Biophysics
A Computational Approach
Concepts, Models, Methods and Algorithms
Lecture 7: RNA, Proteins and DNA

Dieter W. Heermann

Heidelberg University

August 28, 2016

Table of Contents

Introduction I

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

Introduction II

- 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].

Some Measures I

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.

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

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].

Some Measures II

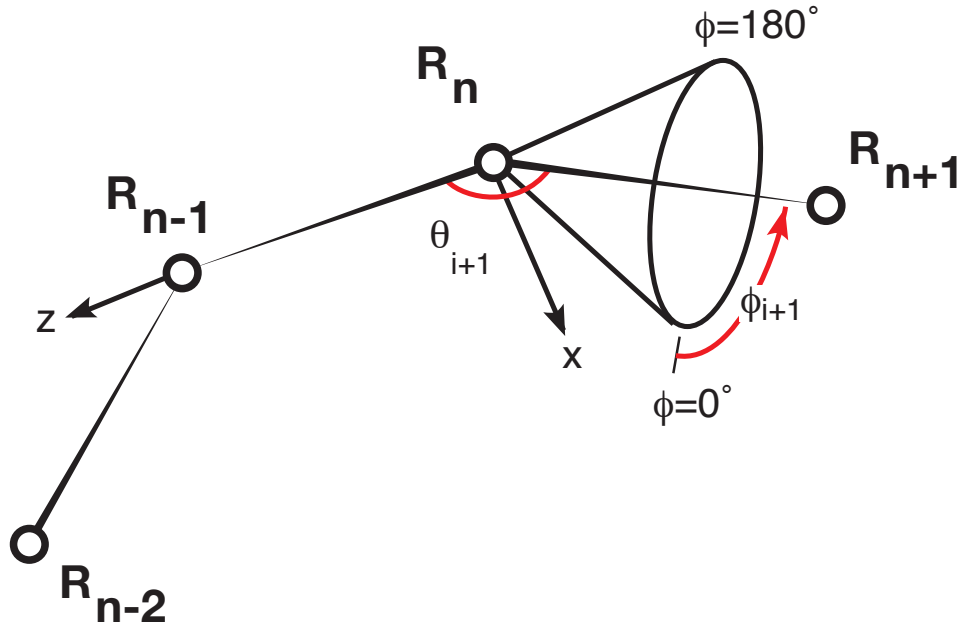


Figure: Angle definition

Some Measures III

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: 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°	t	stable

Some Measures IV

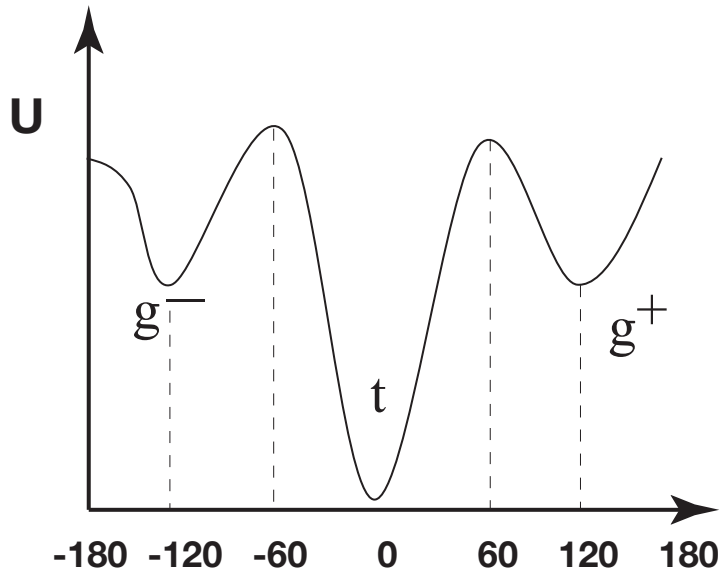


Figure: A typical torsion potential

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.

Freely-Jointed Chain I

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). 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.

Freely-Jointed Chain II

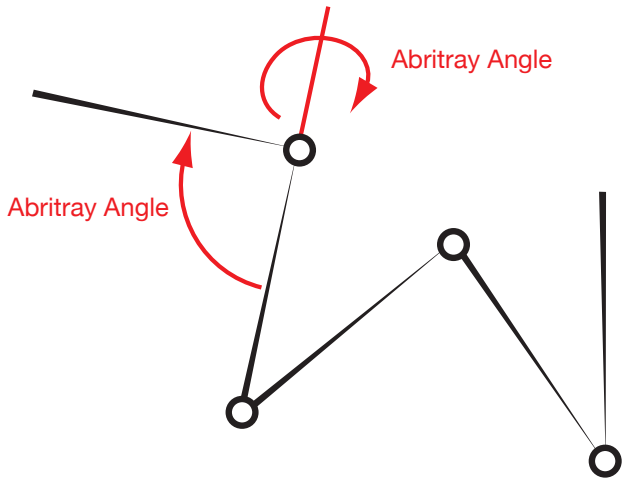


Figure: Freely-jointed chain

Freely-Jointed Chain III

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

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

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

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

which we assume to be a random variable.
Because the $\mathbf{r}_n$ are uncorrelated we must have

$$\langle \mathbf{r}_n \rangle = 0 \tag{5}$$

after averaging over all possible conformations and

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

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

$$\langle \mathbf{R}_e \rangle = \langle \sum_{n=1}^{N} \mathbf{r}_n \rangle = \sum_{n=1}^{N} \langle \mathbf{r}_n \rangle = 0 \tag{7}$$

