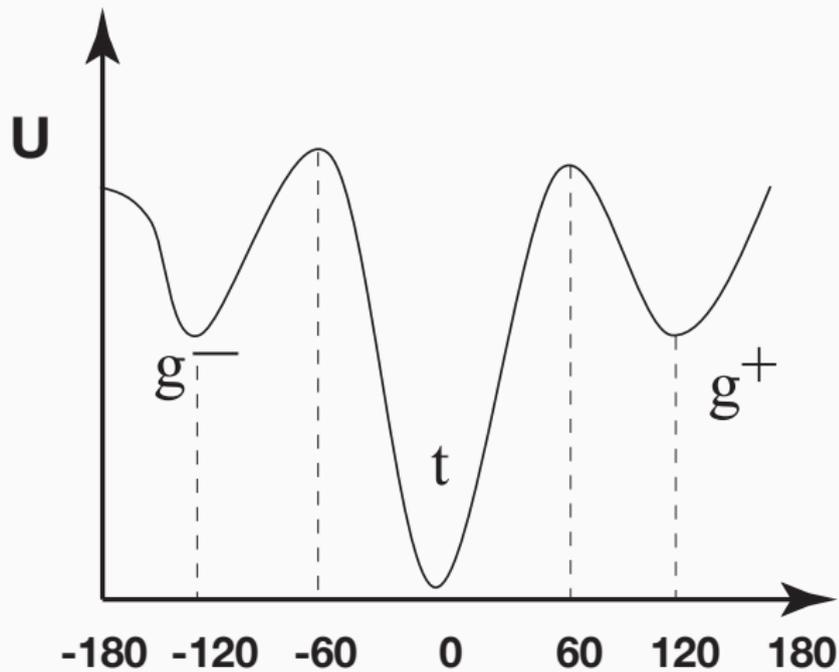


Figure 2: A typical torsion potential

Two special conformations arise if we have pairs of angles:

- $tt$  results in a zig-zag chain
- $g^-g^-$  or  $g^+g^+$  results in a helix.

Polymers are not rigid but can be easily twisted along the bonds of the backbone. This gives rise, at finite temperatures, to different conformations of the polymer.

The simplest model of polymer conformation treats the molecule as a chain of rigid subunits, joined by perfectly flexible hinges [6]. In this *freely jointed chain model* the chain is made up of  $N$  links, each of length  $b$  and  $N + 1$  beads or monomers with no excluded volume (see Figure 3). Thus it corresponds to a random walk where each step has length  $b$  (see Figure 10). This model is the most simple one for a single polymer in solution but is not appropriate to double stranded DNA. This is because individual covalent bonds do not have bending energies that are not small relative to  $k_B T$ . This, however, only applies if we want to describe the macromolecule on the atomistic level. Often, we want to describe the macromolecule on a length scale, where we can safely regard the polymer as flexible.

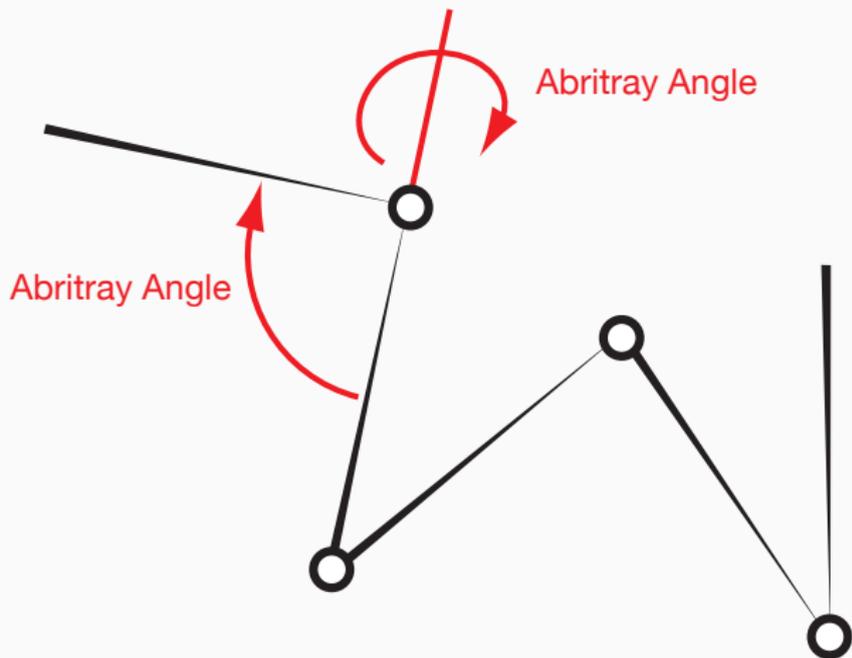


Figure 3: Freely-jointed chain

The joints of the chain are at positions  $\mathbf{R}_n$  and are joined by the link vectors, also called *bonds*

$$\mathbf{r}_n = \mathbf{R}_n - \mathbf{R}_{n-1} \quad . \quad (3)$$

The *end-to-end distance* for a given conformation is given by

$$\mathbf{R}_e = \mathbf{R}_0 - \mathbf{R}_N = \sum_{n=1}^N \mathbf{r}_n \quad (4)$$

which we assume to be a random variable.

Because the  $\mathbf{r}_n$  are uncorrelated we must have

$$\langle \mathbf{r}_n \rangle = 0 \quad (5)$$

after averaging over all possible conformations and

$$\langle \mathbf{r}_n \mathbf{r}_m \rangle = \delta_{nm} b^2 \quad . \quad (6)$$

Here the averaging is done over all possible orientations each having the same weight.

From these two equations we find that the average end-to-end vector is

$$\langle \mathbf{R}_e \rangle = \left\langle \sum_{n=1}^N \mathbf{r}_n \right\rangle = \sum_{n=1}^N \langle \mathbf{r}_n \rangle = 0 \quad (7)$$

and

$$\langle \mathbf{R}_e \mathbf{R}_e \rangle = \left\langle \sum_{n=1}^N \sum_{m=1}^N \mathbf{r}_n \mathbf{r}_m \right\rangle \quad (8)$$

$$= \sum_{n=1}^N \sum_{m=1}^N \langle \mathbf{r}_n \mathbf{r}_m \rangle \quad (9)$$

$$= \sum_{i=1}^N b^2 \quad (10)$$

$$= N b^2 \quad (11)$$

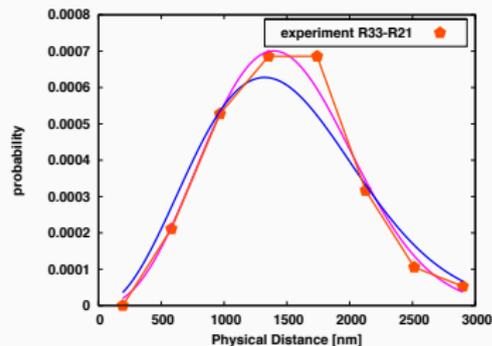
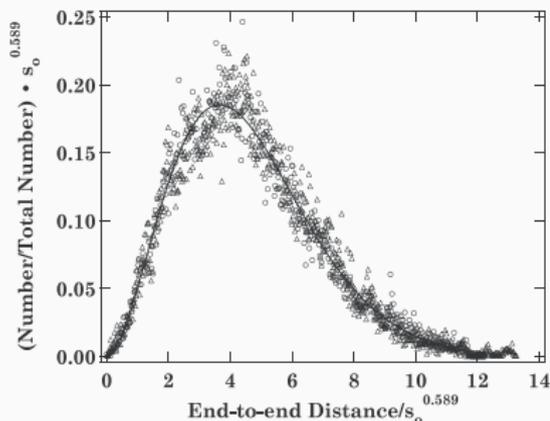
The end-to-end distance square *scales* with the length of the polymer

$$\langle R_e^2 \rangle \propto N \quad (12)$$

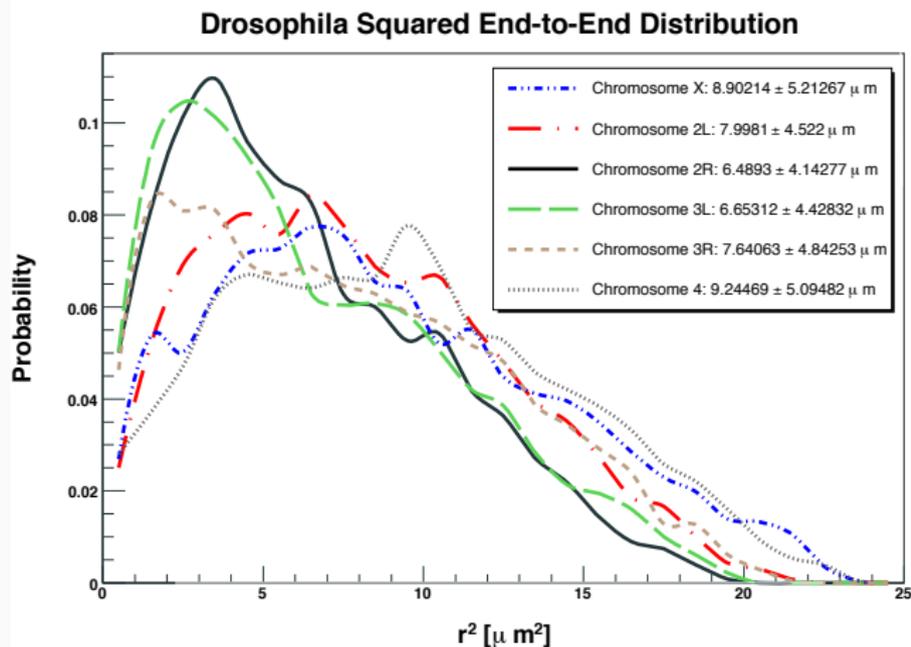
and it measures the average size of the polymer. Since we did not take into account excluded volume effects we anticipate a more general result for the end-to-end distance, if we take these into account, and write

$$\langle R_e^2 \rangle \propto N^{2\nu} \quad (13)$$

introducing the exponent  $\nu$ . Thus  $\nu = 1/2$  here.



**Figure 4:** End-to-end distance distribution for two biological macromolecules. The lower figure shows the comparison of data on human chromosomes (taken from van Driel et. al 2007). The blue curve shows the result for the random walk and the purple the result for a self-avoiding walk fitted to the data points.



**Figure 5:** Histogram representing the distribution of the end-to-end distance for two different contour lengths (548nm circles, 748nm triangles) and how they collapse onto each other. Taken from Dietler et. al. PRL 95, 158105 (2005)

DNA is much stiffer than an alkane chain. Hence DNA has a much larger  $\langle R_e^2 \rangle$  for a given contour length  $Nb$  than does an alkane. Thus we need to parameterize the stiffness of the chain. One such parameterization is the *Kuhn length*  $l_K$

$$\langle R_e^2 \rangle = N_K l_K^2 \quad (14)$$

$$L_c = N_K l_k \quad , \quad (15)$$

where we have introduced two parameters  $N_K < N$ , the *effective number of repeat units* and the Kuhn length  $l_K$ . The Kuhn length thus gives a measure for the statistical segment length.

A conceptually other measure is the *persistence length*  $\xi_p$ . It measures the length along the chain over which the tangent vectors of the chain become de-correlated. It is very useful in describing elastic properties of semiflexible polymers and deals with the rotational-isomeric-states, stiffness, helicity as well as the fact that a real chain can never fold back onto itself.

The persistence length for ideal chains is half of the Kuhn length.

$$\xi_p = l_k/2 \quad L_c \gg l_k \quad (16)$$

and hence

$$\langle R_e^2 \rangle = 2N_p \xi_p^2 \quad L_c \gg l_k \quad (17)$$

$$L_c = N_p \xi_p \quad (18)$$

where  $N_p$  is the contour length of the chain expressed in units of the persistence length. For B-DNA one finds a statistical segment length of 100 – 200 bp and a persistence length of approximately  $\xi_p \approx 50 \text{ nm}$ . Indeed biopolymers differ from artificial polymers in that they are stiff on length scales relevant for the biophysical processes they are involved in.

The probability distribution of the end-to-end vector is a Gaussian in the limit  $N \rightarrow \infty$  (central limit theorem) since it is the sum of independent random variables and we must have

$$P_N(\mathbf{R}_e) = \left( \frac{3}{2\pi N b^2} \right)^{-3/2} \exp \left( -\frac{3\mathbf{R}_e^2}{2N b^2} \right) \quad (19)$$

which is properly normalized.

Note that the conformations of the polymer are random coils. A typical conformation of the chain is shown in Figure 6.

We have introduced two length scales measuring the extend of the chain:  $N$  and  $R_e$ . Another measure is the *radius of gyration*  $R_g$

$$R_g = \sqrt{\frac{1}{N+1} \sum_{n=0}^N \langle (\mathbf{R}_{cm} - \mathbf{R}_n)^2 \rangle} \quad (20)$$

with the center of mass

$$\mathbf{R}_{cm} = \frac{1}{N+1} \sum_{n=0}^N \mathbf{R}_n \quad . \quad (21)$$

For the freely jointed chain model we obtain

$$R_g = \left( \frac{N}{6} \right)^{1/2} b \quad . \quad (22)$$

Thus the ratio between the end-to-end distance and the radius of gyration is constant

$$\frac{\langle R_e^2 \rangle}{\langle R_g^2 \rangle} = 6 \quad . \quad (23)$$



**Figure 6:** A typical random coil conformation

We now look at the free energy of the chain assuming no interaction. Let  $W$  be the number of accessible microstates of the chain. Then  $S(\mathbf{R}_e) = k_B \ln W$  is the entropy associated with a chain with an end-to-end vector  $\mathbf{R}_e$ . Since the system is athermal we need to consider the micro-canonical ensemble for the calculation of the entropy. The entropy difference between a chain held with end-to-end distance  $\mathbf{R}_e$  and one with the end-to-end vector of zero is

$$\Delta S(\mathbf{R}_e) = k_B \log \frac{P(\mathbf{R}_e)}{P(0)} \quad (24)$$

from which we obtain the free energy difference

$$\Delta F_e = -T\Delta S = \frac{3}{2} \frac{k_B T}{Nb^2} R^2 \quad . \quad (25)$$

A step further can be taken by fixing the bond angle  $\theta$  and allowing the torsion angle  $\phi$  to rotate freely. To calculate the end-to-end distance we need to consider the term

$$\langle \mathbf{r}_n \mathbf{r}_m \rangle \quad (26)$$

Since the torsion angle is free only the component that is projected due to the fixed angle contributes so that we have

$$\langle \mathbf{r}_n \mathbf{r}_m \rangle = b^2 (\cos \theta)^{|m-n|} \quad (27)$$

and with this

$$\langle \mathbf{R}_e^2 \rangle = \sum_{n=1}^N \sum_{m=1}^N \langle \mathbf{r}_n \mathbf{r}_m \rangle \quad (28)$$

$$= b^2 \sum_{n=1}^N \sum_{m=1}^N (\cos \theta)^{|m-n|} \quad (29)$$

$$= Nb^2 \frac{1 + \cos \theta}{1 - \cos \theta} \quad (30)$$

Thus the scaling behaviour is the same as for the freely-jointed chain only the Kuhn-length has changed.

We can generalize the above result assuming a finite correlation

$$\lim_{|m-n| \rightarrow \infty} \langle \cos \theta_{nm} \rangle = 0 \quad (31)$$

With this assumption we have

$$\sum_{m=1}^N \langle \cos \theta_{nm} \rangle = C_n \quad (32)$$

and thus

$$\langle \mathbf{R}_e^2 \rangle = b^2 \sum_{n=1}^N \sum_{m=1}^N \langle \cos \theta_{nm} \rangle \quad (33)$$

$$= b^2 N \sum_{n=1}^N C_n \quad (34)$$

$$= Nb^2 C_\infty \quad (35)$$

where  $C_\infty$  is called the *Flory characteristic ratio*.

To make the connection with the persistence length we note that

$$(\cos \theta)^{|m-n|} = \exp \{ |m-n| \ln(\cos \theta) \} = \exp -\frac{|m-n|}{\xi} \quad (36)$$

with

$$\xi = -\frac{1}{\ln(\cos \theta)} \quad (37)$$

we find the persistence length

$$\xi_p = b\xi \quad (38)$$

We consider a chain made up of orientationally uncorrelated (freely-jointed) links where the length of any link vector is no longer constant but has a probability distribution

$$G(\mathbf{r}) = \left(\frac{3}{2\pi b^2}\right)^{3/2} \exp\left(-\frac{3\mathbf{r}^2}{2b^2}\right) \quad (39)$$

with the expectation for the link length being

$$\langle \mathbf{r}^2 \rangle = b^2 \quad . \quad (40)$$

The probability distribution for the end-to-end vector is then

$$P(\mathbf{R}_e) = P(\{\mathbf{r}_n\}) \quad (41)$$

$$= \prod_{n=1}^N \left(\frac{3}{2\pi b^2}\right)^{3/2} \exp\left(-\frac{3\mathbf{r}_n^2}{2b^2}\right) \quad (42)$$

$$= \left(\frac{3}{2\pi b^2}\right)^{3/2} \exp\left(-\sum_{n=1}^N \frac{3(\mathbf{R}_n - \mathbf{R}_{n-1})^2}{2b^2}\right) \quad (43)$$

and hence for the entropy

$$S = \ln P = \sum_{n=1}^N \ln P(\mathbf{r}_n) \quad (44)$$

$$= \text{const} - \frac{3}{2b^2} \sum_{n=1}^N \mathbf{r}_n^2 \quad (45)$$

From this we obtain the free energy

$$F(\{\mathbf{r}_n\}) = E + \frac{3T}{2b^2} \sum_{n=1}^N \mathbf{r}_n^2 \quad (46)$$

with the internal energy  $E$  being independent of  $\{\mathbf{r}_n\}$ .

Hence we obtain the same equilibrium distribution as for the freely-jointed chain.

Eq (43) also results if we start off with the Hamiltonian for a chain of springs

$$H = \frac{3}{2} \frac{k_B T}{b^2} \sum_{n=1}^N (\mathbf{R}_n - \mathbf{R}_{n-1})^2 \quad (47)$$

and we also obtain the scaling of the end-to-end distance

$$\langle R_e^2 \rangle \propto N \quad . \quad (48)$$

A short-coming of the above models (besides they being phantom chains, i.e. no self-avoidance) is that there is no intrinsic stiffness. Intuitively, we expect a bending of the chain to cost energy. A model that provides this is the *worm-like chain model* (WLC). For this we start as above for the freely-rotating chain with a fixed persistence length and simultaneously letting the bond length  $b$  and the angle  $\theta$  go to zero. We are seeking thus a continuous description. We first pull on the results that we have derived before

$$\langle R_e^2 \rangle = \sum_{n=1}^N \sum_{m=1}^N \langle \mathbf{r}_n \mathbf{r}_m \rangle \quad (49)$$

$$= b^2 \sum_{n=1}^N \sum_{m=1}^N (\cos \theta)^{|m-n|} \quad (50)$$

$$= b^2 \sum_{n=1}^N \sum_{m=1}^N \exp\left(-\frac{|m-n|}{\xi_p}\right) \quad (51)$$

Since we want  $b$  to tend to zero we can substitute

$$b \sum_{n=1}^N \rightarrow \int_0^{R_{max}} ds \quad (52)$$

and thus

$$\langle \mathbf{R}_e^2 \rangle = \int_0^{R_{max}} ds \int_0^{R_{max}} ds' \exp\left(-\frac{|s' - s|}{\xi_p}\right) \quad (53)$$

with the result

$$\langle \mathbf{R}_e^2 \rangle = 2\xi_p R_{max} - 2\xi_p^2 \left(1 - \exp\left(-\frac{R_{max}}{\xi_p}\right)\right) \quad (54)$$

We need to consider two case. First we assume that  $R_{max} \gg \xi_p$ , then we recover the freely-jointed chain result

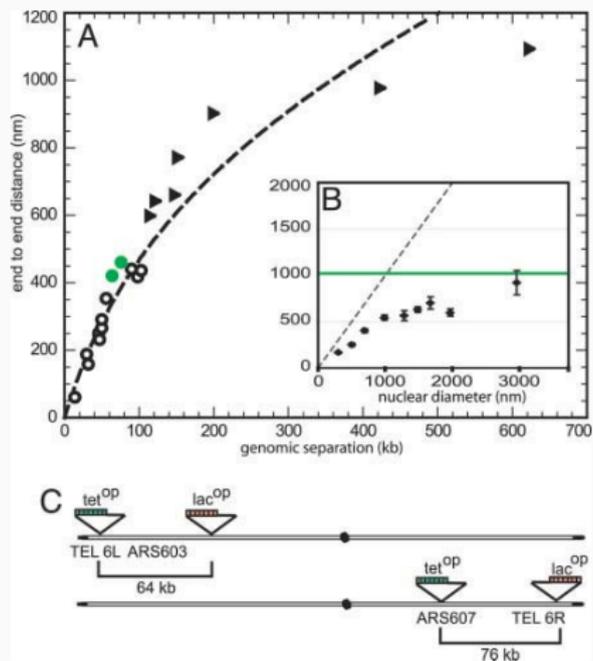
$$\langle \mathbf{R}_e^2 \rangle = 2\xi_p R_{max} \quad (55)$$

Second, if we assume that  $R_{max} \ll \xi_p$  then clearly

$$\langle R_e^2 \rangle \approx R_{max}^2 \quad (56)$$

so that the chain is just like a rod.

In Figure 7 is shown a comparison of the worm-like chain model with data on chromosomal yeast in interphase for small genomic distances.



**Figure 7:** Taken from *Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques*, Kerstin Bystricky, Patrick Heun, Lutz Gehlen, Jörg Langowski, and Susan M. Gasser, PNAS November 23, 2004 vol. 101 no. 47 16495-16500

As we have done before we are seeking to describe the worm-like chain model using a Hamiltonian. The idea is to use a coupling between the bond

$$H = -\epsilon \sum_{n=1}^{N-1} \mathbf{r}_n \cdot \mathbf{r}_{n+1} \quad (57)$$

which is simply the one-dimensional Heisenberg model for ferromagnets. Here  $|\mathbf{r}_n| = b$ . This model can be treated in the continuum limit where  $N \rightarrow \infty$ ,  $b \rightarrow 0$  and  $\epsilon \rightarrow \infty$  with

$$\epsilon/N = \text{constant} \quad , \quad (58)$$

keeping the contour length also constant. Using

$$-\mathbf{r}_n \cdot \mathbf{r}_{n+1} = \frac{1}{2} [(\mathbf{r}_n - \mathbf{r}_{n+1})^2 - 2b^2] \quad (59)$$

we have

$$H = \lim_{b \rightarrow 0; \epsilon, N \rightarrow \infty} \frac{\epsilon b}{2} \sum_{n=1}^{N-1} b \left( \frac{\mathbf{r}_n - \mathbf{r}_{n+1}}{b} \right)^2 . \quad (60)$$

To cross over to the continuum limit we use the tangent vector with the arc length  $s$

$$\frac{\partial \mathbf{r}(s)}{\partial s} = \lim_{b \rightarrow 0} \left( \frac{\mathbf{r}_{n+1} - \mathbf{r}_n}{b} \right) \quad (61)$$

and  $\sum_{n=1}^{N-1} b \rightarrow \int_0^L ds$  to find

$$H = \frac{\kappa}{2} \int_0^L ds \left( \frac{\partial \mathbf{r}(s)}{\partial s} \right)^2 = \frac{\kappa}{2} \int_0^L ds \left( \frac{\partial^2 \mathbf{R}(s)}{\partial s^2} \right)^2 \quad (62)$$

with the *bending modulus*  $\kappa = \epsilon b$ .

Thus the partition function is given by

$$Z = \int \mathcal{D}[\mathbf{r}(s)] \delta(|\mathbf{r}(s)| - 1) \exp(-\beta H[\mathbf{r}(s)]) . \quad (63)$$

The bending modulus must be related to the persistence length. To find this relation we need to calculate the correlation function

$$\langle \mathbf{r}(s)\mathbf{r}(s') \rangle \propto \exp(-|s - s'|/\xi_p) \quad . \quad (64)$$

We can now calculate the mean squared end-to-end-distance and the mean squared radius of gyration

$$\langle R_e^2 \rangle = \left\langle \left( \int_0^L ds \mathbf{r}(s) \right)^2 \right\rangle \quad (65)$$

$$= \int_0^L ds \int_0^L ds' \langle \mathbf{r}(s) \cdot \mathbf{r}(s') \rangle \quad (66)$$

$$= 2\xi_p^2 \left( \frac{L}{\xi_p} - 1 + e^{-L/\xi_p} \right) \quad (67)$$

$$= L^2 f_D \left( \frac{L}{\xi_p} \right) \quad , \quad (68)$$

where  $f_D(x) = 2(x - 1 + e^{-x})/x^2$  being the Debye-function (see Figure 8).

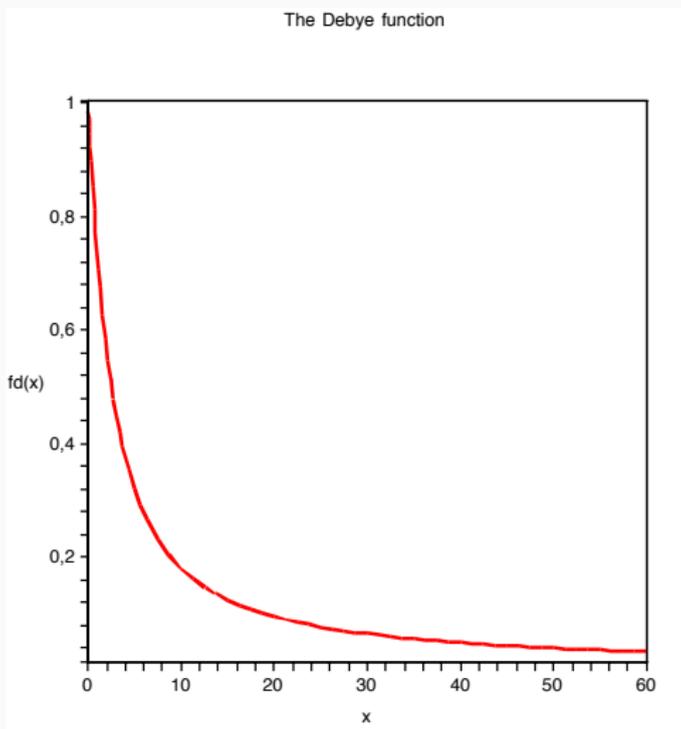


Figure 8: The Debye-function

The self-avoiding random walk (SAW) on a periodic lattice was considered by Orr [7] as a model of a polymer chain. Such a self-avoiding random walk is shown in Figure 11. In one dimension the problem of computing the partition function and other properties such as the end-to-end distance is trivial and unsolved in higher dimensions.

Let  $c_N$  denote the number of  $n$  – step self-avoiding walks (SAW) (equivalent upon translation!). We can easily enumerate on the square lattice  $c_1 = 4$ ,  $c_2 = 12$ ,  $c_3 = 36$  and  $c_4 = 100$  and a simple estimate yields,

$$d^N \leq c_N \leq 2d(2d - 1)^{N-1} \quad (69)$$

$$d^N \leq c_N \leq 2d(2d - 1)^{N-1} \quad (70)$$

In general it is believed to be [8–10]

$$c_N \approx A\mu^N N^{\gamma-1} \quad (71)$$

with  $\gamma$  being a universal exponent ( $d = 2 \gamma = 32/43$ ,  $d = 3 \gamma \approx 7/6$ ,  $d \geq 4 \gamma = 1$ ) and  $\mu$  the *connectivity constant* giving the average number of available steps for an infinitely long walk.

For the partition function we have

$$Z_N \sim q_\mu^N N^{\gamma-1} \quad q_{\text{eff}} < q(\Lambda) \quad (72)$$

and thus for the average end-to-end distance

$$\langle R_e^2 \rangle \propto N^{2\nu} \quad (73)$$

with  $\nu \approx 0.59$  (in 3d) and  $\gamma \approx 1.158$  (in 3d) from numerical calculations.

Polymer Chain

Random Walk

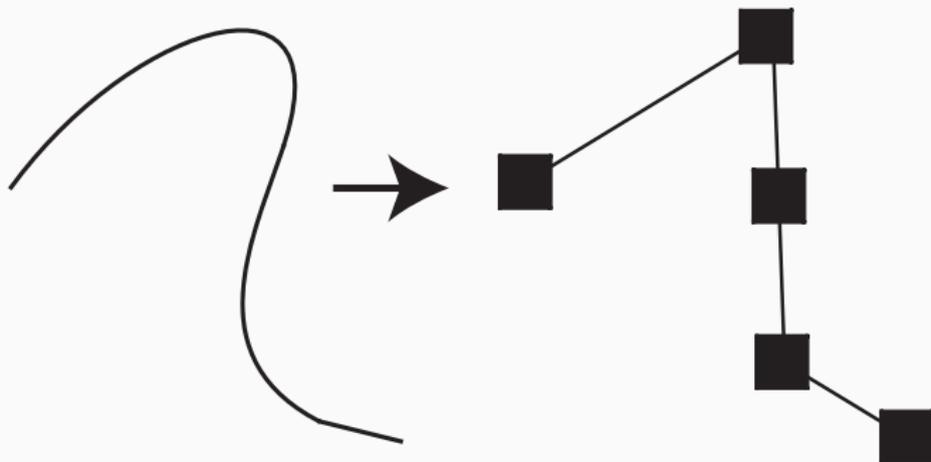
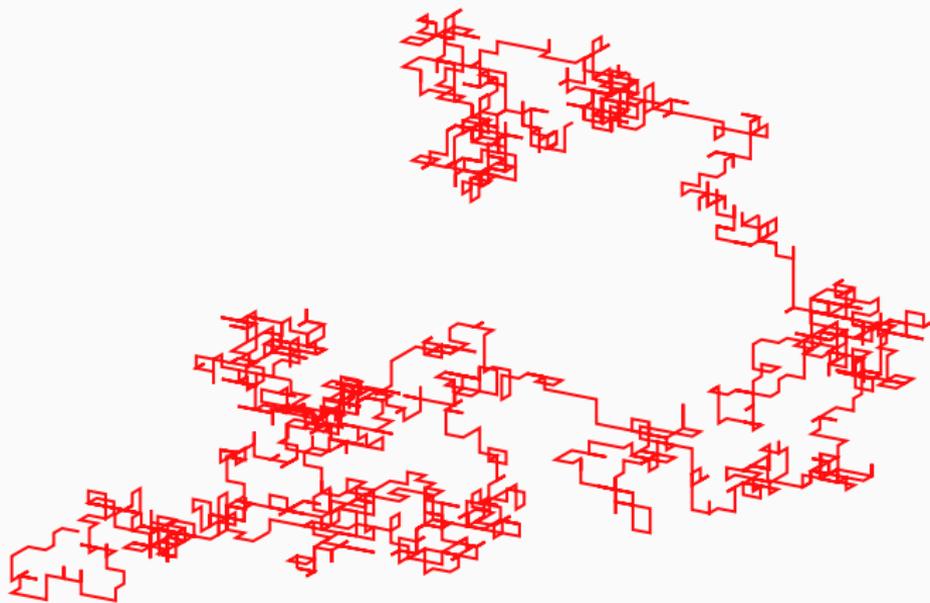


Figure 9: From a continuous to a lattice description



**Figure 10:** A sample of a random walk in three dimensions on a lattice

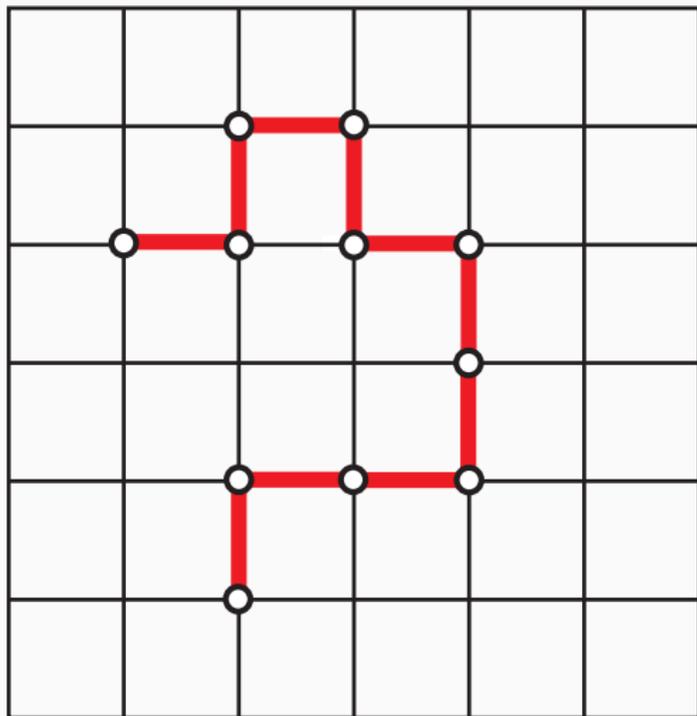


Figure 11: A self-avoiding random walk (SAW)

## Polymers at Membranes

---

- It allows to study the entropic effects on the wall as well as further studies using realistically modeled biopolymers.
- The model interpolates between the united atom model and the bead-spring model. In contrast to these two models it uses non-spherical force fields for the non-bonded. interaction
- The main idea of this approach with a more general form of the force field is to generalize the united atom model in a way that larger atom groups are combined to one construction unit, but the possible anisotropy of these groups is still taken into account.
- As one wants the force field to degenerate into a sphere with increasing distance, we use a con-focal force field inside this interaction volume:

$$\mathcal{H}_{\text{inter}} = V_{\text{abs}} \left( \frac{d_1^{(p)} + d_2^{(p)}}{2} - c \right), \quad (74)$$

where  $d_1^{(p)}$  and  $d_2^{(p)}$  denote the distance of the point  $\mathbf{p}$  to the focal points of the ellipsoid and  $V_{\text{abs}}$  is the absolute potential.

- For convenience we use only a repulsive part

$$V_{\text{abs}}(r) \sim r^{-6} \quad . \quad (75)$$

- The mass of the building units is distributed between the focal points of the ellipsoids in the hard core region of the con-focal potential.
- The main ingredient of the model is the mass matrix of our rod-chains. In order to construct it we, must first calculate the Lagrangian of a single rod  $\mathcal{L}_i = T_i - V_i$  with the kinetic energy  $T_i$  and the potential energy  $V_i$ . The subindex  $i$  marks the position of the rods in the chain. This one-dimensional homogeneous rod  $i$  has the length  $l_i$  starting at  $\vec{a}_i$  and ending at  $\vec{b}_i$ .
- If we suppose that the rods all have the same mass  $m$  and that the velocity of the rod mass scales linearly with the position between the boundaries of the rod, the kinetic energy can be written as

$$\begin{aligned} T_i &= \frac{1}{2} \int_0^{l_i} \frac{m}{l_i} \left( \frac{(l_i - x) \dot{\vec{a}} + x \dot{\vec{b}}}{l_i} \right)^2 dx \\ &= \frac{1}{6} m (\dot{\vec{a}}^2 + \dot{\vec{a}} \dot{\vec{b}} + \dot{\vec{b}}^2). \end{aligned}$$

- Adding the single terms of the rods building the chain we get the Lagrangian  $\mathcal{L}$  of the whole rod chain.
- The equations of motion of the chain can be calculated from the Lagrange equations of the second kind. Since the equations of motion separate in each direction, we have only to solve three tridiagonal  $(N + 1) \times (N + 1)$  matrices per chain which consist of  $N$  rods per time step of the form

$$\mathbf{W}\ddot{\vec{x}} = \vec{F} \quad (76)$$

$$\frac{m}{6} \begin{pmatrix} 2 & 1 & 0 & 0 & \dots \\ 1 & 4 & 1 & 0 & \dots \\ 0 & 1 & 4 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \ddot{x}_0 \\ \ddot{x}_1 \\ \ddot{x}_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} F_{10} \\ F_{11} + F_{21} \\ F_{22} + F_{32} \\ \vdots \end{pmatrix} \quad (77)$$

with the force  $F_{ij}$  on the coordinate  $j$  of the flexible point  $i$  of the chain

$$F_{ij} = -\frac{\partial V_i}{\partial j} \quad (78)$$

and  $\ddot{x}_i$  denote the accelerations of the flexible points of the chain. The flexible points are the link points of the ellipsoids and the end points of the rod chain. The sub-indices mark the positions in the chain: 0 and  $N + 1$  are the end-points

of the chain and the numbers between them denote the linking points of rods in the chain.

- The bonded interactions between neighboring units are given by harmonic length and angle potentials:

$$\mathcal{H}_{bond} = \frac{1}{2}k(r - r_0)^2 \quad (79)$$

$$\mathcal{H}_{angle} = \frac{1}{2}k_\theta(\cos \theta - \cos \theta_0)^2 \quad (80)$$

with the bond lengths  $r$  and the bending angles  $\theta$ . Here  $r_0$  and  $\theta_0$  denote the mean values.

- Consider a wall with a repulsive  $r^6$  potential and a polymer grafted at the wall. The constraint that the polymer is grafted and that one half-space is excluded leads to a competition between the necessity to avoid the wall and the constraint to be fixed at the wall.
- Due to the entropy the monomers would like to stay as far away from the wall as possible.
- In order to do so they exert a pressure on the wall.
- This pressure decreases radially from the grafting point.
- For a theoretical treatment of the pressure we shall regard an elastic wall.
- Let the surface of the wall be described by  $h(x, y)$ . The thermodynamic properties of the chain of length  $N$  grafted at the repulsive wall can be described by the propagator  $G_N(\vec{r}, \vec{r}')$  resulting from the Edwards equation

$$\frac{\partial G_N(\vec{r}, \vec{r}')}{\partial N} = \frac{l^2}{6} \Delta G_N(\vec{r}, \vec{r}') \quad (81)$$

with the  $G_N(r, r') = 0$  at the wall and  $\lim_{N \rightarrow 0} G_N(r, r') = \delta(r, r')$ .

- The partition function is then given by

$$Z_N(l) = \int d\vec{r}' G_N(\vec{r}, \vec{r}') \quad , \quad (82)$$

where the integral extends over all space that is available to the free end. The Greens-function for a planar wall  $h(x, y) = 0$  can then be factorized as

$$G_N^{(0)}(\vec{r}, \vec{r}') = \left( \frac{3}{2\pi Nl^2} \right)^{3/2} \exp \left[ -\frac{3(x-x')^2}{2Nl^2} \right] \exp \left[ -\frac{3(y-y')^2}{2Nl^2} \right] \\ \times \left( \exp \left[ -\frac{3(z-z')^2}{2Nl^2} \right] \exp \left[ -\frac{3(z+z')^2}{2Nl^2} \right] \right) .$$

- The partition function is therefore

$$Z_N^{(0)}(l) = \int_{-\infty}^{+\infty} dx' \int_{-\infty}^{+\infty} dy' \int_0^{+\infty} dz' G_N^{(0)}(\vec{l}, \vec{r}') \quad (83)$$

$$= \operatorname{erf} \left( \frac{l}{2R_g} \right) \quad , \quad (84)$$

where  $R_g = \sqrt{Nl^2/6}$  is the radius of gyration of the free chain and erf the error function.

- To compute the pressure we introduce a small perturbation in  $h$ . We can write the partition function as  $Z_N = Z_N^{(0)} + Z_N^{(1)} + Z_N^{(2)} + \dots$ , where  $Z_N^{(i)}$  is of order  $h^i$  and  $Z_N^0$  as in (83).
- Due to the linearity of (81), each term satisfies the Edwards equation

$$\frac{\partial Z_N^{(i)}}{\partial N} = \frac{l^2}{6} \Delta Z_N^{(i)} \quad i = 0, 1, 2, \dots \quad (85)$$

- The solutions of higher orders are coupled to the constraint. Now we have

$$\begin{aligned} \mathbf{0} &= Z_N(x, y, h) \\ &= Z_N(x, y, \mathbf{0}) + h(x, y) \frac{\partial Z_N}{\partial z}(x, y, \mathbf{0}) + \frac{h^2(x, y)}{2} \frac{\partial^2 Z_N}{\partial z^2}(x, y, \mathbf{0}) + \dots \end{aligned}$$

- For the linear contribution  $Z_N^{(1)}$  we get

$$Z_N^{(1)}(x, y, \mathbf{0}) = -h(x, y) \frac{\partial Z_N^{(0)}}{\partial z}(x, y, \mathbf{0}) \quad , \quad (86)$$

yielding [? ]

$$Z_N^{(1)}(\vec{l}) = \frac{l^2}{6} \int_{\mathbf{0}}^N dn \int dS \frac{\partial G_{N-n}^{\mathbf{0}}}{\partial z}(x, y, \mathbf{0}; \vec{a}) Z_n^{(1)}(x, y, \mathbf{0}) \quad . \quad (87)$$

- Hence, the change in the height is, to first order, due to the work

$$\begin{aligned}\Delta W &= W[h] - W[0] \\ &= -k_B T \log \left[ 1 + \frac{Z_N^1}{Z_N^0} \right] \\ &= \int dS \rho(x, y) h(x, y) \quad ,\end{aligned}$$

where  $\rho(x, y)$  has the symmetric form

$$p(r) = \frac{k_B T}{2\pi(r^2 + l^2)^{3/2}} \left( 1 + \frac{r^2 + l^2}{2R_g^2} \right) \exp \left[ -\frac{r^2 + l^2}{4R_g} \right] \quad (88)$$

with  $r = \sqrt{x^2 + y^2}$ .

- To push at  $\vec{r} = (x, y)$  an elementary volume of  $dV(r) = h(r)dS$  we need the work  $dW = p(r)h(r)dS$ .
- The function  $p(r)$  is the pressure.

- The entire entropic force which the chain exerts on the wall is then given by

$$F = \int_0^\infty dr 2\pi r p(r) = \frac{k_B T}{l} \exp\left[-\frac{l^2}{4R_g^2}\right] \quad (89)$$

$$= \frac{k_B T}{l} \exp\left[-\frac{3}{2N}\right] \quad . \quad (90)$$

- 10 different chain lengths

$N = 20, N = 40, N = 60, N = 80, N = 100, N = 125, N = 150, N = 175, N = 200$   
and  $N = 250$  to study the pressure and the corresponding finite effects.

## Radius vs. Height Distribution

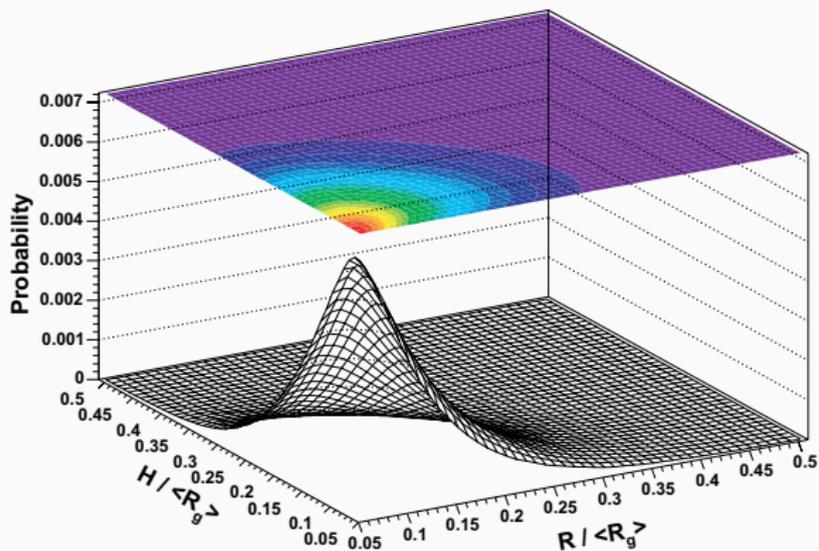
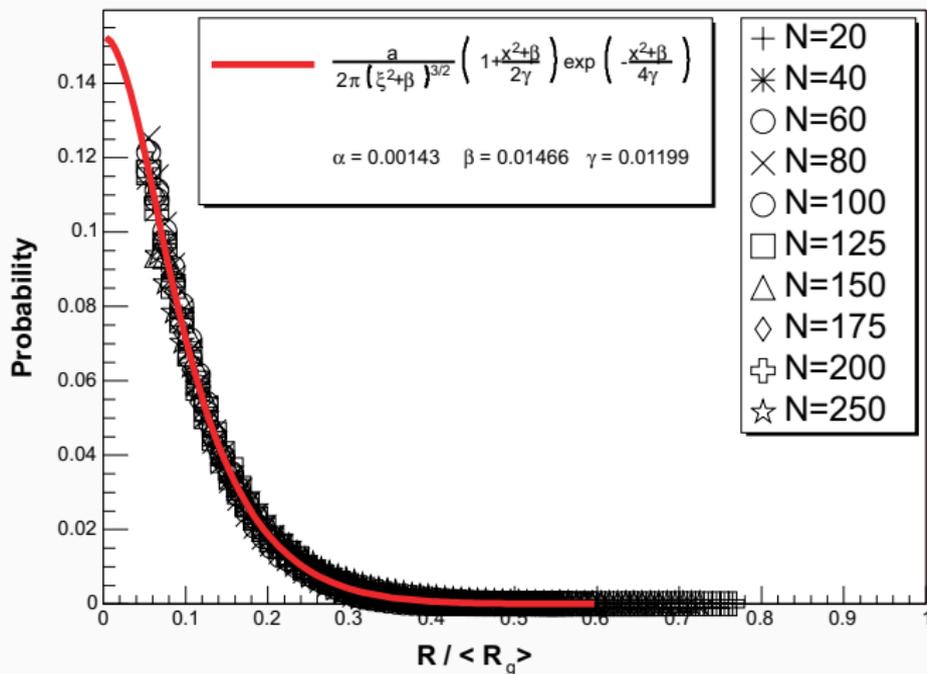
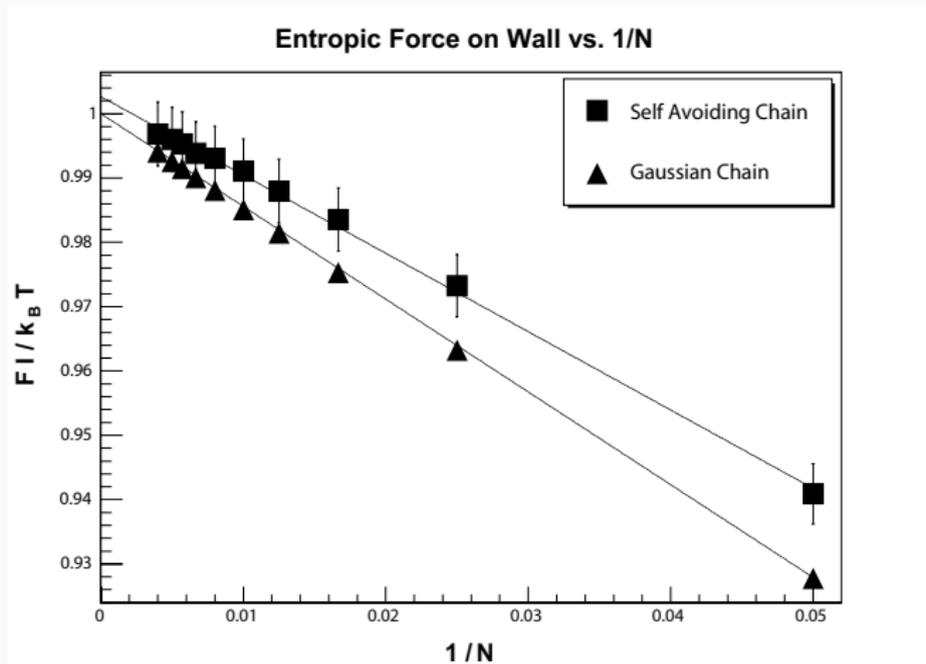


Figure 12: Shown is the distribution of the height with respect to the radius. The values are divided by their respective radii of gyration for reasons of finite-size scaling.

Finite Size Scaled Pressure Distribution for all Chain Lengths



**Figure 13:** Finite-size scaling plot for the pressure distribution. The full curve is the fit using the predicted form of the distribution. The fit is very good over the entire region.





**Figure 14:** Shown is the entropic force exerted on the wall by the polymer. The figure gives the result for the Gaussian and for the chain with self-avoidance.

- Knowing the distribution function, we can calculate the partition function of the polymer with an external force:

$$Z = \int d\vec{r} p_N(\vec{r}) \exp\left(\frac{\vec{f} \cdot \vec{r}}{T}\right). \quad (91)$$

But it is also possible to derive the desired results by fundamental reasoning.

- We only need to introduce the two characteristic lengths for the problem,  $R_F \cong lN^\nu$  and  $\xi_p = T/f$ . In general, the norm of the mean end to end distance vector can be written as:

$$\left| \langle \vec{r}(\vec{f}) \rangle \right| = R_F \varphi_r\left(\frac{R_F}{\xi_p}\right) = R_F \varphi_r(x), \quad (92)$$

where  $\varphi_r(x)$  is a dimensionless function.

- In the case of small forces one expects a linear response of the polymer, so that we can write  $\lim_{x \rightarrow 0} \varphi_r(x) \cong x$ . Using this we get:

$$\left| \langle \vec{r}(\vec{f}) \rangle \right| \cong \frac{R_F^2}{T} f. \quad (93)$$

- If the chain is stretched stronger, we expect deviations from the linear law. Let us assume that the stretched chain is composed of “blobs”, i.e. small chain-balls. Each of these blobs has a size of  $\xi_p$ . In such a blob the external force is just a small perturbation, so we can write for the number of monomers  $g_p$  in the blob:

$$\xi_p \cong l g_p^\nu \quad (94)$$

or:

$$g_p = \left( \frac{T}{fl} \right)^{1/\nu} . \quad (95)$$

- Considering that the number of the blobs must be  $N/g_p$ , one obtains for the three dimensional case:

$$\left| \langle \vec{r}(\vec{f}) \rangle \right| \cong \frac{N}{g_p} \xi_p \cong Nl \left( \frac{fl}{T} \right)^{0.689} . \quad (96)$$

- Hence for large forces the elongation behavior is not linear. For the case of stretched polymers one can look again at the distribution function, which has the form  $\exp(-(r/R_F)^\delta)$ . The resulting entropy is:

$$S(r) = \text{const} + \ln p_N(r) = \text{const} - \left( \frac{r}{R_F} \right)^\delta . \quad (97)$$

- In this case the corresponding elastic free energy amounts to:

$$F_{tot} = T \left( \frac{r}{R_F} \right)^\delta - fr. \quad (98)$$

- If one minimizes this expression, one obtains the wanted relation between force and end-to-end distance:

$$f \cong \delta \frac{T}{R_F} \left( \frac{r}{R_F} \right)^{\delta-1}. \quad (99)$$

- We have seen how to calculate the relation between applied force and resulting elongation for long chains with self-avoiding as well as without self-avoiding. This result is important but not satisfactory. If one considers that the polymer cannot rupture, than the extension should be  $Nl$  for very large forces but in the results above it seems that the polymer chain can be stretched to infinite length. Furthermore, the case of a restricted geometry is not included. Both will be done in the next sections.