Freely-Jointed Chain IV

and

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

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

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

$$= Nb^2 \quad . \quad (11)$$

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

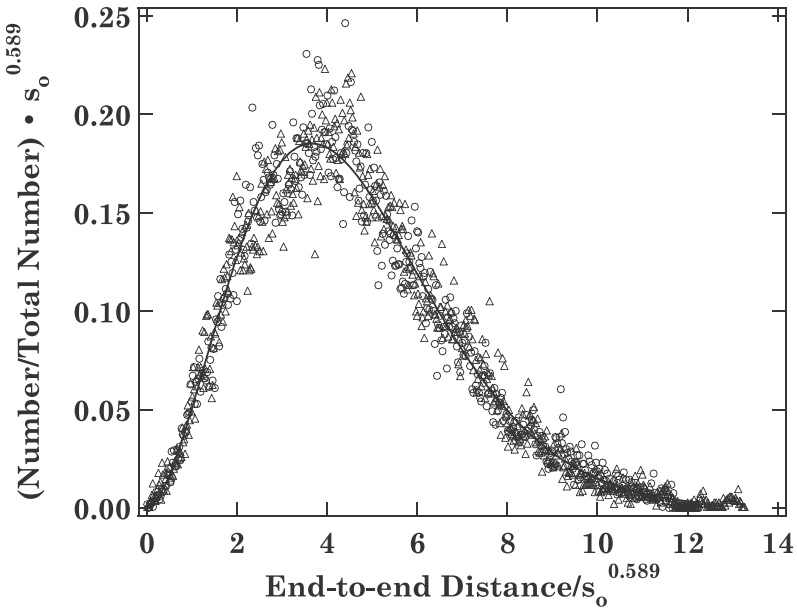
$$\langle R_e^2 \rangle \propto N \quad (12)$$

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 ν. Thus $\nu = 1/2$ here.

Freely-Jointed Chain V



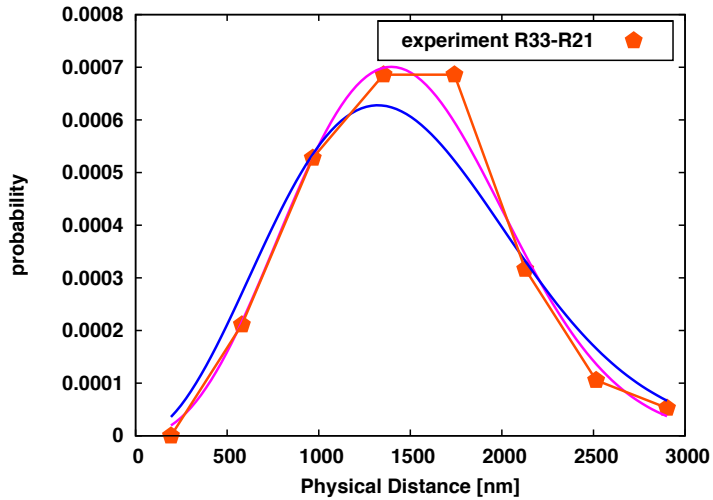


Figure: End-to-end distance distribution for two biological macromolecules. The lower figure shows the comparison of data on human chromosomes (taken from van Driel et. al 2007). The blue curve shows the result for the random walk and the purple the result for a self-avoiding walk fitted to the data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

Freely-Jointed Chain VI

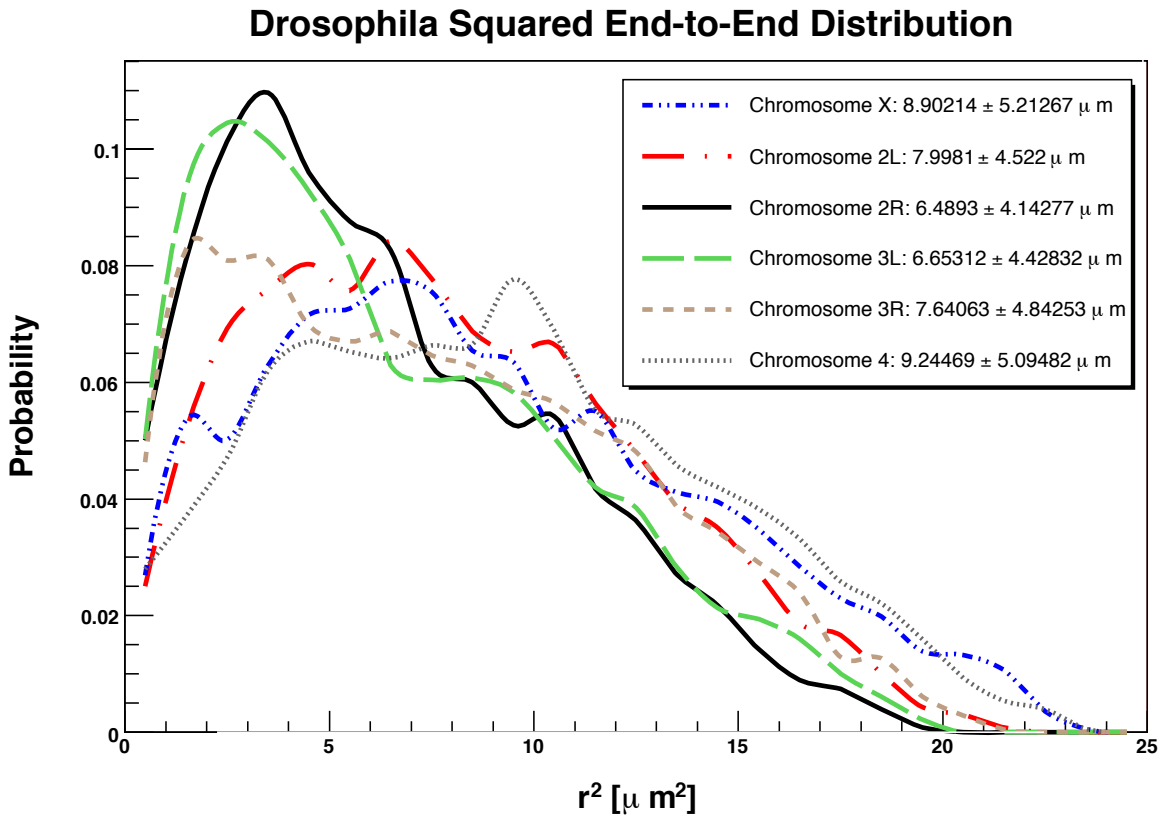


Figure: Histogram representing the distribution of the end-to-end distance for two different contour lengths (548nm circles, 748nm triangles) and how they collapse onto each other. Taken from Dietler et. al. PRL 95, 158105 (2005)

Freely-Jointed Chain VII

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

$$\langle R_e^2 \rangle = N_K l_K^2 \tag{14}$$
$$L_c = N_K l_k \quad , \tag{15}$$

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* ξ_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.

$$\xi_p = l_k/2 \quad L_c \gg l_k \tag{16}$$

and hence

Freely-Jointed Chain VIII

$$\langle R_e^2 \rangle = 2N_p\xi_p^2 \quad L_c \gg l_k \tag{17}$$
$$L_c = N_p\xi_p \tag{18}$$

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 $\xi_p ? 50nm$. 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 $N \to \infty$ (central limit theorem) since it is the sum of independent random variables and we must have

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

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 $\mathbf{R}_e$. Another measure is the *radius of gyration* R_g

Freely-Jointed Chain IX

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

with the center of mass

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

For the freely jointed chain model we obtain

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

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

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

Freely-Jointed Chain X

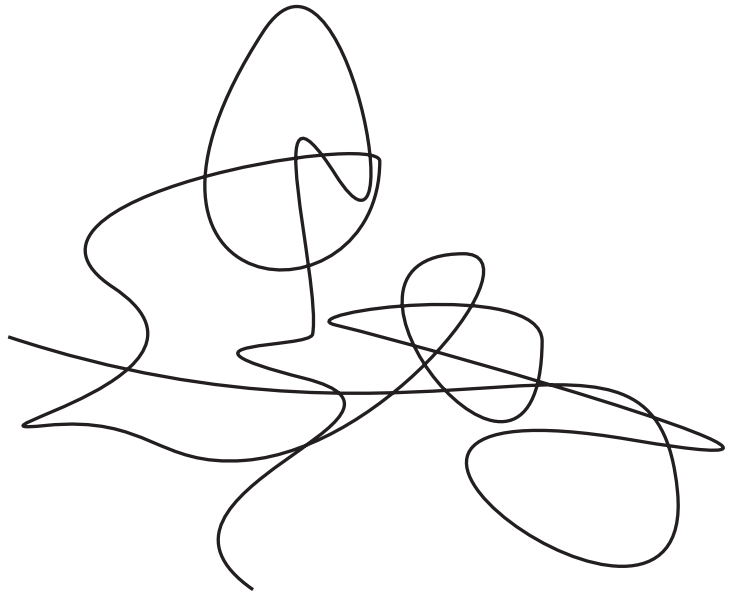


Figure: A typical random coil conformation

Freely-Jointed Chain XI

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

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

from which we obtain the free energy difference

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

Freely-rotating chain I

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

$$\langle \mathbf{r}_n \mathbf{r}_m \rangle \tag{26}$$

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

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

and with this

$$\begin{aligned} \langle \mathbf{R}_e^2 \rangle &= \sum_{n=1}^{N} \sum_{m=1}^{N} \langle \mathbf{r}_n \mathbf{r}_m \rangle & (28) \\ &= b^2 \sum_{n=1}^{N} \sum_{m=1}^{N} (\cos\theta)^{|m-n|} & (29) \\ &= N b^2 \frac{1+\cos\theta}{1-\cos\theta} \quad . & (30) \end{aligned}$$

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

Freely-rotating chain II

We can generalize the above result assuming a finite correlation

$$\lim_{|m-n|\to\infty} \langle \cos\theta_{nm} \rangle) = 0 \tag{31}$$

With this assumption we have

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

and thus

$$\begin{aligned} \langle \mathbf{R}_e^2 \rangle &= b^2 \sum_{n=1}^{N} \sum_{m=1}^{N} \langle \cos\theta_{nm} \rangle & (33) \\ &= b^2 N \sum_{n=1}^{N} C_n & (34) \\ &= N b^2 C_\infty & (35) \end{aligned}$$

where C_∞ is called the *Flory characteristic ratio*.
To make the connection with the persistence length we note that

Freely-rotating chain III

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

with

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

we find the persistence length

$$\xi_p = b\xi \tag{38}$$

Gaussian Chain Model I

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

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

with the expectation for the link length being

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

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

$$\begin{aligned} P(\mathbf{R}_e) &= P(\{\mathbf{r}_n\}) && (41)\\ &= \prod_{n=1}^{N} \left(\frac{3}{2\pi b^2}\right)^{3/2} \exp\left(-\frac{3\mathbf{r}_n^2}{2b^2}\right) && (42)\\ &= \left(\frac{3}{2\pi b^2}\right)^{3/2} \exp\left(-\sum_{n=1}^{N} \frac{3(\mathbf{R}_n - \mathbf{R}_{n-1})^2}{2b^2}\right) && (43) \end{aligned}$$

and hence for the entropy

Gaussian Chain Model II

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

$$= \text{const} - \frac{3}{2\mathrm{b}^2} \sum_{\mathrm{n}=1}^{\mathrm{N}} \mathbf{r}_{\mathrm{n}}^2 \quad . \tag{45}$$