- First we calculate the work that is performed by a force  $\vec{f}$  if the polymer is elongated by  $d\vec{R}$ . This is:

$$\delta A = -\vec{f} \cdot \delta \vec{R} = -\sum_{i=1}^N \vec{f} \cdot d\vec{r}_i = -\sum_{i=1}^N d\varphi_i, \quad (100)$$

where  $\varphi_i = \vec{f} \cdot \vec{r}_i = f \cdot l \cdot \cos(\vartheta_i)$ . So the partition function is

$$Z = \int \exp\left(\sum_{i=1}^N \left(\frac{f \cdot l}{T}\right) \cos(\vartheta_i)\right) \prod_{i=1}^N \sin(\vartheta_i) d\vartheta_i d\varphi_i. \quad (101)$$

- The multidimensional integral can be separated:

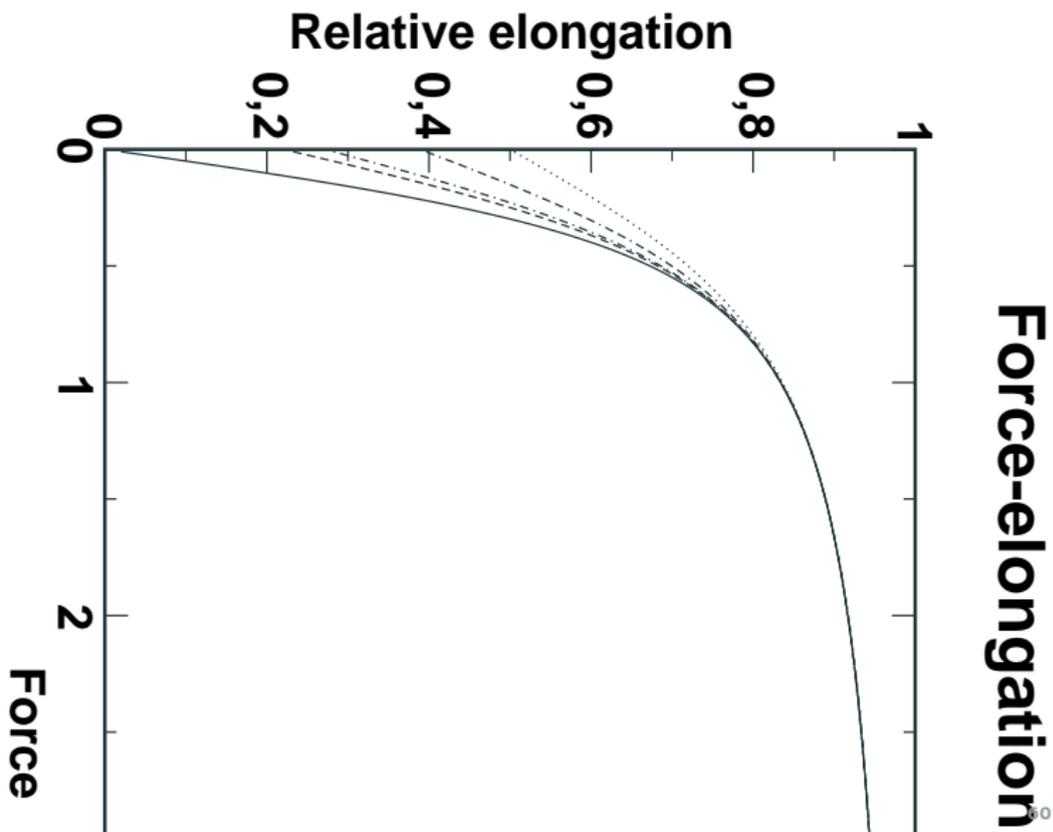
$$Z = \left( \int_0^{2\pi} \int_0^\pi \exp\left(\frac{f \cdot l}{T} \cos(\vartheta)\right) \sin(\vartheta) d\vartheta d\varphi \right)^N. \quad (102)$$

If one introduces  $\beta = \frac{f \cdot l}{T}$ , one gets:

$$Z = \left( \frac{4\pi \sinh(\beta)}{\beta} \right)^N. \quad (103)$$

- The exact force-elongation behavior is:

$$R = \left| \vec{R} \right| = N \cdot l \cdot \left( \coth(\beta) - \frac{1}{\beta} \right). \quad (104)$$





# Modelling of Biopolymers

---

Biopolymers can be modeled for computational purposes in a variety of ways [1]. Depending on the kind of question and the degree of abstraction, one has the basic choice between a model on a lattice or in continuous space. The bond fluctuation model [2] is one of the prominent representatives of a polymer model on the lattice. The main advantage of this type of models is the computational efficiency due to the restricted configuration space. With increasing computer power it was possible to stay closer to reality by simulating polymers by continuum models. Two widely used models of this class are the bead-spring [3] and the united-atom model [4].

In both models monomers or parts of them are considered to be represented by spherical force fields. In the united atom model the  $\text{CH}_2$  groups are modeled by a spherical force field and the bonded interactions by harmonic forces. In this more atomistic model the anisotropic intermolecular potential functions of polyatomic molecules are constructed using spherical force fields. As an effect the inner degrees of freedom of the molecules like the stiff bonds between the units must also be taken into account. As the Newton equations have to be integrated such molecular-dynamic simulations are restricted to small time scales.

Other models have been developed in order to adapt an aspherical model to a molecule's geometry i.e. J. Kushick's and B.J. Berne's model [11] and J.G. Gay's and B.J. Berne's model [12]. They consider ellipsoids as a model for molecules and

calculate the forces between two interacting ellipsoids as a function of the overlap volume.

The continuous backbone mass model in some sense interpolates between of the united atom model and the bead spring model. On the one hand it tries to stay as close as possible to the chemical realistic structure like the united atom model, but on the other hand it integrates out all the inner degrees of freedom just the same as the bead spring model in order to be computationally efficient. In contrast to these two models it uses *non-spherical* force fields for the non-bonded interaction. The main idea of this approach with a more general form of the force field is to generalize the united atom model in a way that larger atom groups are combined to one construction unit, but the possible anisotropy of these groups is still taken into account. The reasoning is that the topology of the monomer has a strong influence on the physical properties. The simplest anisotropic geometrical object one can think of is an ellipsoid of rotational symmetric form and thus it is considered as the interaction volume of the chemical sequences in our model.

As one wants the force field to degenerate into a sphere with increasing distance, we use a con-focal force field inside this interaction volume:

$$H_{\text{inter}} = V_{\text{abs}} \left( \frac{d_1^{(p)} + d_2^{(p)}}{2} - c \right), \quad (130)$$

where  $d_1^{(p)}$  and  $d_2^{(p)}$  denote the distance of the point  $\mathbf{p}$  to the focal points of the ellipsoid and  $V_{\text{abs}}$  is the absolute potential. In the case of the BPA-PC we take only a repulsive part

$$V_{\text{abs}}(r) = r^{-6} \quad (131)$$

into account because from quantum chemical calculations the attractive part proves to be negligible. The calculation of the distances is illustrated in Figure 21.

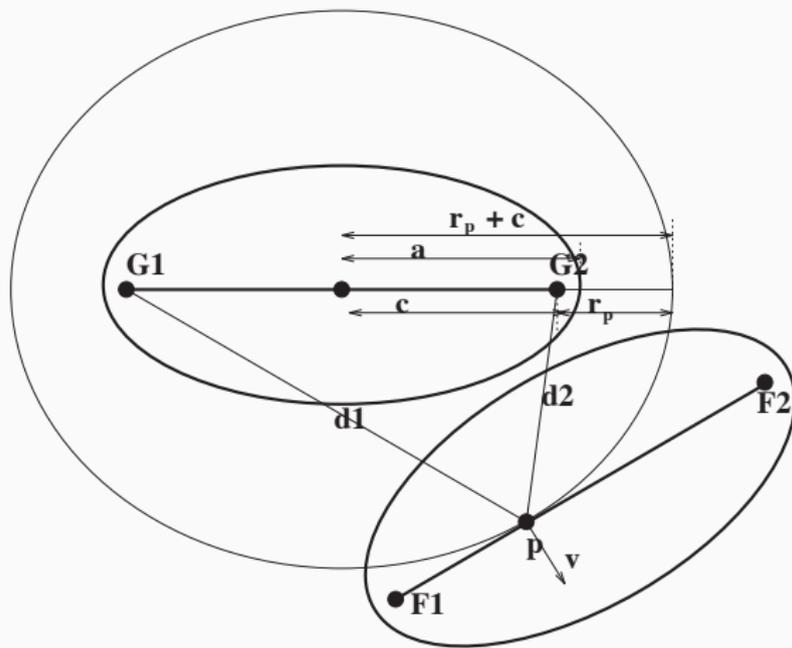


Figure 21: Interaction with a con-focal force field

To be able to predict the folded structure, we crucially depend on an energy function. The energy function of all the parameters are used to describe the protein structure. The task is then to find values of the parameters which minimize this function.

Molecular mechanics describes the energy of a molecule in terms of a simple function which accounts for distortion from ideal bond distances and angles, as well as and for nonbonded van der Waals and Coulombic interactions. Thus, such force field methods ignore the electronic motions to calculate the energy of a system.

To model macromolecular systems the *CHARMM* potential (Chemistry at HARvard Macromolecular Mechanics) [13, 14] , AMBER and GROMOS (GRONingen MOLEcular Simulation System) force fields are often used. They are empirical force field parametrizations that consists in general of six terms:

$$V(\{\mathbf{R}\}) = \sum_{\text{bonds}} c_i (l_i - l_0)^2 \quad (132)$$

$$+ \sum_{\text{bond angles}} c_\alpha (\theta_\alpha - \theta_0)^2 \quad (133)$$

$$+ \sum_{\text{improper torsion angles}} c_\beta (\tau_\beta - \tau_0)^2 \quad (134)$$

$$+ \sum_{\text{dihedral angles}} \text{tri}(\omega) \quad (135)$$

$$+ \sum_{\text{charged pairs}} \frac{Q_i Q_j}{\epsilon r_{ij}} \quad (136)$$

$$+ \sum_{\text{unbond pairs}} c_w \Phi \left( \frac{R_i + R_j}{r_{ij}} \right) \quad (137)$$

where

$$r_{ij} = |\mathbf{R}_i - \mathbf{R}_j| \quad . \quad (138)$$

Here  $\epsilon$  is the dielectric constant and  $Q_i$  are the partial charges. The term  $\text{tri}$  refers to a linear combination of trigonometric functions and multiples of  $\omega$ . The term  $\Phi$  refers to a Lennard-Jones potential. The parameters  $c$  etc. are usually fitted and derived from first principles.

The approach taken by the Molecular Dynamics and the Langevin Dynamics method discussed in the next section is to solve the equations of motion resulting from a force field, such as the one above, numerically.

The total potential energy of a single macromolecule can be divided into bonding and non-bonding parts. The bonding energies are due to the local covalent bonds. We shall now focus on the non-bonding part.

Charged polymers are essential for biology. Many functions depend critically on the activity of DNA, RNA and proteins all of which are charged polymers. A human nuclear DNA molecule carries hundreds of millions of charged groups. Inseparable from the polymer phenomenology is the behavior of the small charged mobile anions which dissociate from their backbones and interact strongly with the polymer chains.

Interactions between charges relevant in biology are almost always affected by the presence of water molecules, ions, and other molecules. The interactions are reduced in strength or are screened.

Non-covalent interactions are much weaker than the covalent bonds. We classify the non-covalent interactions in (decreasing in strength)

- Ionic interactions
- Intermediate dipole-dipole forces
- Hydrogen bonds
- Hydrophobic interactions

- Van der Waals interactions: The force arises from induced dipole and the interaction is weaker than the dipole-dipole interaction.

## ■ Ionic interactions

Let us first discuss the ionic interactions. What we always have to consider is water. Pure  $\text{H}_2\text{O}$  has pH 7.0. Hence even pure water is not the simple dielectric. Many proteins, nucleic acids, and other organic molecules in cells are charged. They give up ions to the solution when they are put in water. A good example is DNA, which has one phosphate ion ( $\text{PO}_4^-$ ) on each nucleotide. The anion is usually  $\text{Na}^+$ .

Important functional ionic groups are for the anionic case

- carboxy
- sulfate
- sulfonate
- phosphate

and for the cationic case

- amino
- imino
- ammonium
- sulfonium

- phosphonium

On a more detailed scale biologically the following ions are of interest:

- **Univalent ions.** The cations  $\text{Na}^+$  and  $\text{K}^+$  are present at roughly 0.1 M concentrations, outside and inside cells, respectively. There are anions (negative ions) at the same concentration to balance the charge.
- **Divalent ions** Charge-2 cations like  $\text{Mg}^{2+}$  and  $\text{Ca}^{2+}$  are present at roughly mM concentrations in cells, and in many biochemistry experiments.

Let  $q_1$  and  $q_2$  be two charges. Then the potential is given by

$$U(r) = \frac{kq_1q_2}{r} \quad (139)$$

with  $k = 8.99 \times 10^9 \text{ N}\Delta\text{m}^2/\text{C}^2$  in vacuum.

**Table 2:** Potential for two charges separated a distance of about 2nm at room temperature

vacuum	$30k_B T$
water	$0.4k_B T$

The charges, which are embedded in a dielectric material (water) interact by a Coulomb interaction which is reduced in strength measured by the dielectric constant. Water molecules have a very large electric dipole moment and are forced to rotate to respond to an alternate external electric field. Hence water as a liquid has a very large dielectric constant 80 at room temperature. This reduction in strength of the Coulomb interaction is due to the polarization of the particles of the dielectric medium - either induced or permanent dipoles around a free charge will be oriented so as to terminate some of the field lines coming from the free charge. The Coulomb interaction continues to have its long-ranged character, just with a reduced strength (dielectric screening of the charge).

$$\lambda_D = \sqrt{\frac{\epsilon_0 \epsilon_r k_B T}{2 N_A e^2 I}} \quad (140)$$

where  $\epsilon_0$  is the permittivity of free space.

The *Bjerrum length*  $l_B$  is the distance at which the Coulomb interaction between two unscreened charges equals the thermal energy

$$l_B = \frac{e^2}{4\pi\epsilon\epsilon_0 k_B T} \quad (141)$$

## Example

The Bjerrum length at which two electron charges have an interaction energy about  $k_B T$  in pure water at standard pressure and temperature is 0.7 nm where we have used

$$k = 9 \times 10^9 \text{ N}\Delta\text{m}^2/\text{C}^2 \quad (142)$$

$$e = 1.6 \times 10^{-19} \text{ C} \quad (143)$$

$$e = 80 \quad (144)$$

$$k_B T = 4 \times 10^{-21} \text{ J} \quad (145)$$

There is another very important screening effect that arises from free ions in solution and which will cluster around charged objects, counter-ions and further reduces the strength and range of the Coulomb interactions. These counter-ions can be associated with the charged objects with energies in excess of  $5k_B T$  and hence are not easily shaken off by thermal energy.

## ■ Charge-dipole and dipole-dipole interactions

Many molecules are electrically neutral but have a permanent dipole because of an asymmetric distribution of the electron cloud around the positively charged nuclei. For example, in HCl, the valence electron of the H atom is donated to the Cl atom, with H carrying a net positive charge, and Cl a net negative charge. Similarly, water has a permanent dipole because the electron density is greater near the more electro negative O atom.

Recall that the *dipole moment* is defined as  $|\mathbf{q}| = qd$  where  $d$  is the separation between 2 charges  $+q$  and  $-q$ .  $\mathbf{p}$  is a vector and points in the direction from  $-q$  to  $+q$ . When a molecule with a dipole moment  $\mathbf{p}$  is placed in an electric field  $\mathbf{E}$ , the dipole has a potential energy

$$U(\theta) = \mathbf{p} \cdot \mathbf{E} = -pE \cos \theta \quad . \quad (146)$$

## ■ Charge-dipole interactions

The electric field from a single point charge at a distance  $r$  from the charge is

$$E = kq/r^2 \quad . \quad (147)$$

The potential energy of a charge-dipole system is

$$U(r) = -\frac{k|q|}{r^2} \cos \theta \quad . \quad (148)$$

The potential energy now falls off as  $1/r^2$ , more rapidly than the charge-charge system. In the absence of thermal motion, the dipole will align with the  $\mathbf{E}$  field, which corresponds to  $\theta = 0$ .

Because of random collisions with the molecules of the surrounding medium, the dipole will undergo a Brownian motion. Here we will consider only the change in the orientation of the dipole as a result of random collisions and write down the Boltzmann probability that the dipole makes an angle  $\theta$  with the  $\mathbf{E}$  field as

$$P(\theta) \propto \exp(-\beta U) = \exp(pE\beta \cos \theta) \quad (149)$$

The average value of the potential energy, averaged over all possible orientations whose probability is given by the Boltzmann distribution, can be written as

$$\langle U \rangle = -\frac{1}{Z} \int_0^\pi d\theta pE \cos \theta e^{pE\beta \cos \theta} 2\pi \sin \theta \quad , \quad (150)$$

where  $Z$  is the normalization constant. Solving this integral yields

$$\langle U \rangle = -pE \left( \coth \left( \frac{pE}{k_B T} \right) - \frac{k_B T}{pE} \right) \quad (151)$$

which, in the limit  $pU \ll k_B T$ , simplifies to

$$\langle U \rangle = -\frac{p^2 E^2}{3k_B T} \quad (152)$$

Substituting for the electric field due to a point charge  $q$  at a distance  $r$  from the charge, we get

$$\langle U \rangle = -\frac{(kqp)^2}{3k_B T r^4} \quad (153)$$

With thermal averaging, the charge-dipole interaction falls off as  $1/r^4$ .

## ■ Dipole-dipole interactions

The interaction energy of two permanent dipoles depends on their relative orientation, and might be expected to be zero overall for a compound if all orientations are possible. This would be true if the molecules were completely free to rotate, but they are not and some orientations are preferred over others. Let  $\mathbf{p}_1$  and  $\mathbf{p}_2$  be two dipoles. The potential energy has the form

$$U(r) = -\mathbf{E}_1 \cdot \mathbf{p}_2 = k \frac{p_1 p_2}{r^3} F(\theta_1, \phi_1, \theta_2, \phi_2) \quad , \quad (154)$$

where  $\mathbf{E}_1$  is the electric field from dipole  $\mathbf{p}_1$  and depends on the angular position of  $\mathbf{p}_2$  relative to  $\mathbf{p}_1$  and their relative orientations. The distance dependence  $1/r^3$  comes from the radial dependence of the electric field  $\mathbf{E}_1$  of dipole  $\mathbf{p}_1$ .

The thermal averaging with Boltzmann probabilities, in the limit  $U \ll k_B T$ , gives

$$\langle U \rangle = -\frac{2}{3} \frac{(k p_1 p_2)^2}{3 k_B T r^6} \quad . \quad (155)$$

Thus the potential energy between two dipoles falls off as  $1/r^6$  power, i.e., dipole-dipole interactions are short-range interactions.

## ■ Van der Waals interactions

Perhaps the most important class of dipole-dipole interactions are the ones where one or both molecules do not have a permanent dipole. These interactions are valid for any two atoms that come into close contact with each other, and are called Van der Waals interactions. .

## ■ Dipole-induced dipole interactions

A molecule with a permanent dipole  $p_1$  can induce a dipole in another polarizable molecule. In this case the induced dipole moment  $p_2^*$  points in the same direction as the inducing electric field  $E_1$ . The potential energy of interaction between  $p_1$  and  $p_2^*$  takes the form

$$U(r) = -\frac{kp_1p_2^*}{r^3} f(\theta) \quad (156)$$

where the minus sign indicates that the interaction is always attractive, since the induced dipole always follows the direction of the instantaneous electric field.  $\theta$  defines the angular position of  $p_2^*$  relative to  $p_1$  and the electric field  $E_1$  is independent of the azimuthal angle  $\phi$ .

The magnitude of  $p_2^*$  depends upon the strength of the electric field at position  $(r, \theta)$

$$p_2^* = \epsilon_0 \alpha_2 E_1(r, \theta) = \epsilon_0 \alpha_0 \frac{kp_1}{r^3} f(\theta) \quad , \quad (157)$$

where  $\alpha_2$  is the polarizability of the second molecule. The interaction potential is given by

$$U(r) = -\alpha_2 \epsilon_0 \frac{k^2 p_1^2}{r^6} f^2(\theta) \quad , \quad (158)$$

where again we have a  $1/r^6$  dependence in the absence of thermal averaging. Hence averaging will also give the same contribution since we only need to average over the angle  $\theta$ .

- **Induced dipole-induced dipole interactions**

A fluctuating electric field environment around each atom induces a fluctuation dipole moment that is proportional to the polarizability of the atom. This instantaneous dipole can then induce a dipole in a neighbouring atom, resulting in an attractive potential that also has a  $1/r^6$  dependence.

- **Short-range repulsive interaction**

As the atoms get too close, at some point there is a strong repulsion from overlapping electron clouds and Pauli's exclusion principle whereby filled electron shells of an atom cannot accommodate any more electrons.

- **Lennard-Jones potential**

A commonly used analytical form that lumps together all dipole-dipole interactions and includes both the attractive and the repulsive terms is the Lennard-Jones potential, where the repulsive term is approximated as having a  $1/r^{12}$  dependence

$$U_{LJ}(r) = 4\epsilon \left[ \left( \frac{r}{\sigma} \right)^{-12} - \left( \frac{r}{\sigma} \right)^{-6} \right] . \quad (159)$$

The atoms can be treated as spheres defined by a van der Waals radius that is a measure of how close another atoms can come before a strong, very short range, repulsive force kicks in.

Some typical Van der Waals radii of atoms are hydrogen 1.2 , oxygen 1.4 , nitrogen 1.6 , and carbon 2 .

## ■ Hydrogen bonds

A very important interaction responsible for the structure and properties of water, as well as the structure and properties of biological macromolecules, is the hydrogen bond. A hydrogen bond is an interaction between a proton donor group D-H and a proton acceptor atom A.

D-H is strongly polar, which means that the electron density is primarily around the electronegative atom (examples, F-H, O-H, N-H, S-H in order of decreasing polarity). The acceptor atom A is also strongly electronegative.

The hydrogen bond interaction is more than just an ionic or dipole-dipole interaction between the donor and the acceptor groups. The distance between the H and A in a hydrogen bond is less than the sum of their respective Van der Waals radii.

The strength of the hydrogen bonds in biological macromolecules ranges from  $2k_B T$  to  $5k_B T$ .

## ■ Hydrophobic interactions

Another very important interaction is the hydrophobic interaction. As the term hydrophobic suggests, this interaction is an effective interaction between two nonpolar molecules that tend to avoid water and, as a result, prefer to cluster around each other.

Unlike all the other interactions that we have studied so far and which are pairwise interactions between atoms or parts of molecules, the nature of the hydrophobic interaction is very different. It involves a considerable number of (water) molecules, and does not arise as a result of a direct force between the nonpolar molecules.

Nonpolar molecules are not good acceptors of the hydrogen bond. When a nonpolar molecule is placed in water, the hydrogen bonding network of water is disrupted. The water molecules therefore reorganize around the solute and make a sort of cage, similar to the structure of water in ice, in order to gain back the broken hydrogen bonds. This reorganization results in a considerable loss in the configurational entropy of water and therefore an increase in the free energy  $G$ . If there are two or more such nonpolar molecules, the configuration in which they are spatially together (clustered together) is preferred because now the hydrogen bonding network of water is disrupted in one (albeit bigger) pocket, rather than in

several small pockets. Therefore, the entropy of water is larger when the nonpolar molecules are clustered together, leading to a decrease in the free energy. At equilibrium, the configuration with the lower free energy and which has a higher Boltzmann probability, is the preferred configuration. Hydrophobic interactions have strengths of a few  $k_B T$  and are comparable in energy to hydrogen bonds.

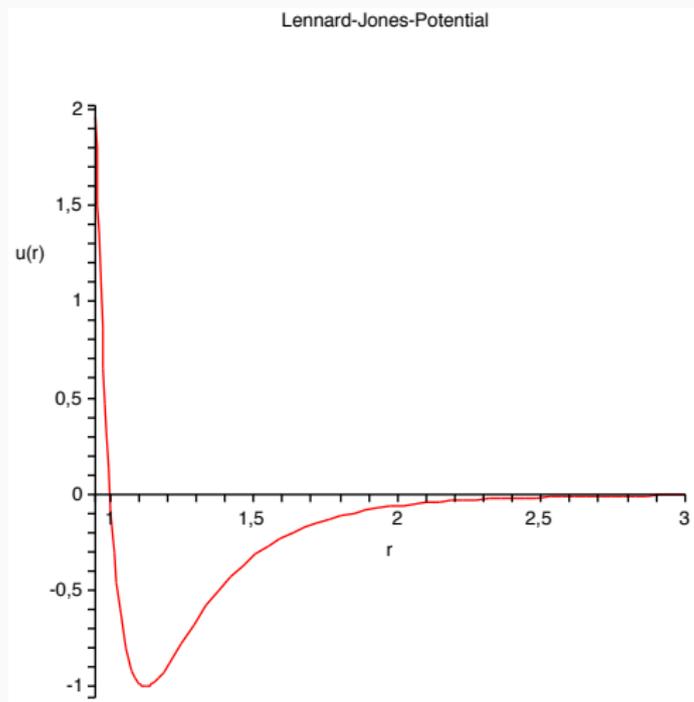


Figure 22: Lennard-Jones potential for the parameter  $\sigma = 1$ .

- Pruning
- Configurational Bias
- Bond Fluctuation Model

If we restrict the chain to a lattice then we need to consider random walks. More precisely we are interested in its trajectory, as this is the polymer chain contour. This idea was proposed by Kuhn. Of course, such a model can only capture ?universal? properties determined by long length scales. Indeed, the standard models used in the statistical mechanics of polymers are combinatorial structures such as random walks, self-avoiding walks, lattice polygons and lattice trees. While lattice models lack atomic details, they contain the fundamental microscopic attributes of polymers in that they show linear connectivity, chain flexibility, excluded volume- and sequence-dependent intra-chain interactions.

Note that here for simplicity we do not take excluded volume into account. This of course can easily be added.

---

## Algorithm 1 Reptation Algorithm

---

- 1: Assume that we have generated a random walk.
  - 2: Choose one of the end points at random and delete this point.
  - 3: Choose one the end points at random.
  - 4: Add the deleted point to the chosen end with a random direction.
-

```
1  a = 0;
   // Monte Carlo Loop
3  for(step=0; step<maxSteps; step++)
   {
5     save     = a;
     t        = selectElement(a);
7     a        = std::get<0>(t);
     b        = std::get<1>(t);
9     c        = std::get<2>(t);

11    position = selectMove(c,polyChain);
     if (acceptMove(position,data)) {
13        p = polyChain[b];
           data.erase(p);
15        polyChain[b] = position;
           data[position] = b;
17    } else {
           a = save;
19    }
   }
```

Code 1: Reptation Algorithm

Let  $W$  denote the set of self-avoiding walks of length  $N$  on a lattice  $\lambda$ . Further let  $G(\lambda)$  be the group of lattice symmetries. The pivot algorithm [15] takes a self-avoiding random walk and pivots the walk to generate a new walk from the set  $W$  such the sequence of generated walks yields a Markov chain which is aperiodic and irreducible with uniform stationary distribution  $\pi$ .

---

## Algorithm 2 Pivot Algorithm (Sokal)

---

- 1: Start with a self-avoiding walk  $\omega_0 \in W$ .
  - 2: Next choose an integer  $i$  uniformly from the set  $\{0, 1, 2, \dots, N - 1\}$ . The site connected with this index is the pivot site  $x = \omega_t(i)$ .
  - 3: Select a lattice symmetry  $g$  uniformly from the symmetry group  $G$ .
  - 4: Set  $\bar{\omega}(k) = \omega_t(k)$  for  $k \leq i$ , and  $\bar{\omega}(k) = g(\omega_t(k))$  for  $k > i$ .
  - 5: **if**  $\bar{\omega}$  is self-avoiding **then**
  - 6:      $\omega_{t+1} = \bar{\omega}$ .
  - 7: **else**
  - 8:     let  $\omega_{t+1} = \omega_t$ .
  - 9:     Goto 2. for the next generation  $t := t + 1$ .
  - 10: **end if**
-

The sequence  $\{\omega_t\}$  is aperiodic and irreducible with uniform stationary distribution  $\pi$ .  
The sequence further is reversible

$$\pi(\omega_i)P(\omega_i, \omega_j) = \pi(\omega_j)P(\omega_j, \omega_i) \quad . \quad (160)$$

Since  $\pi$  is uniform, we need to show that  $P$  is symmetric. Suppose there are  $m$  ways to move, with one pivot, from a self-avoiding walk  $\omega$  to another self-avoiding walk  $\bar{\omega}$ . For  $i = 1, 2, \dots, m$ , consider the pairs  $(x_i, g_i)$ . Each pair gives a transition, using the pivot algorithm from  $\omega$  to  $\bar{\omega}$ .

Thus,

$$P(\omega, \bar{\omega}) = \sum_{i=1}^m P(g = g_i) \cdot P(x = x_i) \quad . \quad (161)$$

Notice that the pairs  $(x_i, g_i^{-1})$ , for  $i = 1, 2, \dots, m$  give one-step transitions from  $\bar{\omega}$  and that  $P(g = g_i) = P(g = g_i^{-1})$  because  $g$  is chosen uniformly. Therefore

$$P(\omega, \bar{\omega}) = \sum_{i=1}^m P(g = g_i) \cdot P(x = x_i) = \sum_{i=1}^m P(g = g_i^{-1}) \cdot P(x = x_i) = P(\bar{\omega}, \omega) \quad . \quad (162)$$

A very simple but useful model for a polymer chain is the united atom model (c.f. Figure 23) In addition to harmonic chain forces which keep the bond lengths next to the equilibrium value, we model the fluctuation of bond angles, again by a quadratic potential. Between monomers which do not participate in mutual bond length or bond angle interactions, Lennard–Jones forces are acting, both to model an excluded volume effect and to hold the polymer system together. Note that we neglect any torsional potential in the present study. To be explicit, the Hamiltonian of the model is of the general form

$$\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2 + \mathcal{H}_3 \quad (163)$$

$$\mathcal{H}_1 = \sum_i \frac{1}{2} k_b (l_i - l_0)^2 \quad (164)$$

$$\mathcal{H}_2 = \sum_i \frac{1}{2} k_\theta (\cos \theta_i - \cos \theta_0)^2 \quad (165)$$

$$\mathcal{H}_3 = \sum_{i < j} u(r_{ij}) \quad (166)$$

where

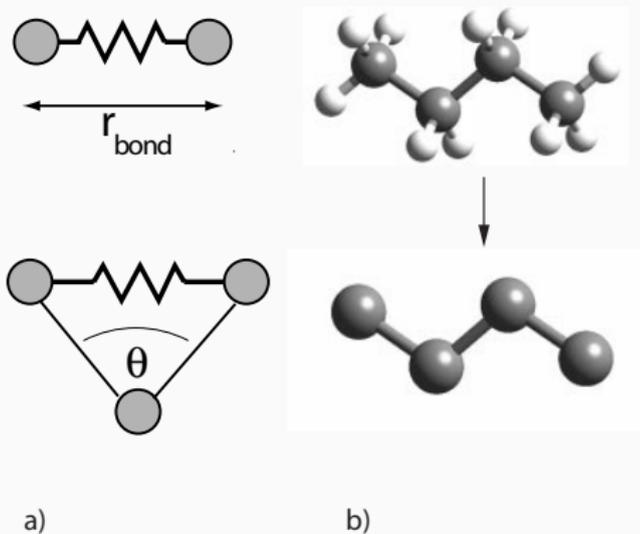
$$u(r_{ij}) = \begin{cases} u_{LJ}(r_{ij}) - u_{LJ}(r_c) - \frac{\partial}{\partial r_c} u_{LJ}(r_c)(r_{ij} - r_c) & r_{ij} < r_c \\ 0 & r_{ij} \geq r_c \end{cases} \quad (167)$$

and

$$u_{LJ}(r_{ij}) = 4\epsilon \sum_{i,j} \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right] \quad (168)$$

Note that the Lennard-Jones part of the potential is cut-off at  $1.5\sigma$  and analytically continued to zero.

The potential consists of the interaction along the chain with  $\mathcal{H}_1$  being bond length potential and  $\mathcal{H}_2$  being the bond angle potential. The interaction part between different chains, as well as from monomers along the chains more than three units apart is given by  $\mathcal{H}_3$ . We did not include the torsional potential part in the interaction purely for computational convenience.



**Figure 23:** The definition of the bond length and the bond angle potential

Excluded volume interactions are simulated by the WCA (Weeks-Chandler-Andersen) potential [16], which was designed to model excluded volume interactions by a short-range repulsive force. It has been used in several other MD studies on polymers [17]. The WCA potential is basically a truncated and shifted Lennard-Jones potential with the following functional form,

$$U_{\text{WCA}}(r) = \begin{cases} 4\epsilon \left( \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 + c_{\text{shift}} \right) & r < r_{\text{cut}} \\ 0 & r \geq r_{\text{cut}} \end{cases} \quad (169)$$

Here  $r_{\text{cut}} = \sqrt[6]{2}$  and  $c_{\text{shift}} = \frac{1}{4}$  are chosen such that the minimum of the potential is  $U_{\text{WCA}}(r_{\text{min}}) = 0$ , the attractive part of the Lennard-Jones interaction being cut off. The WCA potential has two parameters  $\epsilon$  and  $\sigma$ .  $\sigma$  defines the radius of the monomers' hard core.  $\epsilon$  controls the energy penalty of another monomer penetrating this hard core.

Simulating polymers with excluded volume interactions renders the use of a harmonic potential for the backbone potential as in eq. (??) impossible. A harmonic backbone potential in principle allows two adjacent beads to adopt a huge separation larger than their hard-core diameter  $\sigma$ , which would result in the possibility of bond crossings. To

circumvent this problem, it is convenient to use the finitely extensible nonlinear elastic model (FENE) potential.

$$U_{\text{FENE}}(r) = \begin{cases} -\frac{1}{2} k_{\text{FENE}} R_0^2 \log(1 - (r/R_0)^2) & r < R_0 \\ +\infty & r \geq R_0 \end{cases} \quad (170)$$

It is similar to the harmonic potential but grows to infinity at a predefined distance  $R_0$ . The pair potential between two beads (FENE + WCA) is displayed in Fig. ??.

The looping potential is chosen to be the same as in the original model, i.e. a Gaussian with Bernoulli-distributed random variables,

$$U_{\text{loops}} = \frac{1}{2} \sum_{\substack{i < j \\ |i-j| > 1}}^N \kappa_{ij} \| \mathbf{x}_i - \mathbf{x}_j \|^2 .$$

Here, the parameters are the looping probability  $\kappa$  and the interaction strength  $\kappa_{\text{loops}}$  (the  $\kappa_{ij}$  being either this value or zero).

The following parameters are chosen for the simulation runs:

$$R_0 = 1.6\sigma$$

$$\kappa_{loops} = 2.0$$

$$k_{FENE} = 10.0$$

$$\text{temperature } T = 1.0$$

$$\sigma = 1.0$$

$$\text{friction } \Gamma = 0.5$$

$$\epsilon = 20.0$$

$$\text{timestep } t = 0.006$$

Special care is required for the relation between  $R_0$  and  $\sigma$ . If  $R_0$  is too large, other parts of the chain may pass through the gap between two monomers. Setting  $R_0 = 1.6\sigma$  is a reasonable choice to prevent from such bond crossings [17].

Proteins (see Figure 27 for an example ) are the machines and building blocks of living cells. They are polymers of the 20 naturally occurring amino acids listed in table 3. The polymer size can vary from about 50 amino acids monomers with a molecular weight of 5,000 to very large containing 4,000 amino acids monomers with a molecular weight of larger than 513,000

Proteins have several functions in living systems:

- Structural (muscle, tendons, cell membranes, ...)
- Protection/defense (antibodies)
- Regulation (enzymes and hormones)
- Movement (assist other molecules into/out of cells)

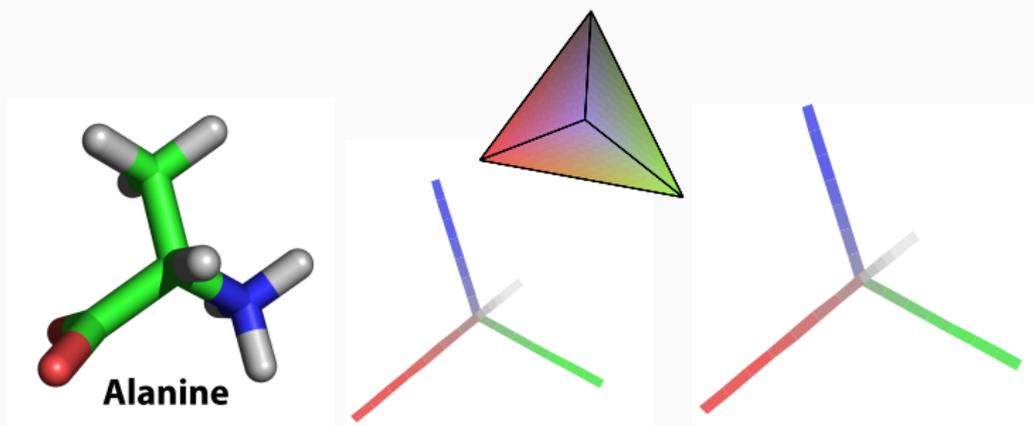
These functions of proteins are a direct consequence of their shape. Recall from Figure ?? that all amino acids have a COO and a NHHH part or a COOH carboxyl and NHH amino part. In addition, there is a side chain usually labeled R. The configuration of the side chain is called rotamer. This is due to the fact that the tetrahedral geometry stays the same and the main degree of freedom is rotation about the carbon bonds. In Figure 24 is shown the amino acid Aniline and its geometry.

**Table 3:** List of the 20 amino acids. The single letter code is used when comparing and aligning sequences of proteins

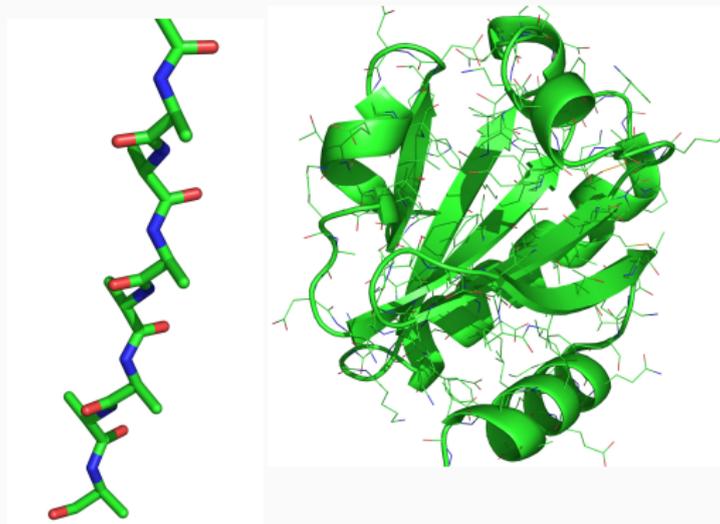


amino acids	3-letter code	single letter code
Alanine	Ala	A
Cysteine	Cys	C
Aspartic Acid	Asp	D
Glutamic Acid	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	Thr	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Tyr	Y

To form a protein, amino acids are bonded together in sequence and fold into a protein. Each protein has a unique three-dimensional structure. It was shown [18] that a protein in its natural environment folds into, i.e. vibrates around, a unique three dimensional structure, the *native conformation*, independent of the starting conformation.



**Figure 24:** The amino acid Alanine. Note that the bond directions for carbon are the same as from the centroid of a tetrahedron to the vertices.



**Figure 25:**  $\beta$ -sheet. The protein thioredoxin contains a five-stranded beta sheet comprised of three parallel strands and three antiparallel strands. The entire protein is shown as a cartoon with the beta strands (three parallel strands and three antiparallel strands) colored red and alpha helices colored yellow.

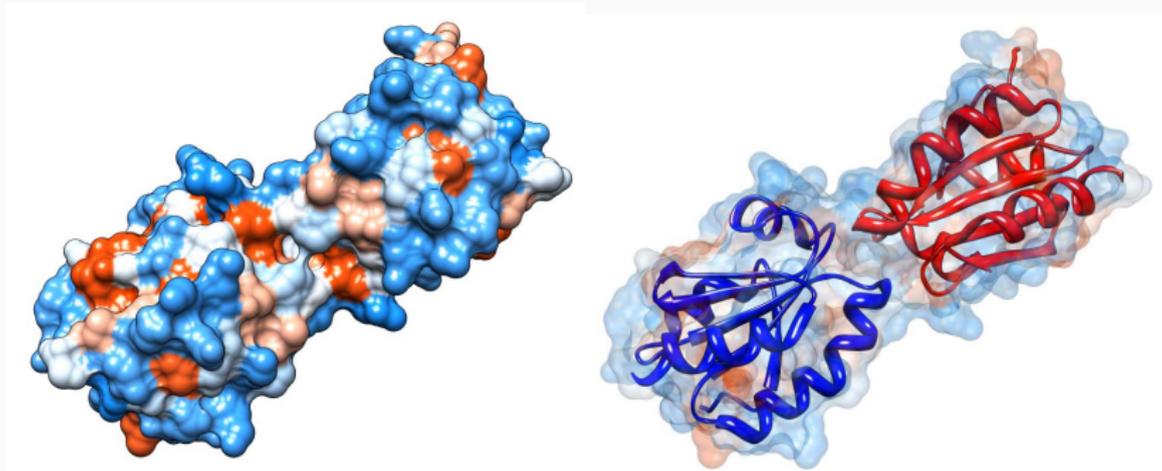


Figure 26: Protein 1f9m

There are four levels of architecture in proteins

- **Primary structure:** The sequence of peptide-bonded amino acids (as in the example: RSDAEPHYLPQLRKDILEVICKYVQIDPEMVTVQLEQKDGDISILEL-NVTLP EAEELK). This is determined by protein synthesis.
- **Secondary structure:** The regular, recurring arrangement in space of adjacent amino acid residues in a polypeptide chain. Two main types of secondary structures have been found in proteins, namely the  $\alpha$ -helices and  $\beta$ -sheets. The  $\alpha$ -helix-complex has already been studied in a previous section. In a  $\beta$ -sheet, two or more polypeptide chains run alongside of each other and are linked in a regular manner by hydrogen bonds between the main chain C=O and N-H groups. Hence, all hydrogen bonds in a  $\beta$ -sheet are between different segments of polypeptide. An example of one strand of a  $\beta$ -sheet is shown in Figure 25. A third type of secondary structure are loops. A loop is a section of the sequence that connects the other two kinds of secondary structures.
- **Tertiary structure:** The spatial arrangement among all amino acids in a polypeptide. The twisted shape is slightly flexible, and the chain folds upon itself.
- **Quaternary structure:** The spatial relationship of polypeptides or subunits. Several proteins interact and form complexes.

From the point of view of polymer physics the protein is simply a polymer consisting of a long chain of amino acid residues, i.e. a polypeptides.

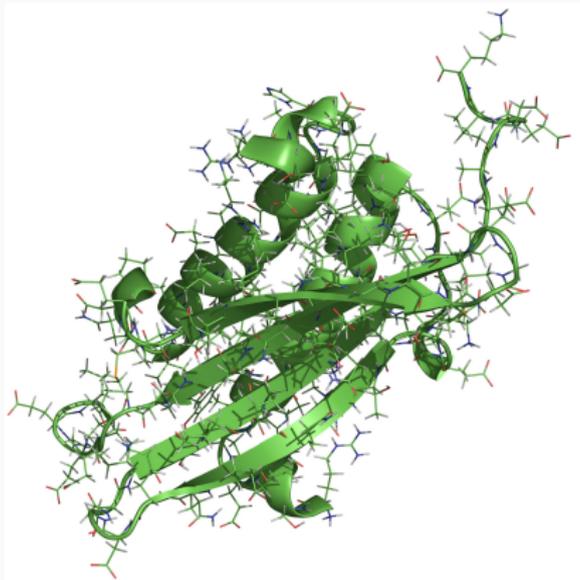


Figure 27: MinE protein showing  $\alpha$ -helices and  $\beta$ -sheets.

An important protein which exists in both monomeric or globular (G-actin) and polymeric or filamentary (F-actin) forms is actin. The filaments can form a network of entangled and crosslinked filaments and is the basis for the cytoskeletal network.

The long-standing question is: how do proteins fold? A protein folds due to the angles  $\phi$  and  $\psi$  between the carbon atom of a residue and the neighboring atoms, i.e. N and CO, in the peptide bond  $-N-C-(CO)-$ . These angles can assume only a few values independently of each other. Denaturants such as urea added to the system caused proteins that are folded in the native conformation to loose tertiary structure and revert to a random coiled state. After removal of the denaturants, the protein folds back into the native conformation.

The protein folding problem entails the mathematical prediction of (tertiary, 3-dimensional) protein structure given the (primary, linear) structure defined by the sequence of amino acids of the protein. With some exceptions, proteins fold spontaneously. What we want to have is a theoretical model that accurately predicts the folding and properties of the fold. The problem lies in the fact that a variety of globally different structures have very low energies, but within a few  $k_B T$  of each other. Hence, we would need a very good energy function for possible predictions and the ensuing dynamics are glassy as we have seen before.

What we would like to predict is for example

- the number of observable thermodynamic states
- the rate of folding

- the effect of specific mutations on the folding rate

Folding is an interesting problem because it involves mathematical modeling and numerical analysis. It is an extremely challenging task which has not been satisfactorily solved to date. Here we can only give a very brief introduction into some current methods.

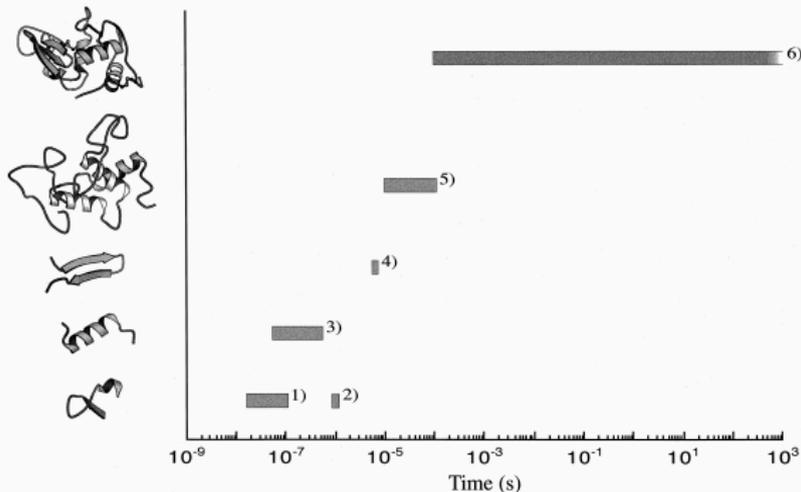
Basically, we need to distinguish between continuous and discrete models. Within continuous space models, a crucial problem is of course the large number of degrees of freedom. The configuration space is an  $n$  dimensional space, where  $n = 3 \times$  number of atoms in molecule. For example, the bacteriorhodopsin has 3576 atoms and hence we have 10728 coordinates! This results in the Levinthal's paradox [19]:

*The 3-D structure of a protein is determined by the dihedral angles. These angles have a few preferred values that correspond to the local minima of torsion energy around each rotation bond. We only have to consider about 10 conformations per AA in a polypeptide chain. This means that we have to examine at least as many as  $10^N$  conformations for a protein with  $N$  amino acids. Assuming that a protein can sample of the order of  $10^{14}$  structures per second, would take this protein about  $10^{26}$  seconds or  $10^{18}$  years to examine all the possible conformations. This is longer than the age of the universe.*

Indeed, the problem of finding the minimum energy configuration is NP-complete under a variety of models. Consequently, it is still impossible to determine the minimum energy structure for larger proteins based on the knowledge of only their sequence.

Since, for the foreseeable future it remains doubtful, that we find a satisfying solution for the molecular mechanics of the folding pathway, starting from the random coil conformation to the folded pattern that will emerge. The standard approach is to investigate models that are reduced in complexity. These can be discrete protein models on a lattice to reduce the conformational degrees of freedom or on the other

end of the spectrum the reduction to paths in a random energy landscape model. We have already touched on the energy landscape models and will here focus on molecular modeling and lattice models.



**Figure 28:** Time scales for the formation of structural elements in protein folding (Taken from O. Bieri and T. Kiefhaber, *Biol. Chem.* 80, 923-929 (1999))

There is an analogy between the spin glass problem and the folding problem. In the considered Hamiltonian there is frustration due to local free energy minima

$$H = - \sum_i (\epsilon_i \sum_j J_{i,i+1}) - \sum_{ij} K_{ij} \quad . \quad (171)$$

Here  $\epsilon_i$  is the energy of the  $i$ th residue,  $J_{i,i+1}$  are nearest neighbour interactions and  $K_{ij}$  are short range interactions between residues. To reach the global minimum simulated annealing is one of the few known algorithms assuring convergence to a global minimum. It is often used in combination with efficient steepest descent methods, such as conjugate gradients, as a way for avoiding getting trapped in local minima. However much hinges on the choice of the annealing schedule.

$$\lim_{T \rightarrow 0} \pi_i(T) = \lim_{T \rightarrow 0} \frac{\exp\{-E_i/kT\}}{\sum_j \exp\{-E_j/kT\}} \quad (172)$$

$$= \lim_{T \rightarrow 0} \frac{\exp\left\{\frac{E^* - E_i}{kT}\right\}}{\sum_j \exp\left\{\frac{E^* - E_j}{kT}\right\}} \quad (173)$$

where,

$$E^* = \min_i E_i = \text{Global min of energy} \quad (174)$$

Thus, the exponents are always either zero or negative. In the limit when  $T \rightarrow 0$  the terms with negative exponents disappear and we get

$$\lim_{T \rightarrow 0} \pi_i(T) = \begin{cases} \frac{1}{N^*} & \text{if } E_i = E^* \\ 0 & \text{otherwise} \end{cases} \quad (175)$$

where

$$N^* = |\{i : E_i = E^*\}| \quad (176)$$

Thus,  $\lim_{T \rightarrow 0} \pi_i(T)$  is uniformly distributed over the set of states of global minimum energy!