From this we obtain the free energy

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

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

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

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

$$\langle R_e^2 \rangle \propto N \quad . \tag{48}$$

Worm-like Chain Model I

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 a continuous description. We first pull on the results that we have derived before

$$\langle \mathbf{R}_e^2 \rangle = \sum_{n=1}^{N} \sum_{m=1}^{N} \langle \mathbf{r}_n \mathbf{r}_m \rangle \tag{49}$$

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

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

Since we want b to tend to zero we can substitute

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

Worm-like Chain Model II

and thus

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

with the result

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

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

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

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

$$\langle \mathbf{R}_e^2 \rangle \approx R_{max}^2 \tag{56}$$

so that the chain a just like a rod.
In figure 7 is shown a comparsion of the worm-like chain model with data on chromosomal yeast in interphase for small genomic distances.

Worm-like Chain Model III

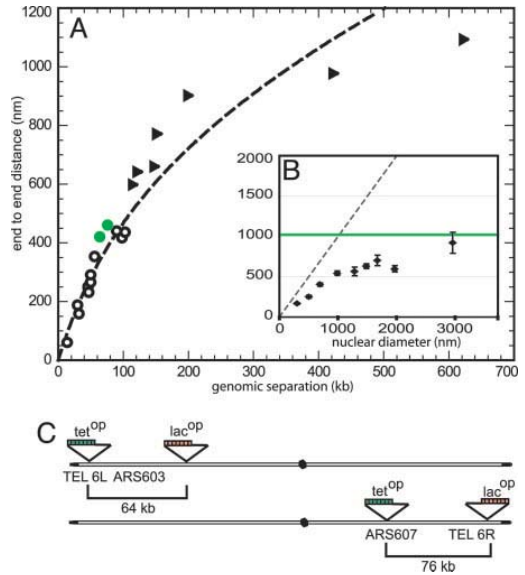


Figure: Taken from *Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques*, Kerstin Bystricky, Patrick Heun, Lutz Gehlen, Jörg Langowski, and Susan M. Gasser, PNAS November 23, 2004 vol. 101 no. 47 16495-16500

Worm-like Chain Model IV

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

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

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 $N \to \infty$, $b \to 0$ and $\epsilon \to \infty$ with

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

keeping the contour length also constant. Using

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

we have

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

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

Worm-like Chain Model V

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

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

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

with the *bending modulus* $\kappa = \epsilon b$.
Thus the partition function is given by

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

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

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

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

Worm-like Chain Model VI

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

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

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

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

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

Worm-like Chain Model VII

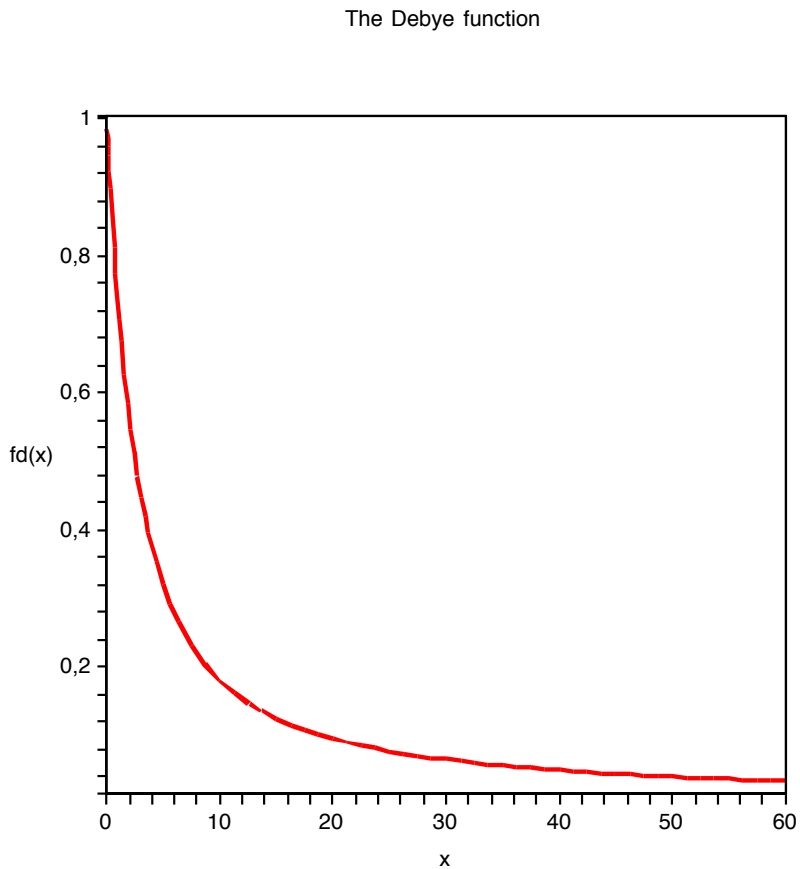


Figure: The Debye-function

Self Avoiding Random Walk I

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 paritition 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,