- 1: choose an initial configuration  $c$
  - 2: **for**  $T = T_0, T_1, \dots, T_m$  decreasing **do**
  - 3:   **for**  $mcs = 0; i < mcsmax$  **do**
  - 4:     choose a new trial configuration  $c_t$ ;
  - 5:     compute  $W = \exp\{E(c)/T - E(c_t)/T\}$ ;
  - 6:     generate a random number  $R$  between 0 and 1
  - 7:     if  $W > R$  , accept the trial configuration as the new configuration of the system;
  - 8:     set  $c = c_t$ ;
  - 9:   **end for**
  - 10: **end for**
-



Along the same line of thought is constructed the *simulated tempering* [?] and Metropolis-coupled Markov chain Monte Carlo [?]. The basic idea is to use  $m$  different chains with distributions  $\pi_i$  ( $i = 1, \dots, m$ ). From time to time we attempt to swap a state from chain  $i$  with one from chain  $j$ . For example we may choose [?]

$$\pi_i = \pi^{1/i}, \quad i = 1, \dots, m \quad (177)$$

i.e., for large  $m$  we would use a nearly uniform distribution. Suppose that we select chains  $i$  and  $j$ , at time  $t$  and we propose to swap  $y_i^t = x_j^t$  and  $y_j^t = x_i^t$ . With probability

$$\min \left\{ 1, \frac{\pi_i(y_i^t)\pi_j(y_j^t)}{\pi_i(x_i^t)\pi_j(x_j^t)} \right\} \quad (178)$$

the move is accepted.

If we are to use a lattice to hold a protein chain, then monomers are represented using uniform size and the bond length is considered uniform. Consider a  $N$ -amino acid polypeptide which is described by a polymer on a lattice in dimension  $D$  with a prescribed symmetry. For the moment, we shall use any general lattice  $\Lambda$  generated by the symmetry group  $G$  that consists only of translations. Each amino acid occupies one site on the lattice, and each peptide bond sits on a bond of the lattice. The folding of lattice proteins amounts to exploration of the ensemble of self-avoiding walk (SAW) configurations. What we are interested in is to count conformational states: How many conformational states are there for the  $N$ -monomer polymer that have a low energy (we will be more precise later).

If we are to enumerate the number of possible conformation one strategy is to use a Monte Carlo method to generate a Markov chain that will give the appropriate distribution at temperatures  $T < \infty$ . Starting with a given chain on our lattice we can change the conformation of the chain using three basic moves as depicted in Figure 29. The repeated application of the move set containing end bends, kink and crankshaft moves respects linear connectivity and is applied such that the condition of excluded volume is maintained. Furthermore this sampling must be ergodic and satisfy detailed balance.

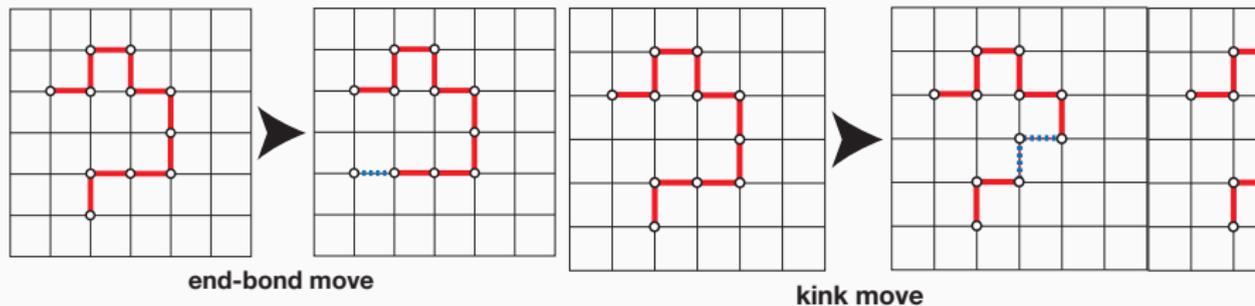


Figure 29: Possible move to change the conformation of a self-avoiding random walk (SAW)

This algorithm will give rise to conformations that can now be studied with respect to mappings of amino acid sequences yielding interaction energies. We will focus here on one model.

The hydrophobic-hydrophilic model [20] is a free energy model that models the belief that a major contribution to the free energy of the native conformation of a protein is due to interactions between hydrophobic amino acids that tend to form a core in the spatial structure shielded from the surrounding solvent by hydrophilic amino acids. The free energy of a conformation (see Figure 30) depends thus on the number of non-adjacent hydrophobic amino acids that occupy adjacent grid points in the lattice.

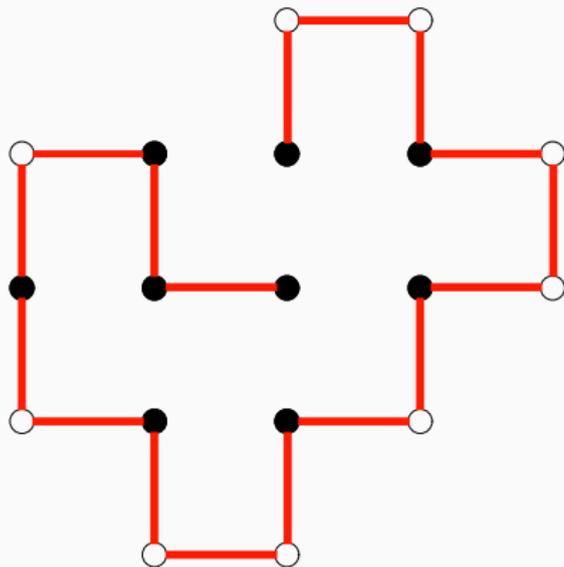
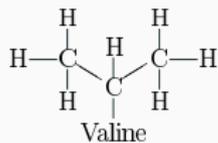
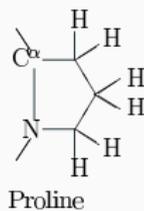
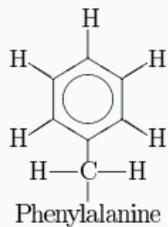
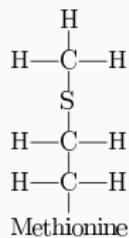
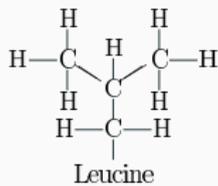
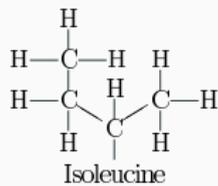
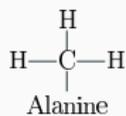


Figure 30: Conformation in the HP model. The black dots denote the hydrophobic acids.

In the *HP Model* the 20 amino acids reduced to a two-letter alphabet, H and P, where H is a hydrophobic amino acid, and P is a polar or hydrophilic amino acid (see Figure 31). The hydrophobic force is presumed to be dominant. For the interaction energy we take the values as shown in table 4.

**Table 4:** Energy in the HP Model

	H	P
H	-1	0
P	0	0



**Figure 31:** The hydrophobic amino acids

On a more abstract footing we start with sequence  $s$ , which is an element of  $\{0, 1\}^*$ , where 0 denotes  $P$  and 1 denotes  $H$ . Each conformation must be self-avoiding. We have connected neighbors:  $i$  and  $j$  are connected, if  $j = i + 1$  or  $j = i - 1$  independent of the conformation. Further, there are topological neighbors:  $i$  and  $j$  not connected and  $\|w(i) - w(j)\| = 1$ . The free energy of conformation is the negative number of HH-neighbors. Thus, we want to maximize HH contacts in hydrophobic core. A conformation is given by

$$w : (1 \dots |s|) \rightarrow Z^d \quad (179)$$

and the energy by

$$E = \sum_{1 \leq i < j \leq N} B_{i,j} \delta(\mathbf{R}_i, \mathbf{R}_j) \quad , \quad (180)$$

where  $\delta(\mathbf{R}_i, \mathbf{R}_j) = 1$  if  $\|\mathbf{R}_i - \mathbf{R}_j\| = 1$  and  $i \neq j \pm 1$  and  $B_{i,j} = -1$  if  $i$  and  $j$  are both H and 0 otherwise. Thus the energy is given by minus the number of topological HH contacts. On a more refined footing the values for the potential  $B$  are taken to be contact energies taken from tables derived from statistics on databases.

Rewriting this model slightly in the form

$$H = \sum_{i < j} \epsilon_{i,j} [\delta(|\mathbf{R}_i - \mathbf{R}_j| - \sigma) - \delta_{j-1,i}] \quad (181)$$

shows that we are dealing with a model that falls into the class of the random heteropolymer models (see 182). Here  $\sigma$  is the nearest neighbour distance. The interaction energy between monomers  $i$  and  $j$ ,  $\epsilon_{i,j}$ , can assume 3 values depending on the type of monomers bounded:  $\{H - H, H - P, P - P\}$ . These values are chosen to minimize the Hamiltonian when  $H$ -like amino acids are buried inside the protein and  $P$ -like amino acids are left on the surface.

One choice (see for example [21]) of the interaction energy (in arbitrary units) is:

$$\epsilon_{HH} = -2.3, \epsilon_{HP} = -1 \text{ and } \epsilon_{PP} = 0.$$

It was shown that the class of the HP-models is NP-complete [22, 23].

Let  $s$  be a sequence and  $c$  be a maximally compact self-avoiding structure. If the sequence has a unique lowest-energy state, or ground state, we say the sequence can *design the structure*. Figure 32 shows a conformation which is very highly designable.



The conformations of a macromolecule give rise to a complicated potential energy surface. The free energy landscape represents the configuration space of energy and entropy available to a macromolecule. Thus local minima or metastable states, basins of attraction and the saddle points separating them, appear. To understand and pave the way for the discussion on protein folding we qualitatively describe the situation. Let us divide the degrees of freedom of the polymer into two sets. The first contains the dihedral bond angles  $\theta_1, \phi_1, \dots, \theta_N, \phi_N$  along the backbone. Into the second set we place everything else from hydrogen bonding, torsion angle energies to rotations of individual side chains. The first set contains the conformational degrees of freedom and the second set the internal degrees of freedom (which allow typical free energy changes on the order of  $k_B T$ ). The energy landscape represents the  $2N$  conformational degrees of freedom. Each configuration is represented by a point on the conformation space such that similar conformations are nearby.

The roughness of an energy landscape may be quantified by the presence of structural hierarchy. Within a closed contour of constant elevation, there exist several closed contours of lower elevation, within each of which are more contours of lower elevation, etc. Landscapes characterized by a hierarchy of sub-valleys within valleys are said to be rugged; trivially hierarchical landscapes, in which each closed contour contains not multiple but a single closed contour of lower elevation, are called smooth.

The conformational energy landscape of a polymer is determined by self-avoidance. To pass to a conformationally near but topologically distant conformation, the polymer must swell and recollapse, overcoming a large energy barrier. A hierarchy of valleys makes the ground state conformation only marginally lower in energy than quasi-degenerate local minima, which act as energetic traps. Further, if the polymer is composed of several different monomers, as for example in a protein, additional constraints arise due to frustration, the inability of chain segments to cooperatively align. Together, these suggest that the conformational landscape of a typical sequence is typically rugged.

We start off with the random energy model (REM) [24–27] for the density of states of a heteropolymer. This model aims to describe differences between the energy spectrum of randomly generated sequences, which are unable to fold, and the energy spectrum of the particular set of sequences that fold to a unique, stable, native conformation.

The random heteropolymer (RHP) can be described by the Hamiltonian

$$H_{\text{RHP}}(\{s_i\}, \{\mathbf{R}_i\}, ) = \sum_{i < j} \epsilon(s_i, s_j) \Delta(|\mathbf{R}_i - \mathbf{R}_j|) \quad . \quad (182)$$

Here  $i$  counts the monomers (residues) along the chain,  $s_i \in \{1, \dots, p\}$  is the species of monomer and  $\mathbf{R}_i$  is the position of monomer  $i$  (see figure 33). Thus  $p = 1$  results in a homopolymer,  $p = 2$  into a co-polymer. The special limit  $p \rightarrow \infty$  stands for a continuous distribution of interaction energies.

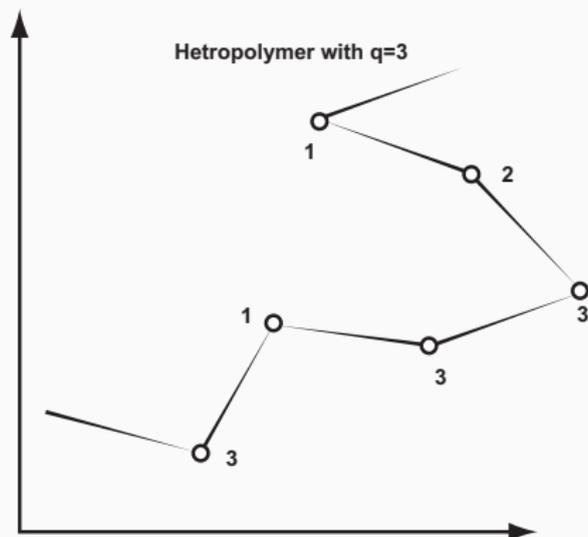


Figure 33: Part of a heteropolymer conformation with  $q = 3$ .

It is understood that the connectivity requirement is met.

$\Delta(r)$  includes the excluded volume effects and is assumed to vanish at larger distances.

We now consider a macroscopic system with energy  $E$  given by the sum of many microscopic energetic terms  $\epsilon$ . If individual terms are considered to be statistically independent we can use the central limit theorem to obtain the energy distribution of the system  $P(E)$

$$P(E) = \frac{1}{\sqrt{2\pi d}} \exp\left(-\frac{(E - \langle E \rangle)^2}{2d^2}\right), \quad (183)$$

where  $d^2 = zNs^2$  with  $z$  being the mean number of contacts per monomer and  $s^2$  being the energetic variance of the individual contacts.  $\langle E \rangle = Nz\bar{\epsilon}$  is the average energy. Note that the basic assumption that contact energies are uncorrelated effectively ignores chain connectivity!

The density of microstates is obtained by multiplication of  $P(E)$  by the total number of microstates. Here, with energy given by the sum of individual contacts, the density of states is given by

$$n(E) = \gamma^N P(E) = \gamma^N \frac{1}{s\sqrt{2\pi Nz}} \exp\left(-\frac{(E - Nz\bar{\epsilon})^2}{2Nzs^2}\right), \quad (184)$$

where  $\gamma$  is the average number of conformations per monomer taking excluded volume into consideration. This implies that the large majority of conformations will have energy between  $Nz\bar{\epsilon} - s\sqrt{Nz}$  and  $Nz\bar{\epsilon} + s\sqrt{Nz}$ . Further, there exists a critical energy value  $E_g$ , obtained from the condition  $n(E_g) = 1$

$$E_g = Nz\bar{\epsilon} - Ns\sqrt{2z \ln \gamma} = \langle E \rangle - \sigma_E \sqrt{2S_0} \quad (185)$$

terms proportional to  $\ln N$  were discarded from the above expression as they are small compared to terms proportional to  $N$  for large  $N$  and  $S_0 = N \ln \gamma$  is the conformational entropy of the chain. Below this point the gaps in the spectrum are too large for the chain to change its conformation in the process of thermal fluctuation.

Using eq. 184 we have that the entropy is given by

$$S(E) = \ln(n(E)) = N \ln \gamma - \ln(s\sqrt{2\pi Nz}) - \frac{(E - Nz\bar{\epsilon})^2}{2Nzs^2} \quad (186)$$

and

$$T(E) = \frac{1}{dS/dE} = -\frac{Nzs^2}{E - Nz\bar{\epsilon}} \quad (187)$$

The corresponding critical temperature is given by

$$T_g = T(E_g) = -\frac{Nzs^2}{E_c - Nz\bar{\epsilon}} = s\sqrt{\frac{Nz}{2N \ln \gamma}} = \frac{s}{\sqrt{2S_0}} \quad (188)$$

Below  $T_g$  the chain will behave like a glass, *freezing* in one of many local minima in its energy surface and never reaching thermal equilibrium. As the energy of the native conformation (the lowest energy in the spectrum),  $E_N$ , for a sequence taken at random is expected to be close to  $E_g$ , it follows that the native conformation of such a sequence would be thermodynamically stable only at a temperature close to  $T_g$  and so would not be able to fold in a reasonable time.

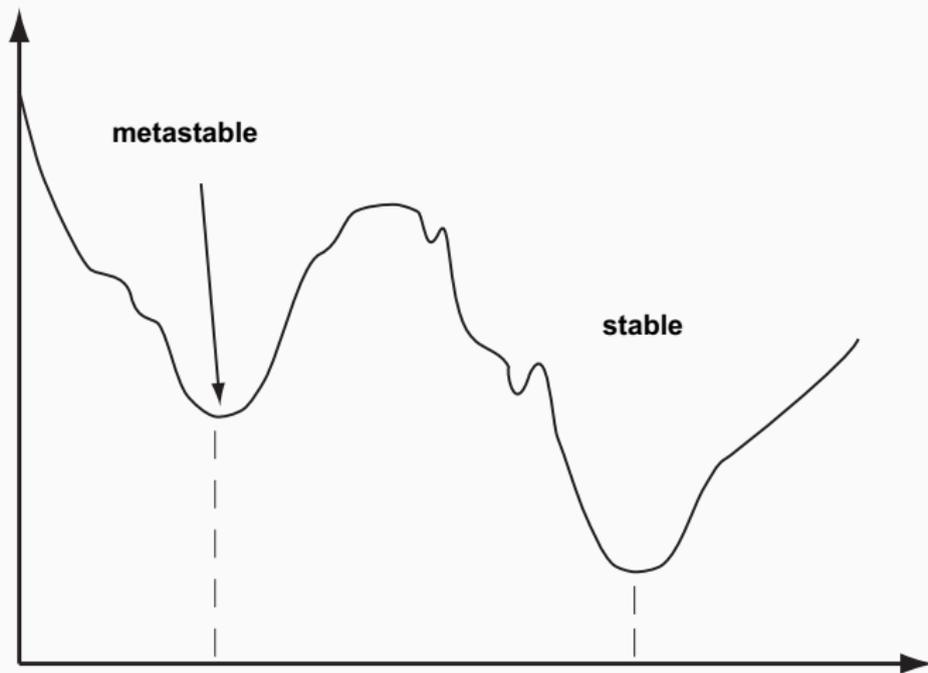
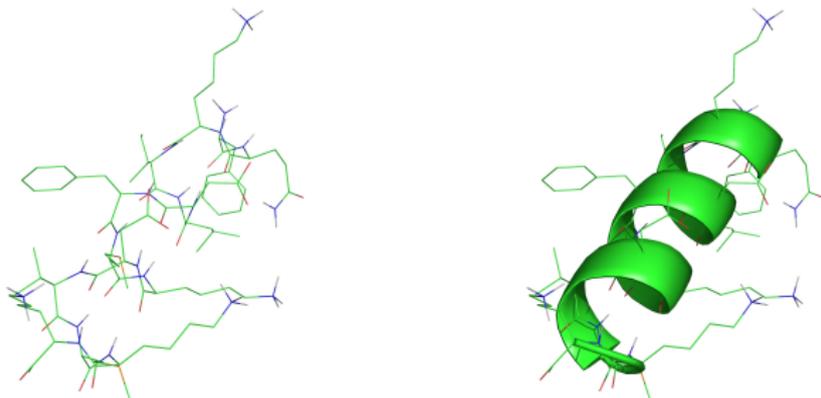


Figure 34: Energy landscape

The  $\alpha$ -helix is the most abundant helical conformation found in globular proteins. In the  $\alpha$ -helix the polypeptide folds by twisting into a right-handed screw, so that all the amino acids can form hydrogen bonds with each other. The helix has maximal intra-chain hydrogen bonding. This high amount of hydrogen bonding stabilizes the structure so that it forms a very strong rod-like structure. The amino group of each AA residue is hydrogen bonded to the carboxyl group of the 4th following AA residue, which is on an adjacent turn of the helix.



**Figure 35:** Helical structure found in polypeptides. To better identify the helical structure the right picture shows a cartoon of the helix.

Along the axis of the helix, it rises 0.15 nm per AA residue, and there are 3.6 residues/turn of the helix. This means, that AA residues spaced 4 apart in the linear chain are quite close to one another in the  $\alpha$ -helix. The screw-sense of any helix can be RH or LH, but the  $\alpha$ -helix found in proteins is always RH. The average length of an alpha helix is about 10 residues.

What we want to consider now is that upon increasing the temperature, the helix structure goes over into a random coil structure [? 28–30].

To describe the macromolecule in terms of helical and non-helical parts, we denote by  $h$  a helical monomer and by  $c$  a coil monomer (see later for the analogy with the Ising model [31]). A conformation is then characterised by a sequence of  $h$  and  $c$ , which we denote by  $\{h, c\}$ . An example for such a sequence is

$$\dots h h c c c h c \dots \quad (189)$$

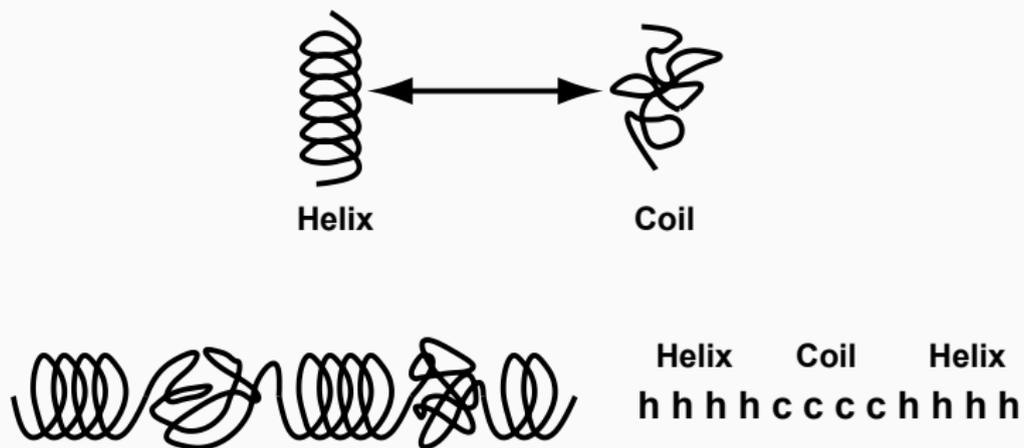


Figure 36: Mapping of the helix-coil transition onto a sequence of symbols

Since there are  $N$  monomers, we have  $2^N$  states. To be able to write down a partition function, we assume that the energies of the  $h$ - and the  $c$ -sequence are independent and that they only depend on the length of the corresponding sequence. Then we can write down individual statistical weights

$$u_i = \exp\{-\beta E_i(c)\} \quad (190)$$

for the  $c$ -sequence with  $i$  coil-like connected monomers. Likewise for the helical sequence

$$v_i = \exp\{-\beta E_i(h)\} \quad (191)$$

Here we have implicitly assumed that the energy is independent of the position within the chain and independent of the neighbouring sequences! Also self-avoidance has been ignored, since we do not take into account that monomers may be linearly located far apart but may get in contact with each spatially. Given all these assumptions we write down the partition function

$$Z_N = \sum_{\{h,c\}} e^{-\beta E(\{h,c\})} \quad (192)$$

$$= \sum_{i,j} \prod_{i,j} u_i v_j \quad . \quad (193)$$

Everything hinges now on the distribution of the  $h$ - and the  $c$ -sequences. Let us write for the sequence  $\{h, c\}$ :

$$i_0, j_1, i_1, \dots, j_M, i_M, j_0 \quad , \quad (194)$$

where  $i$  denotes the length of the  $c$ -sequence and  $j$  the length of the  $h$ -sequence. All  $2M$  inner sequences contain at least one unit

$$M \leq \lfloor N/2 \rfloor \quad (195)$$

with the constraint

$$\sum_{k=0}^M (i_k + j_k) = N \quad . \quad (196)$$

Hence we can write

$$Z_N = \sum_{M=0}^{\lfloor N/2 \rfloor} \sum_{\{i_k, j_k\}} \prod_{k=0}^M u_{i_k} v_{j_k} \quad . \quad (197)$$

From the preceding section it is clear, that if we consider very long chains ( $N \rightarrow \infty$ ) then the free-energy will be proportional to  $N$ , i.e., chain end effects will not play any role

$$Z_N \approx q_{\text{eff}}^N \quad \text{for} \quad N \gg 1 \quad , \quad (198)$$

where  $q_{\text{eff}}$  is the average contribution per monomer to the free-energy.

Let us now look at the generating function

$$\Gamma(x) = \sum_{N=0}^{\infty} Z_N x^{-N} \quad . \quad (199)$$

This series converges for  $x > q_{\text{eff}}$  and diverges for  $x \rightarrow q_{\text{eff}}$

$$\Gamma(x) < \infty \quad x > q_{\text{eff}} \quad (200)$$

$$1/\Gamma(x) = 0 \quad x = q_{\text{eff}} \quad . \quad (201)$$

Hence the partition function is the largest root of eq 201. So, let us look at the  $\Gamma$  in more detail

$$\Gamma(x) = \sum_{N=0}^{\infty} x^{-N} \sum_{M=0}^{\lfloor N/2 \rfloor} \sum_{\{i_k, j_k\}} \prod_{k=0}^M u_{i_k} v_{j_k} \quad (202)$$

$$= \sum_{M=0}^{\infty} \sum_{N=2M}^{\infty} \sum_{\{i_k, j_k\}} \prod_{k=0}^M u_{i_k} x^{-i_k} v_{j_k} x^{-j_k} \quad (203)$$

$$= \sum_{M=0}^{\infty} \sum_{i_0=0}^{\infty} \frac{u_{i_0}}{x^{i_0}} \sum_{j_0=0}^{\infty} \frac{v_{j_0}}{x^{j_0}} \prod_{k=1}^M \sum_{i_k=0}^{\infty} \frac{u_{i_k}}{x^{i_k}} \sum_{j_k=0}^{\infty} \frac{v_{j_k}}{x^{j_k}} \quad (204)$$

The sums over  $i_k$  and  $j_k$  do not depend on  $k$  any more. Only the ends can have a different weight. For  $k \geq 1$  we can define

$$U(x) \equiv \sum_{i=1}^{\infty} u_i x^{-i} \quad (205)$$

$$V(x) \equiv \sum_{j=1}^{\infty} v_j x^{-j} \quad (206)$$

which converge in  $q_{\text{eff}} < x < \infty$ , since  $\Gamma(x)$  converges. With this we have

$$\Gamma(x) = U_0 V_0 \sum_{k=0}^{\infty} (UV)^k \quad (207)$$

$$= U_0 V_0 \frac{1}{1 - UV} \quad (208)$$

$\Gamma(x)$ ,  $U(x)$  and  $V(x)$  are positive and monotone decreasing functions of  $x$  since the statistical weights are positive and real. It follows that  $1/\Gamma(x)$  is a continuous and monotonically decreasing function in  $q_{\text{eff}} < x < \infty$ , since  $\Gamma(x)$  and  $1/\Gamma(x) = 0$  for  $x = q_{\text{eff}}$ . Since

$$U_0 V_0|_{x=q_{\text{eff}}} \neq 0 \quad (209)$$

we have

$$U(q_{\text{eff}})V(q_{\text{eff}}) = 1 \quad (210)$$

In a chain composed of six units only four contribute with hydrogen bonds to the helical structure. In general, we have that for  $j$  consecutive helical states only  $(j - 2)$  are formed by hydrogen bonds. Hence we need three states in our model:

- a coil-like state
- a helical state with hydrogen bond.
- a helical state without hydrogen bond,

Corresponding to these three states we need statistical weights

$$\text{coil} \quad - \quad u \quad (211)$$

$$\text{h with h - bond} \quad - \quad w \quad (212)$$

$$\text{h without h - bond} \quad - \quad v \quad (213)$$

If we take as a reference the coil state then we have the weights

$$\text{coil} \quad - \quad u/u = 1 \quad (214)$$

$$\text{helix with h - bond} \quad - \quad w/u = s \quad (215)$$

$$\text{helix without h - bond} \quad - \quad v/u = \sigma^{1/2} \quad (216)$$

For the sequences of  $h$  and  $c$  we get

$$\begin{array}{ll} c - \text{sequence} & u_i = u^i & 1 \\ h - \text{with h - bond} & v_j = v^2 w^{j-2} & v_j = \sigma s^{j-2} \\ h - \text{without h - bond} & v_1 = v & v_1 = \sigma^{1/2} \end{array} \quad (217)$$

From the experimental point of view one can determine the relative number of unbroken hydrogen bonds  $\theta$  (which is proportional to the number of  $w$  statistical weights).

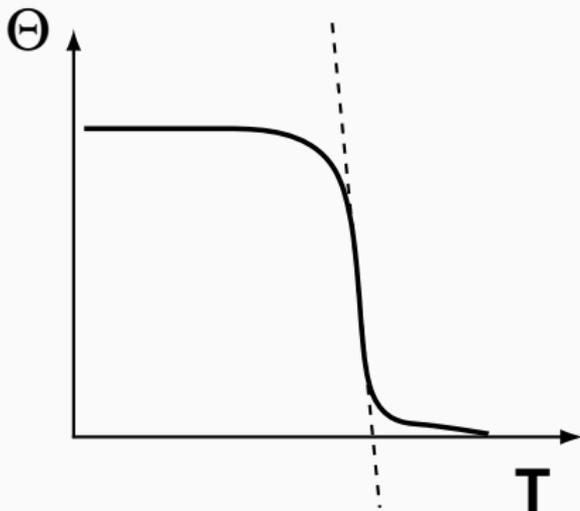


Figure 37: Dependence of the order parameter on the temperature for the helix-coil transition

With the above defined statistical weights and using eq 193 we have

$$Z_N = \sum_{ij} \prod_{ij} u_i v_j = \sum_{ij} \prod_{ij} \sigma s^{j-2} \quad (218)$$

We obtain  $\theta$  by taking the derivative with respect to  $s$

$$\theta = \frac{1}{N-2} \frac{\partial \ln Z_N}{\partial \ln s} \quad (219)$$

Since  $Z_N \propto q_{\text{eff}}^N$  for  $N \gg 1$  we find

$$\theta = \frac{1}{N} \frac{\partial \ln q_{\text{eff}}^N}{\partial \ln s} = \frac{s}{q_{\text{eff}}} \frac{\partial q_{\text{eff}}}{\partial s} \quad (220)$$

We determine  $q_{\text{eff}}$  from eq 210

$$q_{\text{eff}}^3 - q_{\text{eff}}^2(u+w) + q_{\text{eff}}(wu - uv) + uvw + v^2 u = 0 \quad (221)$$

with the solution

$$q_{\text{eff}} = \frac{1}{2} \left\{ w + u + \sqrt{(w - u)^2 + 4uv^2/w} \right\} \quad (222)$$

and for  $\theta$

$$\theta = \frac{1}{2} \left\{ 1 + \frac{s - 1}{\sqrt{(s - 1)^2 + 4\sigma s}} \right\} . \quad (223)$$

In figure 37 is shown the qualitative result for the order parameter  $\theta$ . A comparison to the experimental findings is shown in figure 38.

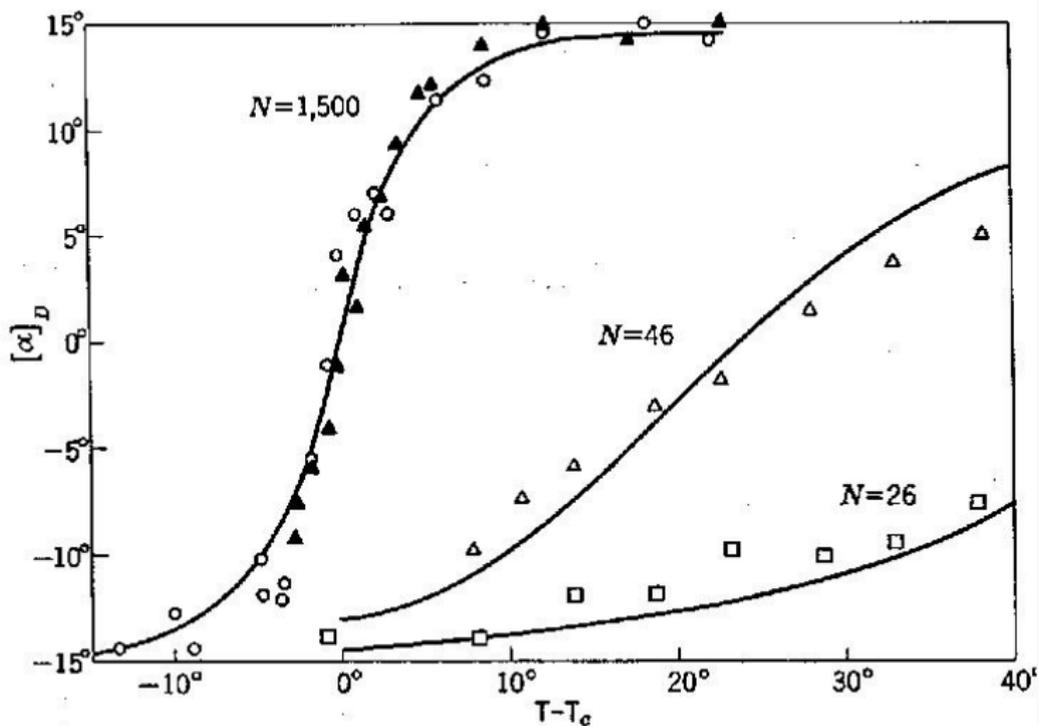


Figure 38: Experimental results for the helix-coil transition

We can make contact with the Ising model by setting

$$\sigma = e^{-4J/k_B T} \quad (224)$$

$$s = e^{2H/k_B T} \quad (225)$$

to find

$$\theta = \frac{1}{2} \left\{ 1 + \frac{\sinh(H/k_B T)}{\sqrt{\sinh^2(H/k_B T) + e^{-4J/k_B T}}} \right\} . \quad (226)$$

Hence the average magnetization per spin  $\langle m \rangle$  can be looked upon as the helical fraction

$$\langle m \rangle = 2\theta - 1 = \frac{\sinh(H/k_B T)}{\sqrt{\sinh^2(H/k_B T) + e^{-4J/k_B T}}} . \quad (227)$$

In this context the Ising model appears as a special case of the  $\alpha$ -helix model.

DNA melting refers to the dissociation of the two strands of the double helix by an increase of temperature. The melting is thus a highly cooperative thermal disruption of the hydrogen bonds between complementary bases in the double helix. At the equilibrium melting temperature half of the bonds are disrupted. Dissociation can occur also through a change of pH.

The melting or denaturation of DNA is a thermodynamic reversible phase transition. The order of the transition is still debated due to the effect to the entropy of loops embedded in the chain. Existing experimental studies of the thermal denaturation of DNA yield sharp steps in the melting curve suggesting, that the melting transition is first order. Here we present the Poland-Scheraga-model [?] and the zipper-model [?].

The Poland-Scheraga-model considers the DNA molecule as composed of an alternating sequence of bound and denaturated states as depicted in figure 39. Consider two strands, made of up monomers, each representing one persistence length of a single strand. Typically a bound state is energetically favored over an unbound one, while a denaturated segment or loop is entropically favored over a bound one.



Figure 39: Poland-Scheraga-model of the DNA-melting transition

Within the Poland-Scheraga-model the segments that compose the chain are assumed to be non-interacting with one another, i.e. excluded volume effects are not taken into account. This assumption considerably simplifies the theoretical treatment and enables one to calculate the resulting free energy.

Analogous to the  $\alpha$ -helix model we can define statistical weights

coil sequence in a loop	$\delta(i)\sigma$	
coil sequence at the end of the chain	1	(228)
helix sequence	$s^j$	

The statistical weight of a bound sequence of length  $k$  is

$$s^j = \exp(-jE/k_B T) \quad . \quad (229)$$

On the other hand the statistical weight of a denaturated sequence of length  $i$  is given by the change in entropy due to the added configurations arising from a loop of length  $2i$ . For large  $i$  the free energy will be proportional to the entropy of a closed loop  $S(i)$

$$S(i)/k_B = \ln \delta(i) = ai - c \ln i + b \quad , \quad (230)$$

which is the typical form of the entropy for polymer chains with two free ends and excluded volume effects. It follows

$$\delta(i) = e^{S(i)/k_B} \propto i^{-c} \approx \kappa^i / i^c \quad , \quad (231)$$

where  $s$  is a non-universal constant, and the exponent  $c$  is determined by the properties of the loop configurations. For simplicity, we set  $a = 1$ .

The model is most easily studied within the grand canonical ensemble where the total chain length  $N$  is allowed to fluctuate. The grand canonical partition function is given by

$$Z = \sum_{N=0}^{\infty} G(N)x^N = \frac{V_0(x)U_N(x)}{1 - U(x)V(x)} \quad , \quad (232)$$

with

$$U(x) = \sum_{i=1}^{\infty} \frac{\kappa^i}{i^c} x^i, \quad V(x) = \sum_{j=1}^{\infty} s^j x^j \quad (233)$$

and  $V_0(x) = 1 + V(x)$ ,  $U_L(x) = 1 + U(x)$ . In the thermodynamic limit,  $L \rightarrow \infty$

$$\ln Z \simeq N \ln x_1. \quad (234)$$

Here  $x_1$  is the value of the fugacity in the limit  $\langle N \rangle \rightarrow \infty$ . This is the lowest value of the fugacity for which the partition function diverges, i.e., for which

$$U(x_1)V(x_1) = 1 \quad . \quad (235)$$

It is thus clear that the nature of the denaturation transition is determined by the dependence of  $x_1$  on  $s$ . The transition takes place when  $x_1$  reaches  $1/\kappa$ . Its nature is determined by the behaviour of  $U(x)$  in the vicinity of  $x_c$ . This is controlled in turn by the value of the exponent  $c$ .

We can again define an order parameter  $\theta$  to be

$$\theta = \frac{1}{1 + \frac{\sigma s}{x_1} \sum_{i=1}^{\infty} x_1^{-i} i^{1-c}} . \quad (236)$$

From the above we get the determining equation for  $x_1$

$$\sum_{i=1}^{\infty} x_1^{-i} i^{-c} = \frac{x_1 - s}{\sigma s} . \quad (237)$$

Since  $x_1(s) \geq 1$  we have a lower bound  $x_1(s_c) = 1$  with

$$s_c = \frac{1}{1 + \sigma \zeta(c)} , \quad (238)$$

with  $\zeta(c)$  being the Riemann Zeta-function.

We can distinguish three regimes:

- 1 For  $c \leq 1$ ,  $U(x_c)$  diverges, so that  $x_1$  is an analytic function of  $s$  and no phase transition takes place.

- 2 For  $1 < c \leq 2$ ,  $U(x_c)$  converges but its derivative diverges at  $x_1 = x_c$ . Thus the transition is continuous.
- 3 For  $c > 2$ ,  $U(z)$  and its derivative converge at  $x_1 = x_c$  and the transition is first order.

The value of the exponent  $c$  can be obtained by enumerating random walks, which return to the origin, so that  $c = d\nu$ . For ideal random walks this yields  $c = d/2$ . Thus there is no transition at  $d \leq 2$ , a continuous transition for  $2 < d \leq 4$  and a first order transition only for  $d > 4$ .

On the other hand, for self-avoiding random walks the excluded volume interaction modifies the exponent to  $c = 3/2$  for  $d = 2$  and  $c \simeq 9/5$  for  $d = 3$ . The transition is thus sharper, but still continuous, in three dimensions.

## Zipper Model of DNA Melting

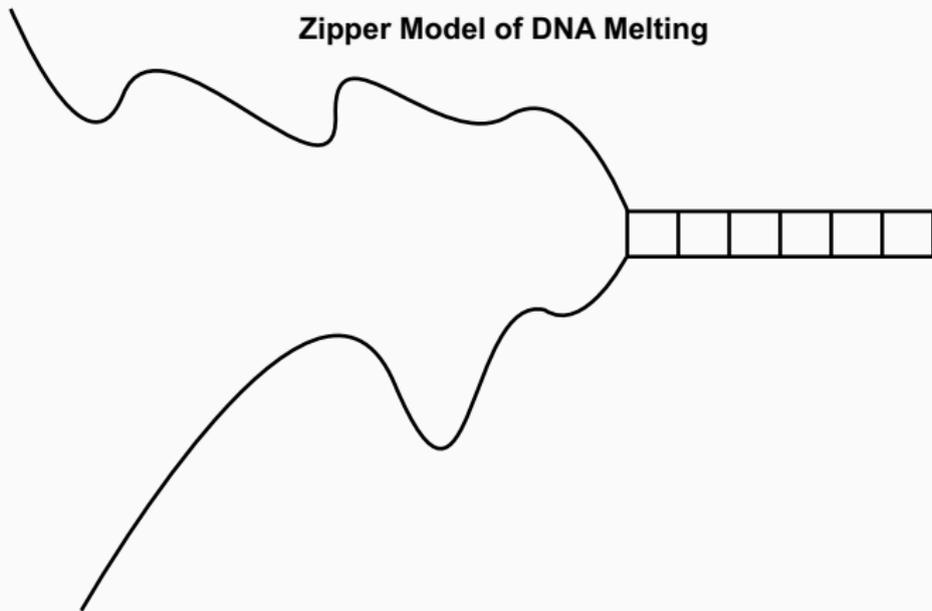


Figure 40: Kittel zipper model of the DNA-melting transition

Kittels zipper-model [?] describes the breaking up of the double helix starting from the end. The zipper is comprised of  $N$  parallel bonds. The bonds can only break up successively starting from one end of the chain (see figure 40). In this model it is assumed to be impossible to break up a bond anywhere with the chain except for the one right next to the one that broke last.

If the bonds  $1, \dots, p$  are broken then the energy to break the  $p + 1$  bond is  $\epsilon$ . The last element of the chain is considered unbreakable.

We assume, that an open bond can take on  $G$  orientations (due to rotational degrees of freedom,  $G \approx 10^4$ ). To break up the first  $p$  bonds we need an energy  $p\epsilon$ , and this will give a contribution of

$$G^p e^{-p\epsilon/k_B T} \quad (239)$$

to the partition function. Thus

$$Z_N = \sum_{p=0}^N G^p e^{-p\epsilon/k_B T} = \frac{1 - x^N}{1 - x} \quad , \quad (240)$$

with  $x = Ge^{-\epsilon/k_B T}$ . We define as the order parameter the average number of open or broken bonds

$$\langle \theta \rangle = \frac{\sum_p p x^p}{\sum x^p} = x \frac{d}{dx} \ln Z_N \quad (241)$$

$$= \frac{N x^N}{x^N - 1} - \frac{x}{x - 1} \quad (242)$$

which is shown in figure 41. We can expand the order parameter in the neighbourhood of the critical point  $x_c = 1$  using

$$\epsilon \equiv |x - x_c| \propto |T - T_c| \ll 1 \quad . \quad (243)$$

With this

$$\langle \theta \rangle = G \frac{d\epsilon}{dG} \frac{d}{d\epsilon} \ln Z_N \quad (244)$$

$$= \frac{1}{2} N \left( 1 + \frac{1}{6} N \epsilon - \frac{1}{360} N^3 \epsilon^3 + \dots \right) \quad . \quad (245)$$

For  $T = T_c$  we have

$$\frac{1}{N} \frac{d\langle\theta\rangle}{d\epsilon} = \frac{1}{12} N - \frac{1}{240} N^3 \epsilon^2 \quad (246)$$

for  $N \gg 1$  and  $\epsilon \ll 1$ . At the critical point the order parameter reaches a value

$$\frac{\langle\theta\rangle}{N} = \frac{1}{2} \quad (247)$$

The slope diverges in the thermodynamics limit. The critical temperature is given by

$$T_c = \frac{\epsilon}{k_B} \ln G \quad (248)$$

For  $G > 1$  we find that the critical temperature is finite.

## Order parameter for the Kittel zipper model

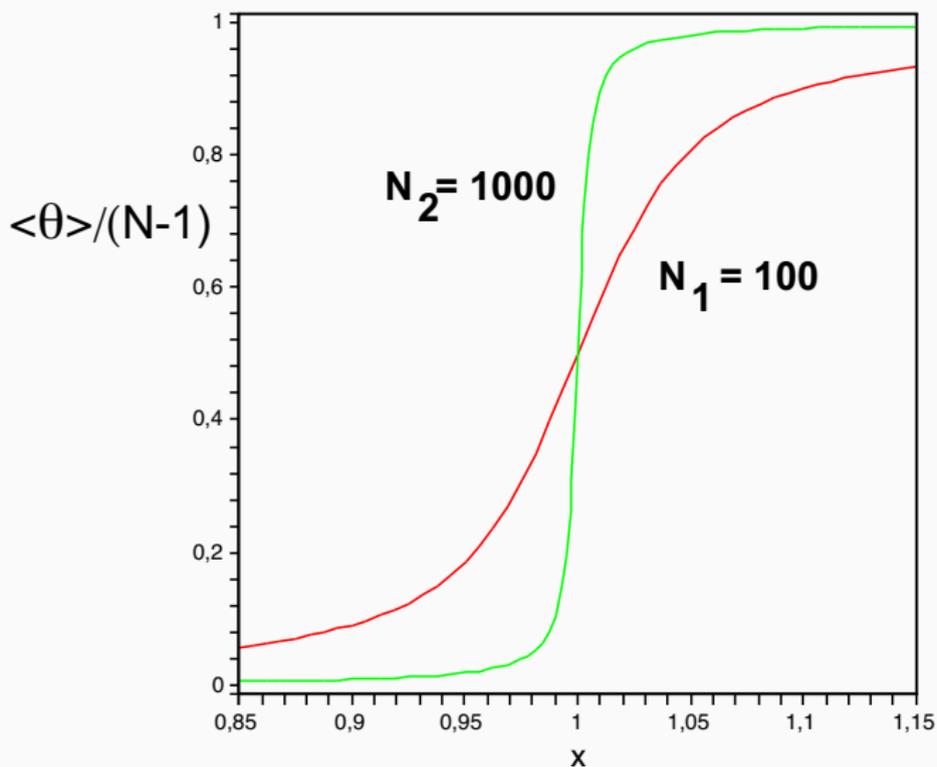


Figure 41: Kittel zipper-model of the DNA-melting transition

A protein aggregate together with its wrapped DNA comprises a nucleosome core particle with a radius of about 5nm and a height of about 6nm. With its linker DNA it is the fundamental chromatin repeating unit. It carries a large electrostatic charge [32]. Whereas the structure of the core particle has been resolved up to high atomic resolution [33], there is still considerable controversy about the nature of the higher-order structures to which they give rise. When stretched the chromatin string appears to look like beads-on-a-string in electron micrographs.

The beads-on-a-string structure can be seen clearly when chromatin is exposed to very low salt concentrations, and is known as the 10-nm-fiber, since the diameter of the core particle is about 10nm. With increasing salt concentration, i.e. heading towards physiological conditions ( $c \approx 100\text{mM}$ ), this fiber appears to thicken, attaining a diameter of 30nm. The absence of the extra linker histones (H1 or H5) leads to more open structures; so it is surmised that the linker histones act near the entry-exit point of the DNA; they carry an overall positive charge and bind the two strands together leading to a stem formation [34–36]. Increasing the salt concentration decreases the entry-exit angle  $\alpha$  of the stem as it reduces the electrostatic repulsion between two strands.

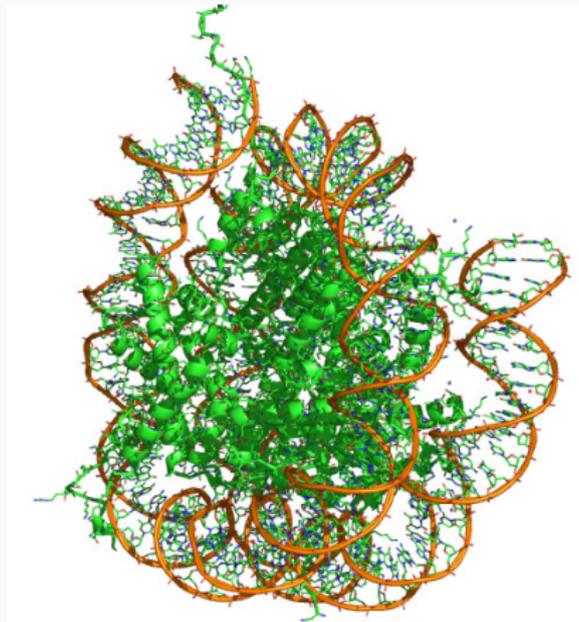


Figure 42: Nucleosome

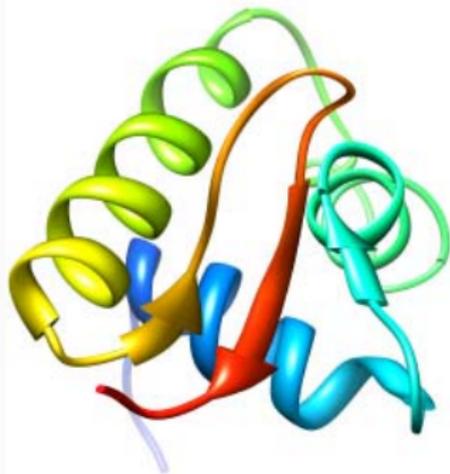
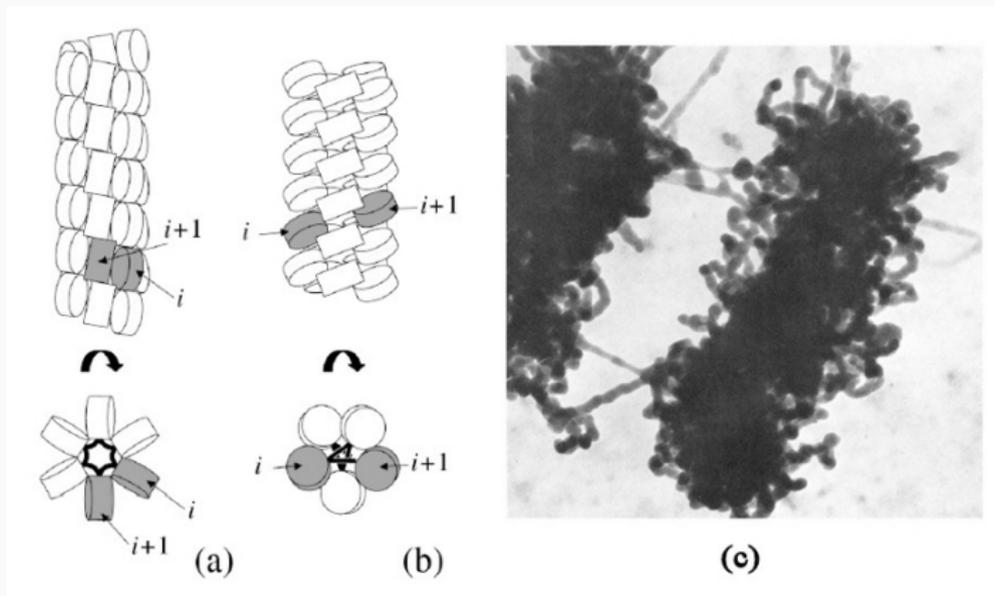


Figure 43: Histone H1



**Figure 44:** Image taken from C. L. Woodcock, S. A. Grigoryev, R. A. Horowitz, and N. Whitaker, Proc. Natl. Acad. Sci. USA 90:9021-9025.