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

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

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

$$c_N \approx A\mu^N N^{\gamma-1} \tag{71}$$

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

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

and thus for the average end-to-end distance

Self Avoiding Random Walk II

$$\langle R_e^2 \rangle \propto N^{2\nu} \tag{73}$$

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

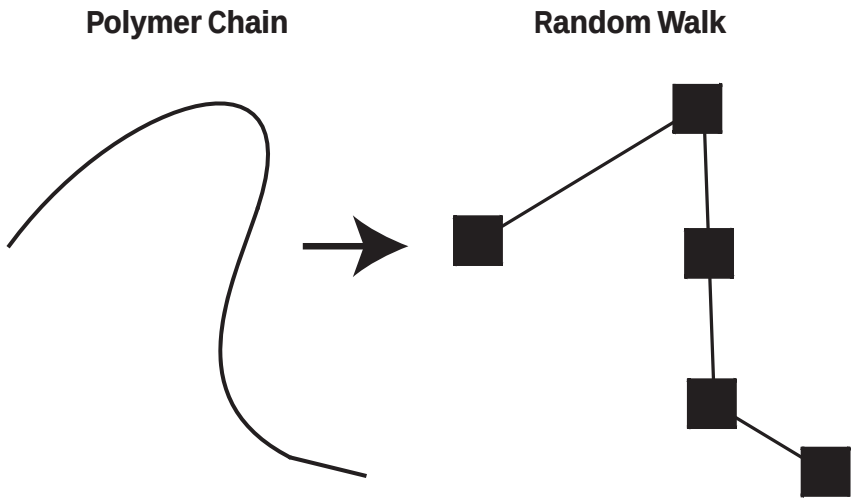


Figure: From a continuous to a lattice description

Self Avoiding Random Walk III

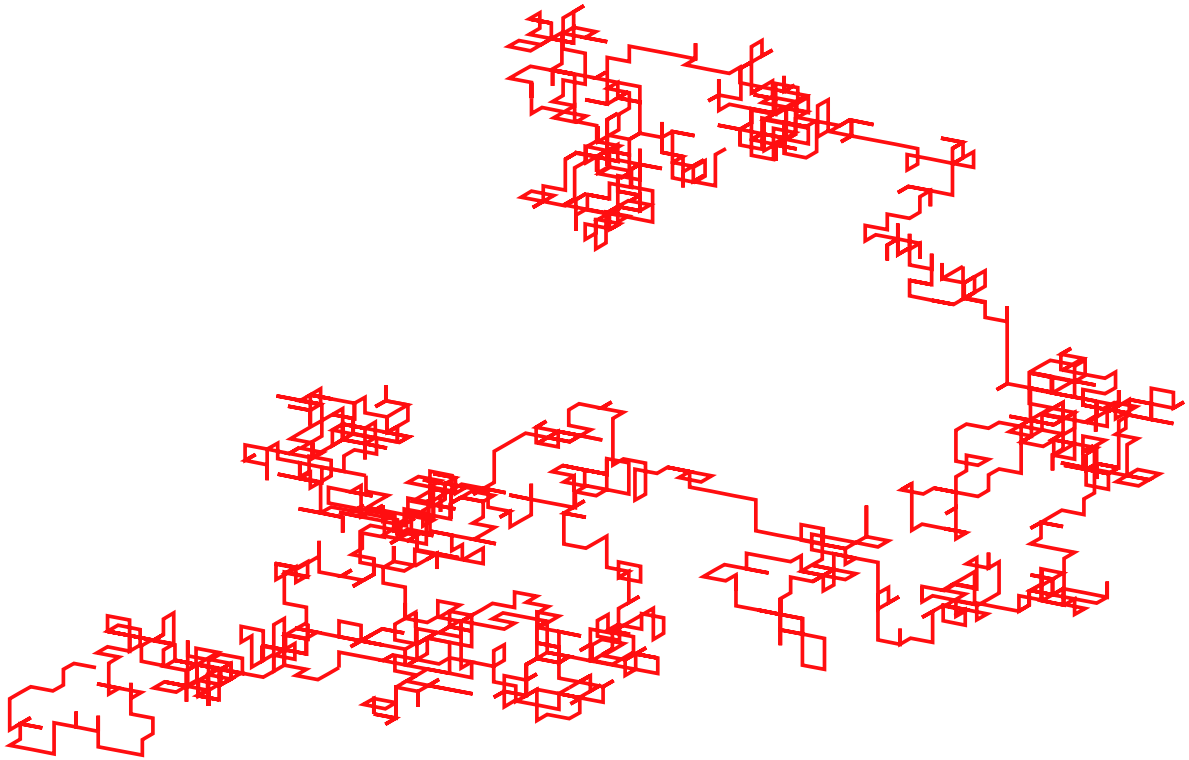


Figure: A sample of a random walk in three dimensions on a lattice

Self Avoiding Random Walk IV

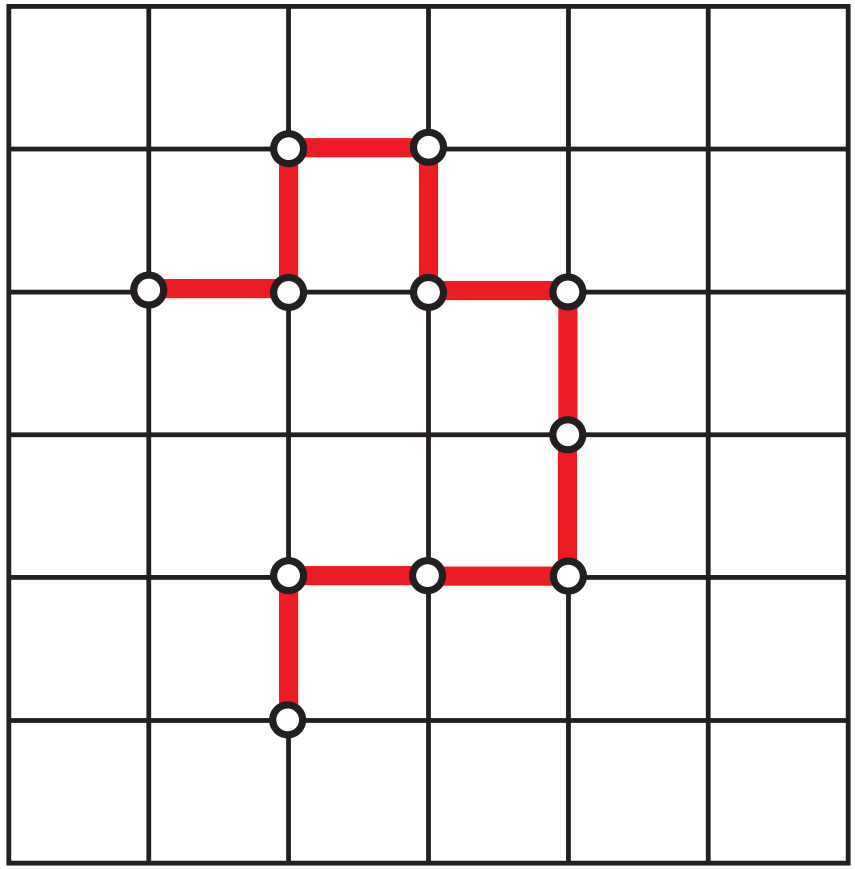


Figure: A self-avoiding random walk (SAW)

Modelling of Biopolymers I

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 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:

$$H_{\mathbf{inter}} = V_{\mathbf{abs}} \left(\frac{d_1^{(p)} + d_2^{(p)}}{2} - c \right), \tag{74}$$

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

$$V_{\mathbf{abs}}(r) = r^{-6} \tag{75}$$

into account because from quantum chemical calculations the attractive part proves to be negligible. The calculation of the distances is illustrated in figure 12.

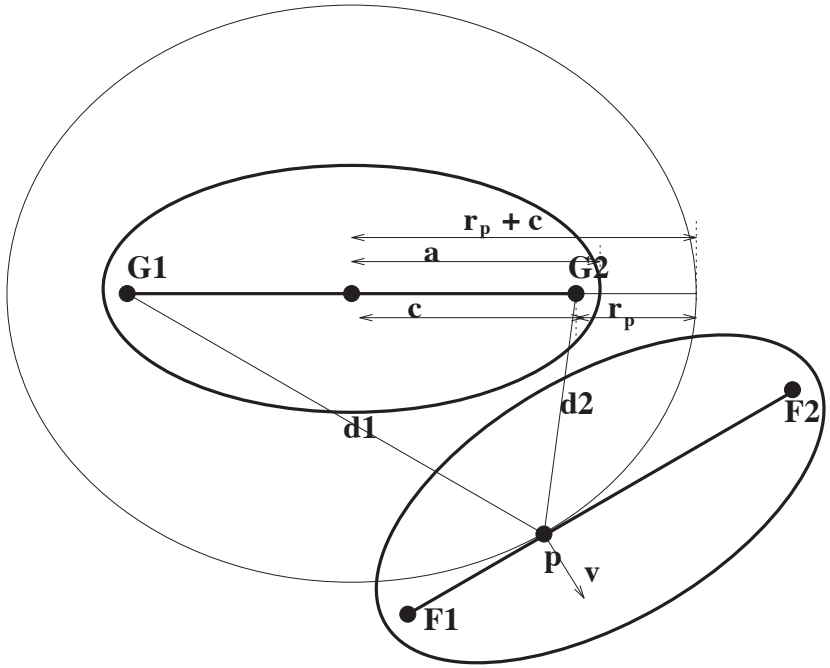


Figure: Interaction with a con-focal force field

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:

Modelling of Biopolymers VI

$$
\begin{aligned}
V(\{\mathbf{R}\}) &= \sum_{\text{bonds}} c_i (l_i - l_0)^2 && (76)\\
&+ \sum_{\text{bond angles}} c_\alpha (\theta_\alpha - \theta_0)^2 && (77)\\
&+ \sum_{\text{improper torsion angles}} c_\beta (\tau_\beta - \tau_0)^2 && (78)\\