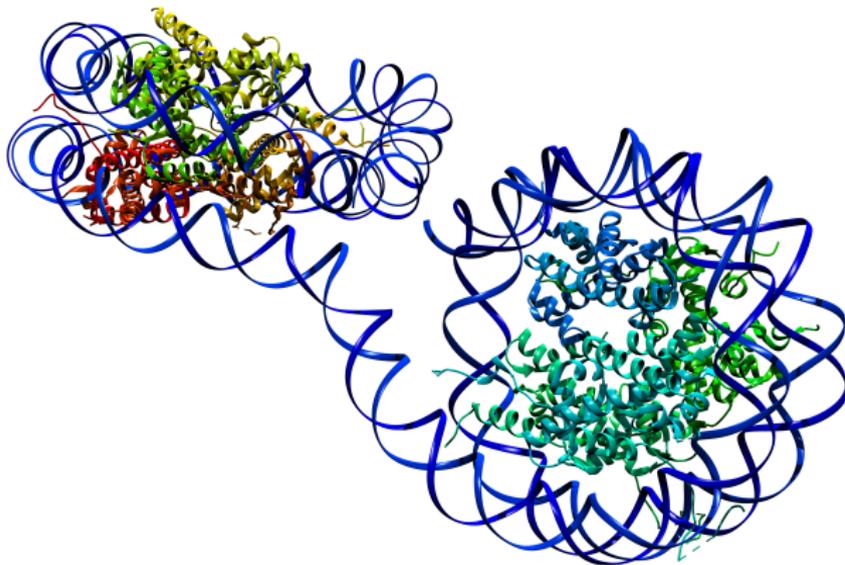
Following Woodcock [37] et al. and Schiessel et al. [38] we consider four consecutive nucleosomes (cf. Figures 45 and 46):  $N_0, N_1, N_2$  and  $N_3 \in \mathbb{R}^3$  within the chain.  $N_3$  is a function of  $N_0, \dots, N_2$  by fulfilling the following conditions:

- i  $\sphericalangle((N_0 - N_1), (N_2 - N_1)) = \alpha$ ;
- ii  $\|N_2 - N_1\| = b_2, \|N_0 - N_1\| = b_1, \|N_3 - N_2\| = b_3$ , with  $b_1, \dots, b_3 = b$ ;
- iii

$$P := \{r \in \mathbb{R}^3 \mid \exists \lambda, \mu \in \mathbb{R}, \text{ such that } r = N_1 + \lambda(N_0 - N_1) + \mu(N_2 - N_1)\}$$

$$P' := \{r \in \mathbb{R}^3 \mid \exists \lambda', \mu' \in \mathbb{R}, \text{ such that } r = N_1 + \lambda'(N_2 - N_1) + \mu'(N_3 - N_1)\}$$

$$\sphericalangle(P, P') = \beta.$$



**Figure 45:** Two nucleosomes from a crystal structure database motivating the basic definitions of the two-angle model.

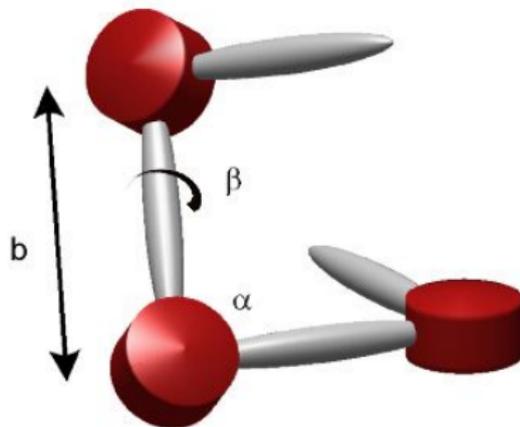


Figure 46: Basic definition of the two-angle model.

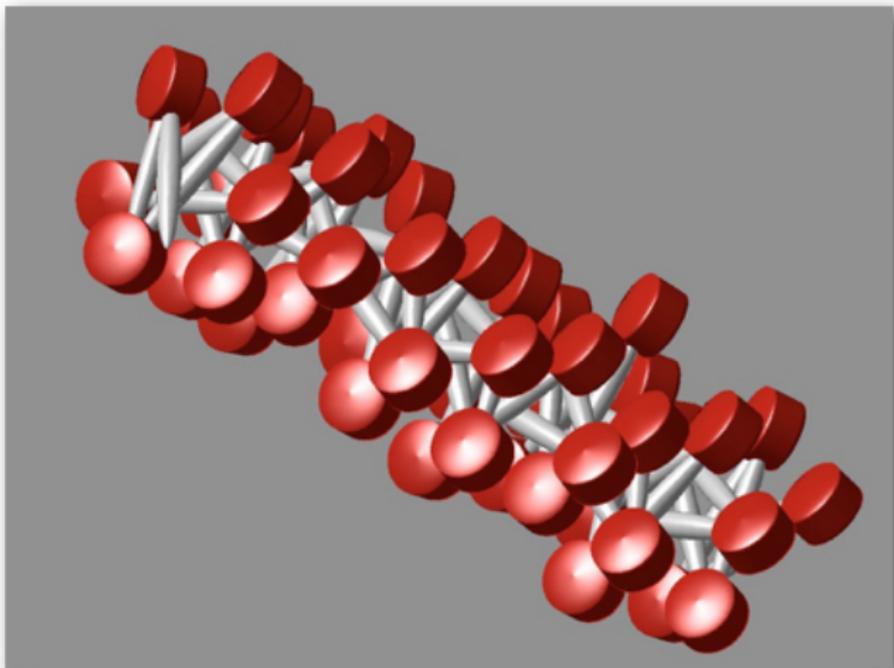
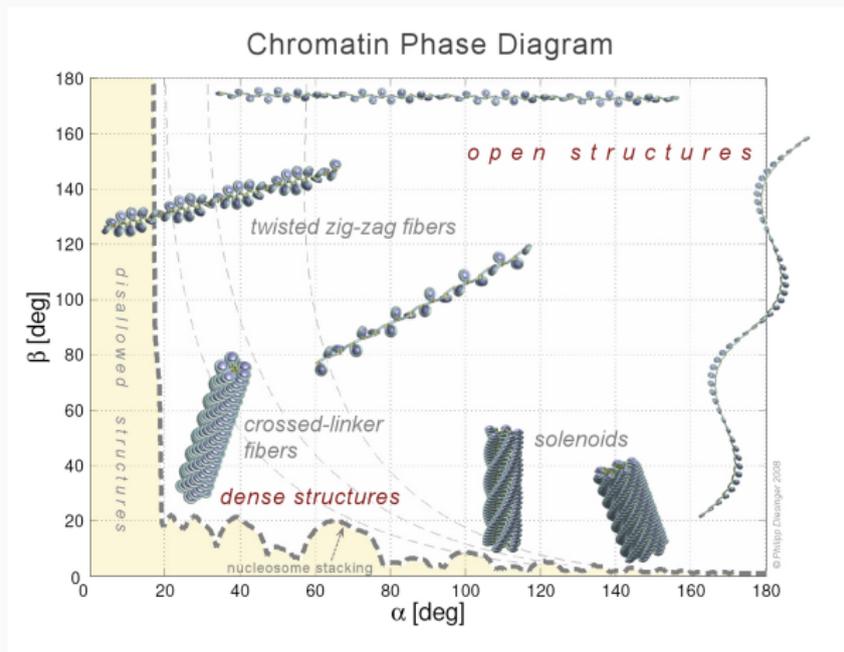
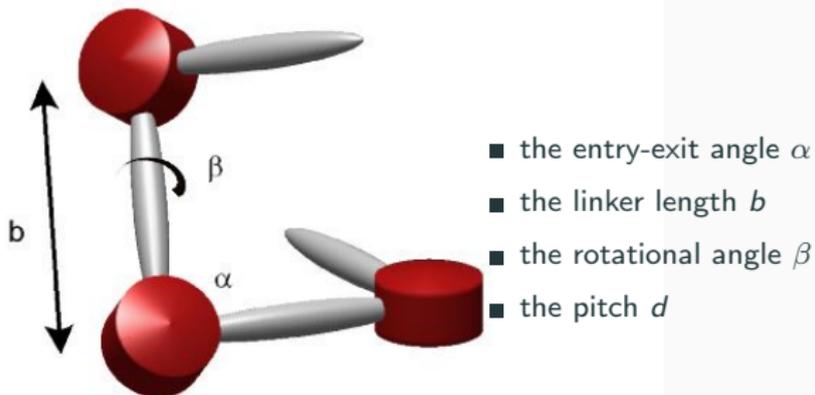


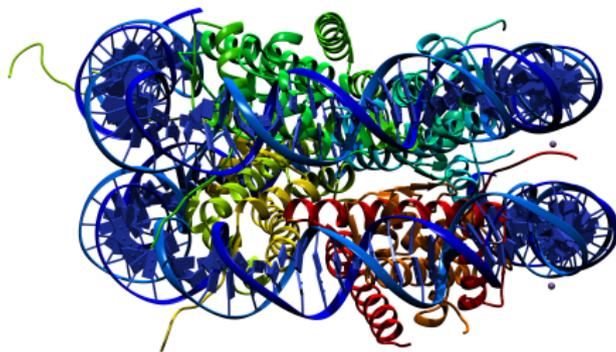
Figure 47: Two-angle model chain.



**Figure 48:** The solenoid and crossed-linker structures are most important. The dotted line is the function  $\zeta(\alpha)$  which represents the border of the forbidden region due to excluded volume.

## Basic definitions of the extended two-angle model





**Figure 49:** Nucleosome from a crystal structure database motivating to include the pitch in the basic definitions of the two-angle model.

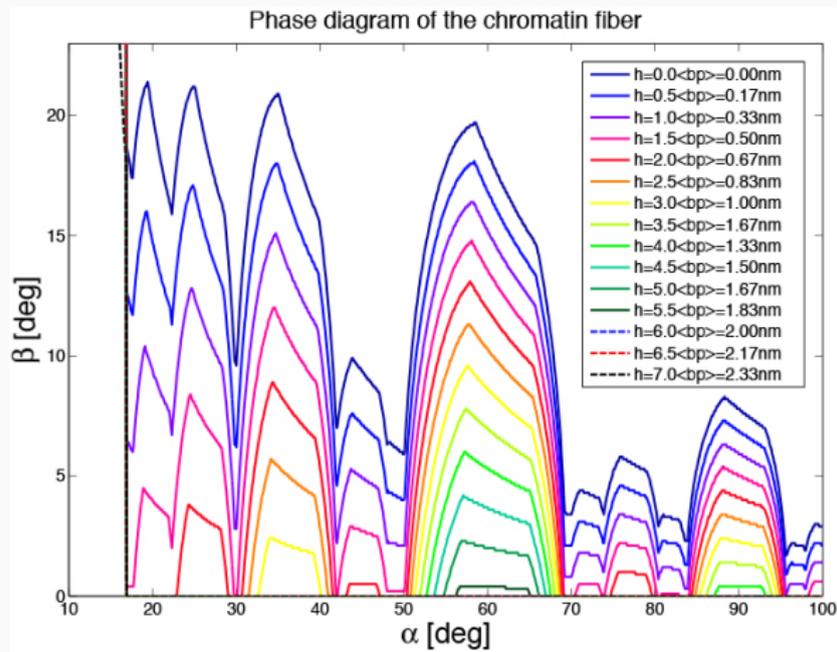
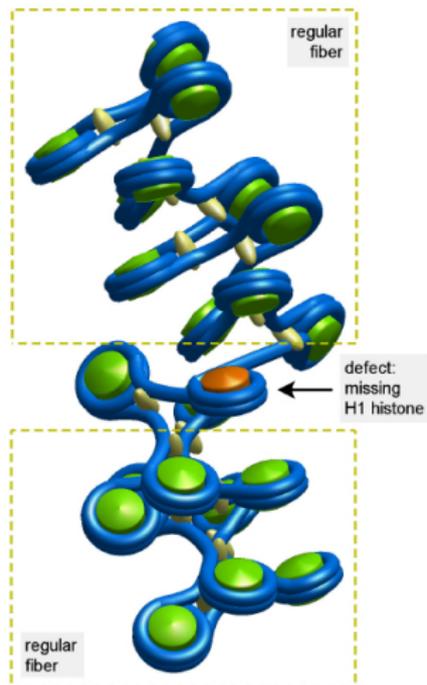
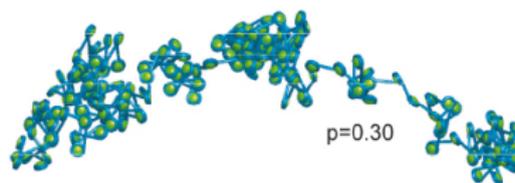
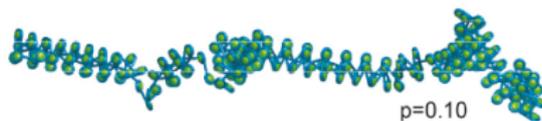
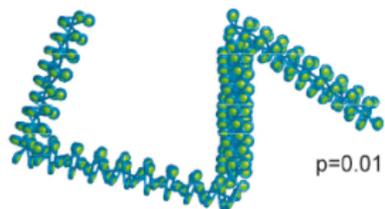


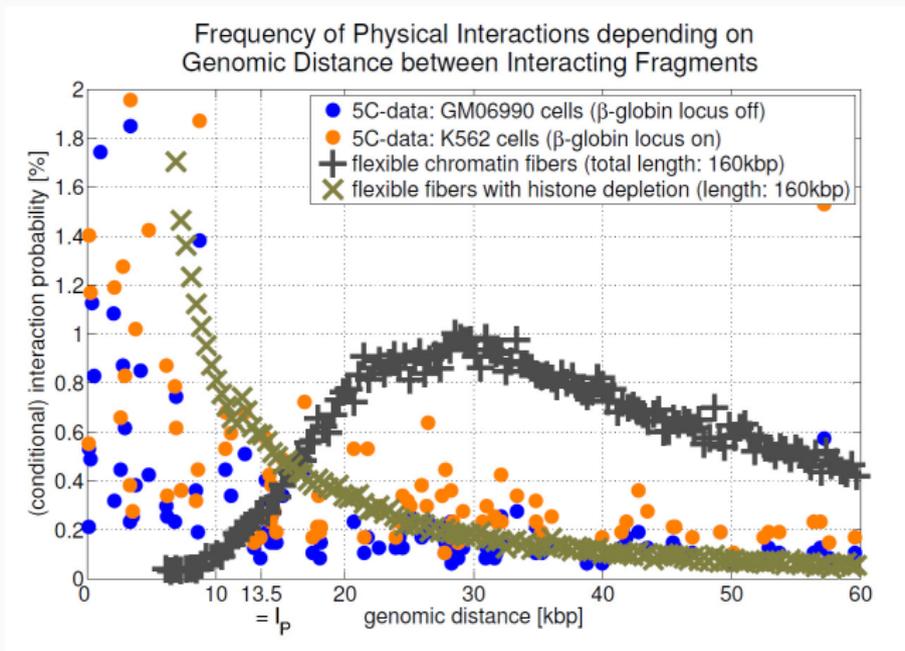
Figure 50: Phase diagram resulting from the model

# Effect of the Missing H1 Histone I



# Effect of the Missing H1 Histone II





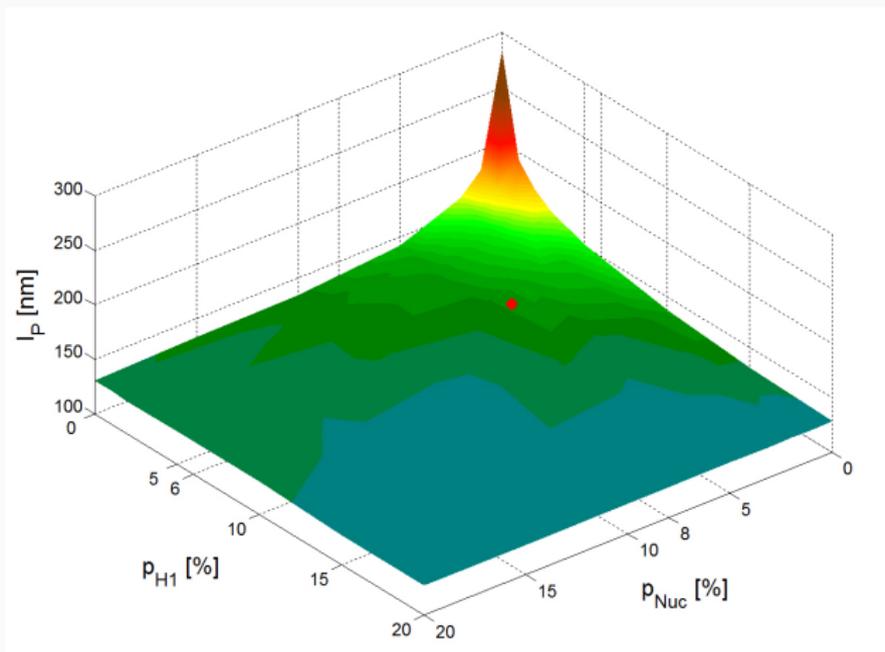


Figure 51: Definition of persistence length (and problems associated with it)

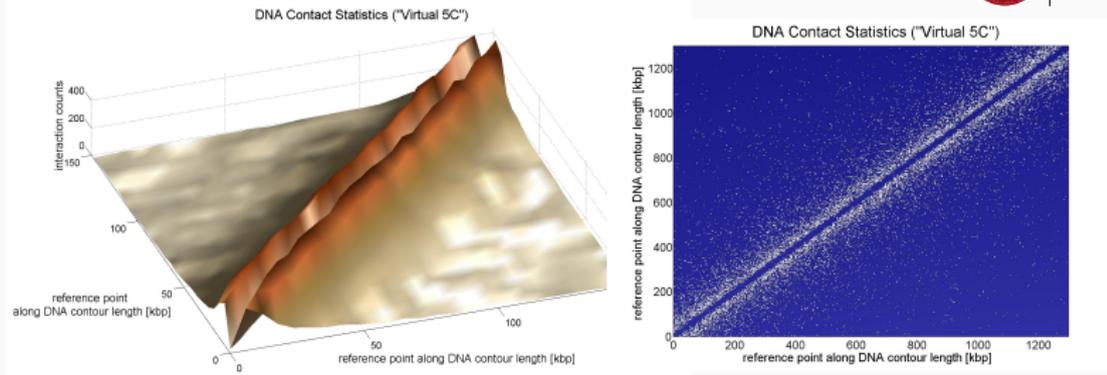


Figure 52: Contact Statistics

The last section was dedicated to the derivation of a general formula for the mean square displacement between two arbitrary beads of the chain where each bead may interact with any other via harmonic potential. This quantity turned out only to depend on the matrix  $K$ , or more accurately speaking, on its inverse. The matrix  $K$  contains all information about the interactions. Now we want to specify this matrix. Our model assumes the chromatin fibre to have a random walk backbone, meaning that  $\kappa_{ij} = \kappa$  with  $|i - j| = 1$ . Furthermore the chromatin forms loops whose size and positions are randomly distributed along the chain. On a more general footing we can restrict the possible loop sizes  $\ell$  to a certain range  $[l_1, l_2]$ . Within this range all loops are chosen randomly by setting:

$$\kappa_{ij} = \begin{cases} \kappa & \text{with probability } \mathcal{P} \\ 0 & \text{with probability } 1 - \mathcal{P} \end{cases}, \quad \text{if } l_1 \leq |i - j| \leq l_2$$
$$\kappa_{ij} = 0 \quad \text{otherwise}$$

Note that we can set  $\kappa = 1$  as it only scales the mean square displacement in eq. (??). Thus our model has two adjustable parameters, namely the chain length  $N$  and the probability  $\mathcal{P}$ .

The resulting matrices  $K$  represent an ensemble of diagonally dominated band random matrices and each matrix of this ensemble represents a loop configuration. This ensemble of random matrices has been investigated recently [? ]. We are interested in the ensemble average of the mean square displacement, i.e. in the quantity

$$\begin{aligned}\langle r_{IJ}^2 \rangle &= \langle \langle r_{IJ}^2 \rangle_{\text{thermal}} \rangle_{\text{loops}} \\ &= 3 \left( \langle \sigma_{JJ} \rangle_{\text{loops}} + \langle \sigma_{II} \rangle_{\text{loops}} - 2 \langle \sigma_{IJ} \rangle_{\text{loops}} \right) .\end{aligned}$$

This average is a quenched average and is equivalent to averaging over the ensemble of random matrices given by the above constraints. In sec. 181 we also consider the case of the annealed ensemble and give an explanation why we use the quenched one here.

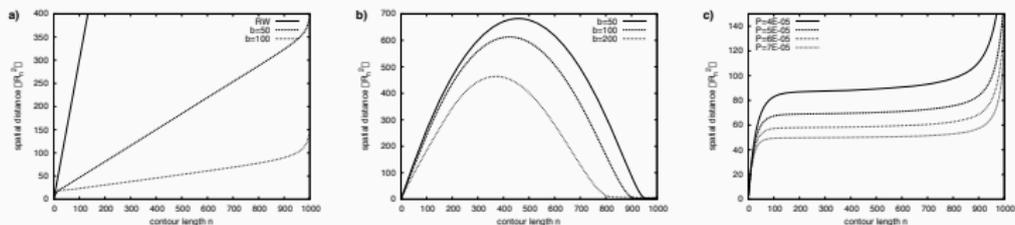
As our model already assumes that the chromatin fibre is translational invariant (as we do not take into account genomic content), we are only interested in the mean square displacement  $\langle R_n^2 \rangle$  for two beads separated by  $n = |i - j|$ .

The average over the ensemble of random matrices cannot be performed analytically, so we have to use a representative subset of the ensemble and numerically calculate the inverse matrix and thereby the mean square displacement.

As noted earlier, our polymer model makes use of coarse-graining, since it is impossible to model such a long fibre in detail. Restrictions are given by computing

time, which basically depends on the size of the matrix  $K$ . For our calculations we chose a matrix size of  $N = 1000$  as a good compromise between computing time and not too coarse graining. Using a coarse-graining approach implies that we neglect details on a scale below the effective segment length being 150 kb in the following figures. Therefore we cannot resolve those loops that have been investigated in some gene-expression systems like the  $\beta$ -globin locus. As we are interested in large scale chromatin organization, it is justified to neglect these loops as they have no effect for the levelling-off at large genomic distances, but only lead to a rescaling of the effective Kuhn length.

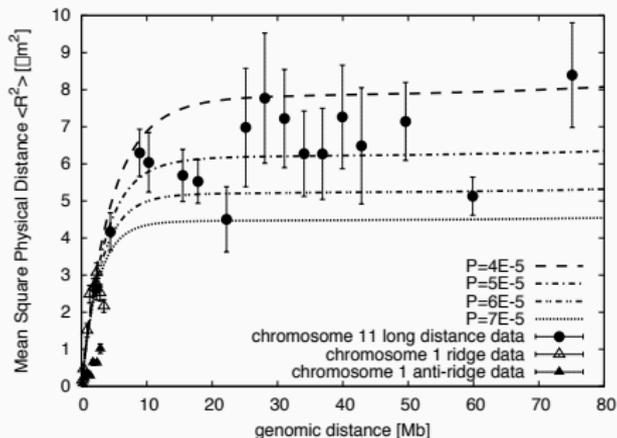
The first point of interest is which loops are necessary for the observed experimental behaviour. Do small loops (in the order of 100 kb to 1 Mb) already lead to the levelling-off, or are loops on all scales up to 80 Mb needed? Restricting the loop sizes to a range  $\ell \in [1, s]$  only leads to a rescaling of the effective segment length, we still have  $\langle R_n^2 \rangle \sim n$  (fig. 53a). In fig. 53b) we analyzed the ensemble where only large loops are allowed. Obviously, large loops are responsible for forcing the collapse of the chain, but the overall behaviour of the mean square physical distance does not fit the experimental data. Therefore loops on all scales are needed to obtain the levelling-off observed in experiment.



**Figure 53:** Mean square displacement between two chain segments in relation to their contour length for the Random Loop Model for different allowed loop sizes. The chain length is always  $N = 1000$ . In figure **a)** only loops smaller than a certain size  $s$  are allowed. The basic scaling behaviour is still  $\langle R_n^2 \rangle \sim n$  with a changed effective contour length compared to the free random walk. For this plot  $\mathcal{P}$  was chosen that the average number of loops per configuration is 100. Figure **b)** is for large loops where only loops of sizes  $\ell$  in a range  $[N - s, N]$  are allowed. While large loops seem to be responsible for the collapse of the chain, they alone cannot explain the experimental data. As in **a)**  $\mathcal{P}$  was chosen that the mean number of loops per configuration is 100. Figure **c)** shows the results for the situation where loops of all sizes are allowed. The levelling-off to  $\langle R_n^2 \rangle \sim O(1)$  can already be achieved by a small number of loops

The characteristic features of the mean square displacement allowing loops on all scales can be seen in fig. 53c). At short contour lengths the mean square displacement grows similar to a random walk, but soon a levelling-off can be observed due to the attractive long-range interactions which is fairly  $\sim O(1)$ . While the contour length approaches  $N$ , the mean square displacement again rises to a random walk like behaviour. This is a chain end effect which is not of interest to us, as experiments only measure intra-chain distances. It is due to the construction of the loops, as the probability for having a loop with a larger size becomes increasingly small.

Thus, adding long-range interactions forcing the polymer to form loops yields completely different traits than a simple random walk or self-avoiding walk model. Note that the probabilities  $\mathcal{P}$  are chosen very small, meaning that a few loops suffice to obtain this levelling-off. The number of independent randomly chosen entries  $\kappa_{ij}$  is  $\mathcal{C} = (N - 1)(N - 2)/2$  for a  $N \times N$ -matrix and therefore the average number of loops per configuration is given by  $\mathcal{C} \cdot \mathcal{P}$ . With  $\mathcal{P} = 4 \times 10^{-5}$  and  $N = 1000$  one has an average of about 20 loops.



**Figure 54:** Experimental data compared to the Random Loop Model. The data is taken from [?] and includes short and long distance sets. The results of the Random Loop Model are shown for  $N = 1000$  and different values of  $\mathcal{P}$ .

In fig. 54 the model is compared to the experimental data for different values of  $\mathcal{P}$ . Here one has to introduce two new scaling parameters, the segment length in physical units (e.g. nm) and the segment length in base pairs. The data is shown for a segment length of 300 nm and 150 kb. The latter is the size of the fluorescent markers used in experiments, therefore it does not make any sense to model on a more detailed scale. As mentioned above, using this coarse-graining approach all details on length scales smaller than 150 kb are neglected. The model can quite well explain the levelling-off at genomic distances above a few mega-base pairs as well as the rise at small genomic distances. As we have shown that on small genomic distances we have a globular-state-like behaviour [? ], this random-walk-based model does not yield perfect results here.

In a recent publication [? ] we already mentioned that plotting  $\langle R_n^2 \rangle$  versus  $n$  is not a very sensitive method to check for the correctness of a model. Looking at the cumulant relation between higher-order moments,

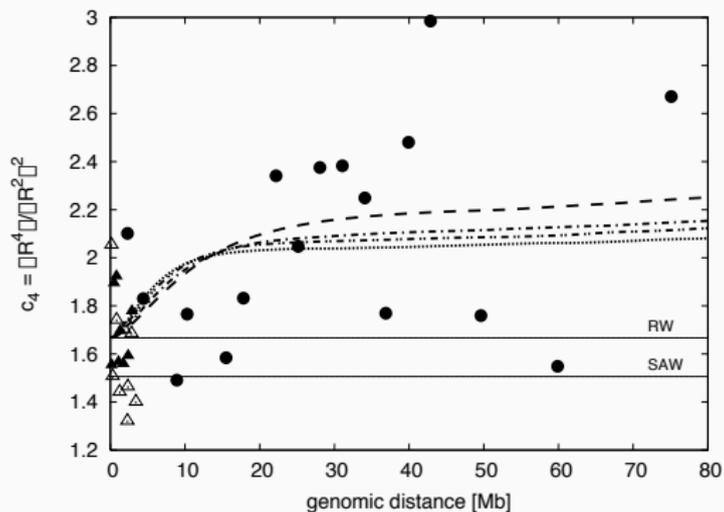
$$c_4 = \frac{\langle R^4 \rangle}{\langle R^2 \rangle^2} \quad (257)$$

gives much stronger evidence, as this expression is related to the distribution of the distances and not only its average value. Furthermore it has the advantage that the

physical length scale divides out. Eq. (257) can be easily evaluated for a Gaussian Chain, where  $c_4^{\text{RW}} = 5/3$ . To obtain the value of the cumulant relation for a self-avoiding walk one has to use the expression for the distance probability density obtained by scaling arguments [? ? ],

$$P_{\text{SAW}}(R_N) = AR_N^{\mu+2} \exp\left(-D R_N^{\frac{1}{1-\nu}}\right), \quad \mu = 0.28 \quad (258)$$

Numerical integration gives  $c_4^{\text{SAW}} \approx 1.506$ , so both RW and SAW yield a constant expression. Fig. 55 shows that this constant has a value significantly below the fluctuations of the data. Here our model is in better agreement with experiments. We should point out here the importance of averaging over the disorder of loops. The cumulant expression  $c_4$  only averaged over the thermal ensemble given by equations (??) and (??) is the same as for a pure random walk, namely 5/3. It is the average over the loop configurations that changes this behaviour, bringing it in better agreement with the data.



**Figure 55:** The cumulant expression  $c_4 = \langle R^4 \rangle / \langle R^2 \rangle^2$  of the experimental data compared to SAW, RW and Random Loop Model. The data shows significant differences to a RW and a SAW (horizontal lines), and is fluctuating around the cumulant of the Random Loop Model.

In systems with disorder one has to perform averages both over a set of statistical variables and over a set of random variables representing the disorder [? ]. In the case of the Random Loop Model the partition sum  $\mathcal{Z}$  depends on the statistical variables  $\mathbf{x}_1, \dots, \mathbf{x}_N$  and a set of random variables representing the disorder  $\{\kappa_{ij}\}$  with  $|i - j| > 1$ . Above we have performed the average first over the statistical variables, while the average over the disorder has been taken over the quantity  $\langle r_{ij} \rangle^2$ . This corresponds to a process of quenched averaging. Using this method we assume that the time the cell needs to go into a new disorder configuration is much longer than the time needed for the cell to come into thermal equilibrium,  $\tau_{\text{eq}} \ll \tau_{\text{dis}}$ . At least to our knowledge, such time scales are not known inside the cell nucleus. Time evolution measurements cannot be performed on a cell using **FISH** markers as the cell has to be fixated before applying imaging techniques. Different configurations can only be observed by looking at different cells. It seems to us more reasonable to use a quenched type of average for comparison to biological data, as the loops are functional complexes which have to persist a while in order to properly fulfill their tasks.

Nevertheless, it is interesting to consider the other case, the annealed ensemble. This type of average should be applied if  $\tau_{\text{dis}} \ll \tau_{\text{eq}}$ . The annealed average of the partition sum can be written as

$$\langle \mathcal{Z} \rangle_{\text{ann}} = \sum_{\{\kappa_{ij}\}} \mathcal{Z}(\{\mathbf{x}_k\}, \{\kappa_{ij}\}) p(\{\kappa_{ij}\}) \quad , \quad (259)$$

where the sum is over all possible configurations of disorder, and  $p$  is the probability of one such configuration. We restrict our calculation to the case that loops of all sizes are allowed and that the spring constant is  $\kappa$  for adjacent beads as well as for loops. Assuming that the  $\kappa_{ij}$  are i.i.d. Bernoulli as before, the average over the disorder can be carried out exactly,

$$\langle \mathcal{Z} \rangle_{\text{ann}} = \int d\mathbf{x}_1 \dots d\mathbf{x}_N \exp(-U_{\text{Gaussian}}) \times \prod_{i < j-1} \left[ \mathcal{P} \left( e^{-\frac{1}{2} \kappa \|\mathbf{x}_i - \mathbf{x}_j\|^2} - 1 \right) + 1 \right] .$$

Introducing the effective potential

$$U_{\text{eff}} = \frac{1}{2} \kappa \sum_{i=0}^{N-1} r_{i,i+1}^2 - \sum_{|i-j| > 1} \log \left[ 1 + \mathcal{P} \left( e^{-\frac{1}{2} \kappa r_{ij}^2} - 1 \right) \right] ,$$

where  $r_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2$ , we can rewrite the partition sum as

$$\langle \mathcal{Z} \rangle_{\text{ann}} = \int \int d\mathbf{x}_1 \dots d\mathbf{x}_N \exp(-U_{\text{eff}}) \quad (260)$$

The effective potential has two parts: Adjacent beads with  $|i - j| = 1$  keep their attractive harmonic potential, while *all* non-adjacent beads interact via a pairwise attractive potential  $V(r)$ . This potential is characterized by a minimum at  $r = 0$ , while for large  $r$  it reaches a plateau at  $V(r \rightarrow \infty) = -\log(1 - \mathcal{P})$ . In a low temperature approximation, a series expansion around  $r = 0$  up to second order gives

$$V(r) = \frac{1}{2} \mathcal{P} \kappa r^2 \quad (261)$$

– a harmonic potential with effective spring constant  $\mathcal{P} \kappa$ .

The partition sum in (260) cannot be evaluated analytically and therefore we do not obtain an expression for the mean square displacement in the annealed case. One could obtain results using extensive and time-consuming MD or MC simulations. It will be left for future investigations.

In most cases one cannot solve the model presented above analytically. Using the quenched ensemble one cannot calculate the average over the disorder, while using the annealed ensemble one cannot obtain the partition sum after having performed the disorder average. Therefore we calculated sample averages for the quenched case above. There are two special cases where the model can be solved exactly. These are the limiting cases where no disorder is present.  $\mathcal{P} = 0$  is the situation of a normal Gaussian Chain with spring constant  $\kappa$ . It is well known that the mean square distance between two beads separated by  $n$  monomers is given by  $\langle R_n^2 \rangle = \frac{3}{\kappa} n$ . The other limit,  $\mathcal{P} = 1$ , corresponds to a fully connected network of beads. Assuming that all beads interact with spring constant  $\kappa$ , we can solve this problem analytically. Here we basically do not deal with a linear chain any more. The interaction matrix  $K = (k_{ij})_{i,j}$  in this case writes

$$k_{ij} = \begin{cases} N\kappa & \text{for } i = j \\ -\kappa & \text{for } i \neq j \end{cases} \quad (262)$$

By an easy calculation one can show that the inverse matrix is given by

$$\sigma_{ij} = \begin{cases} \frac{2}{(N+1)\kappa} & \text{for } i = j \\ \frac{1}{(N+1)\kappa} & \text{for } i \neq j \end{cases} \quad (263)$$

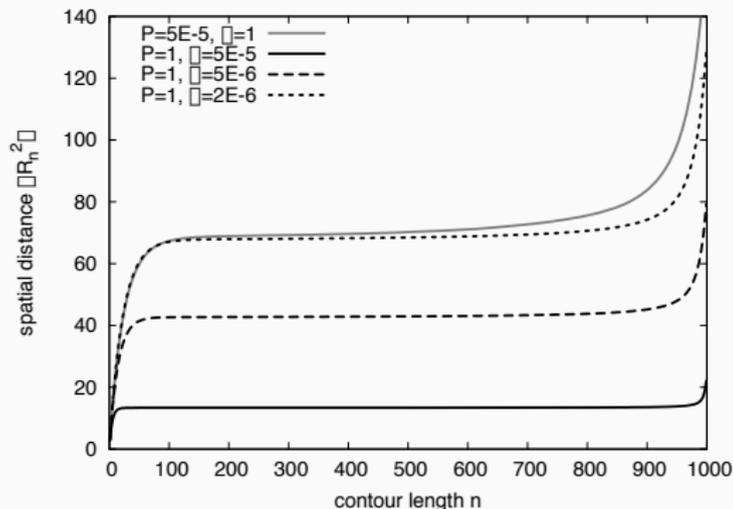
Recall our definition of the chain at the beginning of sec. ???: Although we have an  $N \times N$ -matrix our chain has  $N + 1$  beads, as we set  $\mathbf{x}_0 \equiv 0$ . Inserting into eq. (??) yields

$$\langle R_n^2 \rangle \equiv \langle r_{ij}^2 \rangle = \frac{3}{(N + 1)\kappa/2} \quad (264)$$

Within this system two beads are interacting with an effective harmonic potential with  $\kappa_{\text{eff}} = (N + 1)\kappa/2$ .

Of major interest is the case where  $\mathcal{P} = 1$ , but where adjacent beads interact with a different spring constant than loops, i.e.  $\kappa_{ij} = \kappa$  for  $|i - j| = 1$  and  $\kappa_{ij} = \hat{\kappa}$  for  $|i - j| > 1$ . We were not able to solve this case analytically. One might take this system as a model for the low-temperature limit of the annealed case in eq. (261) where  $\kappa$  is replaced by an effective interaction  $\hat{\kappa} = \mathcal{P}\kappa$ . On a more general footing this case might also be regarded as a model for a system where the random attraction with probability  $\mathcal{P}$  and loop spring constant  $\kappa$  has been replaced by an average attraction with probability  $\mathcal{P} = 1$  and loop spring constant  $\mathcal{P}\kappa$ . It is clear a priori that such a potential will lead to a collapse of the chain, as all beads are interconnected. In fig. 56 we chose  $\hat{\kappa} = \kappa = 1$  and  $\mathcal{P} = 4 \times 10^{-5}$  as the reference curve. In comparison with the case of average attraction ( $\mathcal{P} = 1, \kappa = 1, \hat{\kappa} = \mathcal{P}$ ) the levelling-off is much less pronounced. Of course it is possible to come into close agreement with the reference curve by choosing another interaction constant. For our reference curve one would

have to lower  $\hat{\kappa}$  by about one order of magnitude, corresponding to  $\mathcal{P} \sim 2 \times 10^{-6}$  ( $< 1$  loop per chain!). Although one could fit the data with these averaged attraction potential, we see no biological reason for such a potential to exist in the cell.



**Figure 56:** The Random Loop Model (RLM) compared to a system where the random attraction (setting  $\kappa_{ij} = 1$  with probability  $\mathcal{P}$ ) has been replaced by an average attraction (setting all  $\kappa_{ij} = \mathcal{P}\kappa$  for  $|i - j| > 1$ ). Shown are the RLM reference curve for  $\mathcal{P} = 5 \times 10^{-5}$  (grey line), the corresponding system with average attraction (black line) and two systems with smaller average attraction.



## Excercises

---

- Exercise 1: Find all possible random walks without self-intersections on the square lattice for length  $N=1,2,3, \dots$  and compute their mean square displacement.
- Exercise 2: Write a program to generate configurations self-avoiding lattice polymers using the reptation and pivot algorithms. Calculate mean-square end-to-end lengths and radii of gyration as function of the number of chain segments. Compare your results with the mean-field predictions.
- Exercise 3: **Random walk Metropolis updating**  
Assume that  $p_{xy} = g(y - x)$  for some arbitrary density. Clearly  $y$  is choose as  $y = x + z$  with  $z$  drawn from  $g$ , i.e. the proposed moves have the random walk character. Often,  $g$  is taken to uniform or gaussian. Use this idea to generate conformations of a linear chain in continuum. Compute the auto-correlation function for the radius of gyration.
- Exercise 4: **Independence Sampler**  
An interesting choice for  $p$  is  $p_{xy} = g(y)$ , i.e., the new candidate is drawn independent of the current state. Repeat the above exercise and compare the auto-correlation.

Exercise 5: A nucleosome is has 146 bp of DNA and wraps around a proteins making 1.75 helical turns with helix radius of 5 nm. The pitch is 3 nm. Compute the bending free energy of the DNA in units of  $k_B T$ .

Exercise 6: **Peyrard-Bishop model of DNA**

The melting of DNA can be approached from different point of view. We start from the Hamiltonian [39]

$$H = \sum_{i=1}^N \left\{ \frac{J}{2} (x_i + 1 - x_i)^2 + V(x_i) \right\} \quad (265)$$

where the variables  $x_i$  can take on real values representing the difference of the actual distance between two bases in base pair  $i$  and their equilibrium distance. The harmonic interaction represents the rigidity of the molecule due to in part to the stacking interaction between consecutive base pairs. The potential  $V(x_i) = B(e^{-Rx_i} - 1)^2$  is a Morse potential with the parameters  $B$  and  $R$ . It describes the hydrogen bonds between two bases in a base pair.  $B$  gives the strength of the potential and  $R$  is the width of attracting well of the potential.

## Exercise 7: Polyelectrolytes

The perhaps simplest model for charged flexible polymers or polyelectrolytes is obtained using the Flory ansatz for the free energy

$$\beta F = \frac{R^2}{Nb^2} + \frac{kN^2q^2}{eR}$$

where the interaction term is basically giving an electrostatic potential energy  $k^2/R$  to each of the  $N^2$  interactions between the charged monomers, and is up to a numerical constant the electrostatic energy of a sphere of radius  $R$  with charge  $Nq$  dispersed through it. How does the radius  $R$  scale with  $N$  for this case?



## Bibliography

---

## References

---

- [1] K. Binder. *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*. 1995.
- [2] K. Binder J. Baschnagel and W. Paul. *J. Chem. Phys.*, 95:6014, 1991.
- [3] M. H. Kalos M. Bishop and H. L. Frisch. *J. Chem. Phys.*, 70(1299), 1979.
- [4] J. P. Ryckaert and A. Bellemans. *Chem. Phys. Lett.*, 30(123), 1975.
- [5] E. Creighton. Understanding protein folding pathways and mechanisms, in protein folding. pages 157–170. Amer. Ass. Adv. Sci., Washington, 1990.
- [6] J. Flory. *Statistical Mechanics of Chain Molecules*. Interscience, 1969.
- [7] 12 W. J. Orr, *Trans. Faraday Soc.* **43**. 1947. 1947.
- [8] J. M. Hammersley and 23 K. W. Morton, *J. Roy. Stat. Soc. B16*. 1954. 1954.
- [9] J. M. Hammersley and 23 K. W. Morton, *J. Roy. Stat. Soc. B16*. P. 1954. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press Ithaca, New York, 1979).
- [10] 35 D. S. McKenzie, *Phys. Rep.* C27. 1976.

- [11] J. Kushnick and 1362 B. J. Berne, *J. Chem. Phys.* **64**. 1976. 1976.
- [12] J. G. Gay and 3316 B. J. Berne, *jchem* **74**. 1981. 1981.
- [13] pp. 44-55 Levinthal *J. Chem. Phys.* Vol. 65. B. 1968.
- [14] pp. 44-55 Levinthal *J. Chem. Phys.* Vol. 65. A. pages 271–277, 1968. D. MacKerell Jr., B. Brooks, C. L Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. CHARMM: The energy function and its parametrization with an overview of the program. In P. v. R. Schleyer, editor, *The Encyclopedia of Computational Chemistry*, volume 1, pages John Wiley and Sons: Chichester, 1998.
- [15] N. Madras and A.D Sokal. The pivot algorithm: a highly efficient monte carlo method for the self-avoiding walk. *J. Stat. Phys.*, 50:109–186, 1988.
- [16] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.*, 54(12):5237–48, 1971.
- [17] K. Kremer and G.S. Grest. *The Journal of Chemical Physics*, 92:5057– 5086, 1990.
- [18] E. Haber C. B. Anfinsen and F. H. White. The kinetics of the formation of native ribonuclease during oxidation of the reduced polypeptide domain. *Proceedings of the National Academy of Science*, 47:1309–1314, 1961.

- [19] pp. 44-55 Levinthal J. Chem. Phys. Vol. 65. 1968. 1968.
- [20] 24:1501 K. A. Dill. *Theory for the folding and stability of globular proteins*. Biochemistry. 1985. 1985.
- [21] C. Tang H. Li, R. Helling and 666 N. Wingreen, Science **273**. 1996.
- [22] B. Berger and 5(1):27-40 T. Leighton. *Protein folding in the hydrophobic-hydrophilic (HP) model is NP-complete*. Journal of Computational Biology. 1998. 1998.
- [23] C. Papadimitriou A. Piccolboni P. Crescenzi, D. Goldman and 5(3):423-465 M. Yannakakis. *On the complexity of protein folding*. Journal of Computational Biology. 1998. 1998.
- [24] B. Derrida. Random-energy model: An exactly solvable model of disordered systems. pages 2613–2626, 1981. Phys. Rev. B **24**.
- [25] D. J. Gross and **240** 431-452 M. Mezard *The simplest spin glass*. Nucl. Phys. B. 1984. 1984.
- [26] J. Bryngelson and P. Wolynes. Spin glasses and the statistical mechanics of protein folding. pages 7524–7528, 1987. Proc. Natl. Acad. Sci. USA 84.
- [27] E. Shakhnovich and A. Gutin. Formation of unique structure in polypeptide chain. *Biophys. Chem.*, 34:187–199, 1989.

- [28] BH; Bragg JK Zimm. D. 1959. Poland, H.A. Scheraga, Theory of Helix-Coil Transition in Biopolymers (Academic Press, New York, 1970).
- [29] BH; Bragg JK Zimm. C. 1959. R. Cantor, P.R. Schimmel, Biophysical Chemistry (Freeman, New York, 1980).
- [30] BH; Bragg JK Zimm. A. 1959. Y. Grosberg, A.R. Khokhlov, Statistical Physics of Macromolecules (AIP Press, New York 1994).
- [31] 81 253 E. Ising, Z. Phys. 1923.
- [32] A.V. Sivolob N. Khraounov, A.I. Dragan and A.M. Zagariya. *Biochim. Biophys. Acta*, page 1351, 1997.
- [33] C. J. Geyer and 909-920 E. A. Thomposon, J. Am. Statis. Ass **90**. Luger et al. pages 251–260, 1995. 1997. Nature. 389:.
- [34] C. J. Geyer and 909-920 E. A. Thomposon, J. Am. Statis. Ass **90**. Bednar et al. pages 4173–1417, 1995. 1998. Proc. Natl. Acad. Sci. USA. 95:18.
- [35] C. J. Geyer and 909-920 E. A. Thomposon, J. Am. Statis. Ass **90**. Dustin, i. pages 15–21, 1995. P. Furrer, A. Stasiak, J. Dubochet, J. Langowski & E. Egelman. 1991. J. Struct. Biol. 107:.

- [36] C. J. Geyer and 909-920 E. A. Thomposon, *J. Am. Statis. Ass* **90**. Stasiak, a. 1995. A. Adrian, J. Bednar, P. Furrer, B. ten Heggeler Bordier, W. Wahli & J. Dubochet. 1992. Basel, *Experientia* 48, A79.
- [37] C. J. Geyer and 909-920 E. A. Thomposon, *J. Am. Statis. Ass* **90**. Woodcock, c. pages 9021–9025, 1995. L.,Grigoryev, S. A.,Horowitz, R. A., and Whitaker, N. 1993 : *Proc. Natl. Acad. Sci. USA.* 90:.
- [38] W. M. Gelbart Schiessel and Robijn Bruinsma. *Biophys J.*, 80(4):1940–1956, 2001.
- [39] M. Peyrard T. Bauxois and A. R. Bishop. *Phys. Rev E*, 47(684), 1993.



# Index

---

$\nu$ , critical exponent, 16

Alanine, 99

alpha-helix, 129

AMBER, 67

bending modulus, 35

beta sheet, 100

Bjerrum length, 73

bond, 13

bond angle, 6

chain length, 6

CHARMM, 67

co-polymer, 124

conformation, 6

connectivity constant, 39

contour length, 6

cytoskeletal network, 104

degree of polymerization, 6

designability, 120

dielectric constant, 73

dielectric screening, 73

dipole moment, 75

divalent ions, 72

DNA melting, 146

effective number of repeat units, 18

end-to-end distance, 13

Flory characteristic ratio, 25

freely-jointed chain model, 11

freely-rotating chain, 23

Gaussian chain model, 27

GROMOS, 67

helix-coil transition, 129

heteropolymer, 123

homopolymer, 124

HP Model, 117

Independence Sampler, 191

Kuhn length, 18

lattice polymer models, 85

Metropolis-coupled Markov chain Monte Carlo, 112

monomer, 5

native conformation, 99

network, cytoskeletal, 104

NP-complete, 120

permittivity of free space, 73

persistence length, 18

Peyrard-Bishop model, 156, 192

Pivot Algorithm, 88

pivot algorithm, 88

Poland-Scheraga-model, 146

protein folding, 105

proteins, 96

radius of gyration, 20

random energy model, 123

Random walk Metropolis updating, 191

REM, random energy model, 123

repeating unit, 5

reptation algorithm, 86, 87

rotamer, 96

SAW, 38

self avoiding random walk, 38

simulated tempering, 112

univalent ions, 72

Van der Waals interactions, 78

Worm-like Chain Model, 30

worm-like chain model, 30

Yamakawa, 18

Zimm-Bragg Mode, 129

zipper-model, 153