&+ \sum_{\text{dihedral angles}} \text{tri}(\omega) && (79)\\
&+ \sum_{\text{charged pairs}} \frac{Q_i Q_j}{\epsilon r_{ij}} && (80)\\
&+ \sum_{\text{unbond pairs}} c_w \Phi \left(\frac{R_i + R_j}{r_{ij}} \right) && (81)
\end{aligned}
$$

where

$$
r_{ij} = |\mathbf{R}_i - \mathbf{R}_j| \quad . \qquad (82)
$$

Here ϵ is the dielectric constant and Q_i are the partial charges. The term tri refers to a linear combination of trigonometric functions and and multiples of ω. The term Φ 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.

- Prnuning
- 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

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

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

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

Code 1: Reptation Algorithm

Algorithm 2 Verdier-Stockmayer Algorithm

...

Lattice Polymer Models: Pivot Algorithm

Let W denote the set of self-avoiding walks of length N on a lattice λ. 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 π.

Algorithm 3 Pivot Algorithm (Sokal)

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

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

$$\pi(\omega_i)P(\omega_i,\omega_j) = \pi(\omega_j)P(\omega_j,\omega_i) \quad . \tag{83}$$

Since π 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 ω to another self-avoiding walk $\bar{\omega}$. For $i = 1, 2, ..., m$, consider the pairs (x_i, g_i). Each pair gives a transition, using the pivot algorithm from ω to $\bar{\omega}$.
Thus,

$$P(\omega,\bar{\omega}) = \sum_{i=1}^{m} P(g = g_i) \cdot P(x = x_i) \quad . \tag{84}$$

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

$$P(\omega,\bar{\omega}) = \sum_{i=1}^{m} P(g = g_i)\cdot P(x = x_i) = \sum_{i=1}^{m} P(g = g_i^{-1})\cdot P(x = x_i) = P(\bar{\omega},\omega) \quad . \tag{85}$$

United Atom Model I

A very simple but useful model for a polymer chain is the united atom model (c.f. Figure 13) 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

$$\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2 + \mathcal{H}_3 \tag{86}$$

$$\mathcal{H}_1 = \sum_i \frac{1}{2} k_b (l_i - l_0)^2 \tag{87}$$

$$\mathcal{H}_2 = \sum_i \frac{1}{2} k_\theta (\cos\theta_i - \cos\theta_0)^2 \tag{88}$$

$$\mathcal{H}_3 = \sum_{i<j} u(r_{ij}) \tag{89}$$

where

United Atom Model II

$$u(r_{ij}) = \begin{cases} u_{LJ}(r_{ij}) - u_{LJ}(r_c) - \frac{\partial}{\partial r_c} u_{LJ}(r_c)(r_{ij} - r_c) & r_{ij} < r_c \\ 0 & r_{ij} \geq r_c \end{cases} \tag{90}$$

and

$$u_{LJ}(r_{ij}) = 4\epsilon \sum_{i,j} [(\frac{\sigma}{r_{ij}})^{12} - (\frac{\sigma}{r_{ij}})^{6}] \tag{91}$$

Note that the Lennard-Jones part of the potential is cut-off at 1.5σ 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.

United Atom Model III

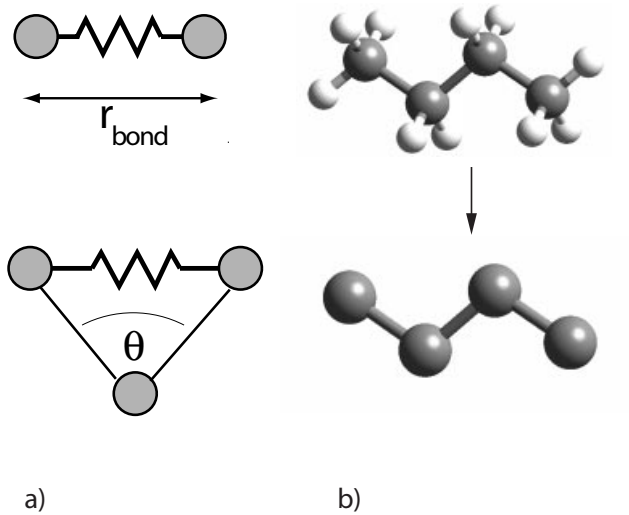


Figure: The definition of the bond length and the bond angle potential

United Atom Model IV

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,

$$U_{\mathrm{WCA}}(r) = \begin{cases} 4\epsilon \left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + c_{shift} \right) & r < r_{cut} \\ 0 & r \geq r_{cut} \end{cases} \tag{92}$$

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

$$U_{\mathrm{FENE}}(r) = \begin{cases} -\frac{1}{2}\, k_{\mathrm{FENE}}\, R_0^2\, \log(1 - (r/R_0)^2) & r < R_0 \\ +\infty & r \geq R_0 \end{cases} \tag{93}$$

United Atom Model V

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,

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

Here, the parameters are the looping probability ¶ and the interaction strength κ_{loops} (the κ_{ij} being either this value or zero).
The following parameters are chosen for the simulation runs:

$$R_0 = 1.6\sigma \qquad \kappa_{loops} = 2.0$$
$$k_{\text{FENE}} = 10.0 \qquad \text{temperature } T = 1.0$$
$$\sigma = 1.0 \qquad \text{friction } \Gamma = 0.5$$
$$\epsilon = 20.0 \qquad \text{timestep } t = 0.006$$

Special care is required for the relation between R_0 and σ. 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 I

Proteins (see figure 17 for an example) are the machines and building blocks of living cells. They are polymers of the 20 naturally occuring amino acids listed in table 2. 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 14 is shown the amino acid Analine and its geometry.

Proteins II

Table: 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

Proteins I

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.

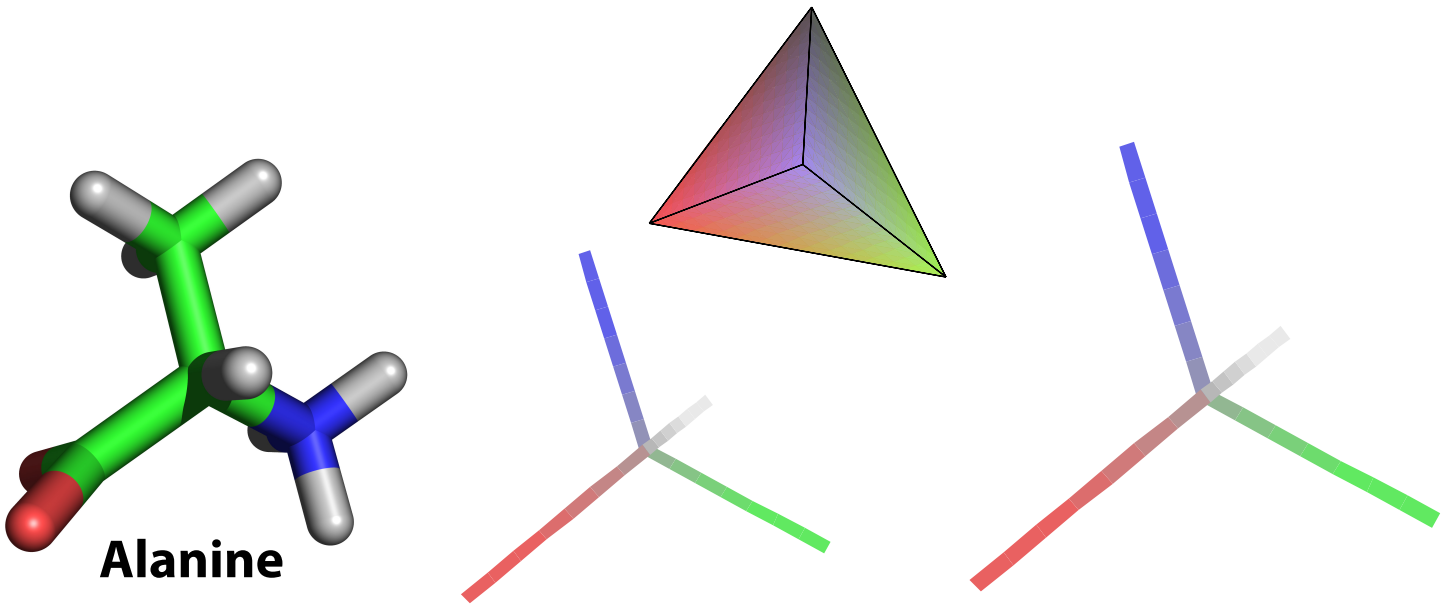


Figure: The amino acid Alanine. Note that the bond directions for carbon are the same as from the centroid of a tetrahedron to the vertices.

Proteins II

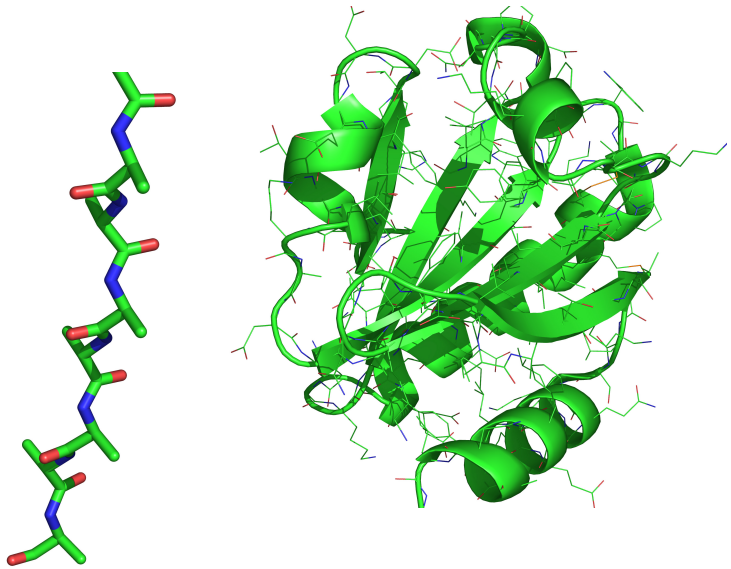


Figure: β-sheet. The protein thioredoxin contains a five-stranded beta sheet comprised of three parallel strands and three antiparallel strands. The entire protein is shown as a cartoon with the beta strands (three parallel strands and three antiparallel strands) colored red and alpha helices colored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

Proteins III

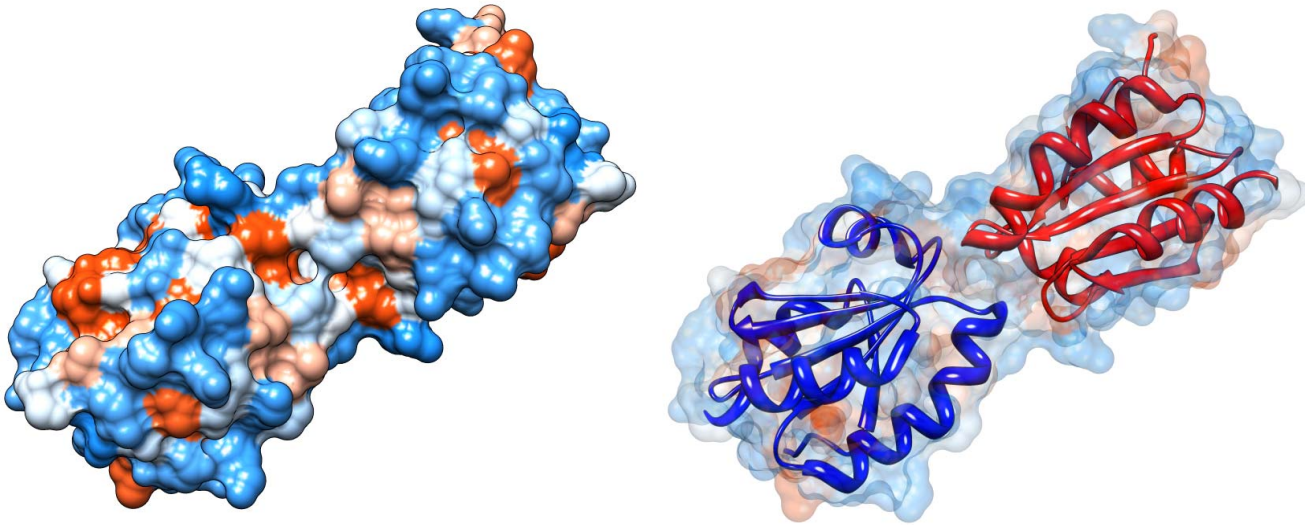


Figure: Protein 1f9m

There are four levels of architecture in proteins

- **Primary structure**: The sequence of peptide-bonded amino acids (as in the example: RSDAEPHYLPQLRKDILEVICKYVQIDPEMVTVQLEQKDGDISILEL-NVTLPEAEELK). 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 α-helices and β-sheets. The α-helix-complex has already been studied in a previous section. In a β-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 β-sheet are between different segments of polypeptide. An example of one strand of a β-sheet is shown in figure 15.
  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.

Proteins V

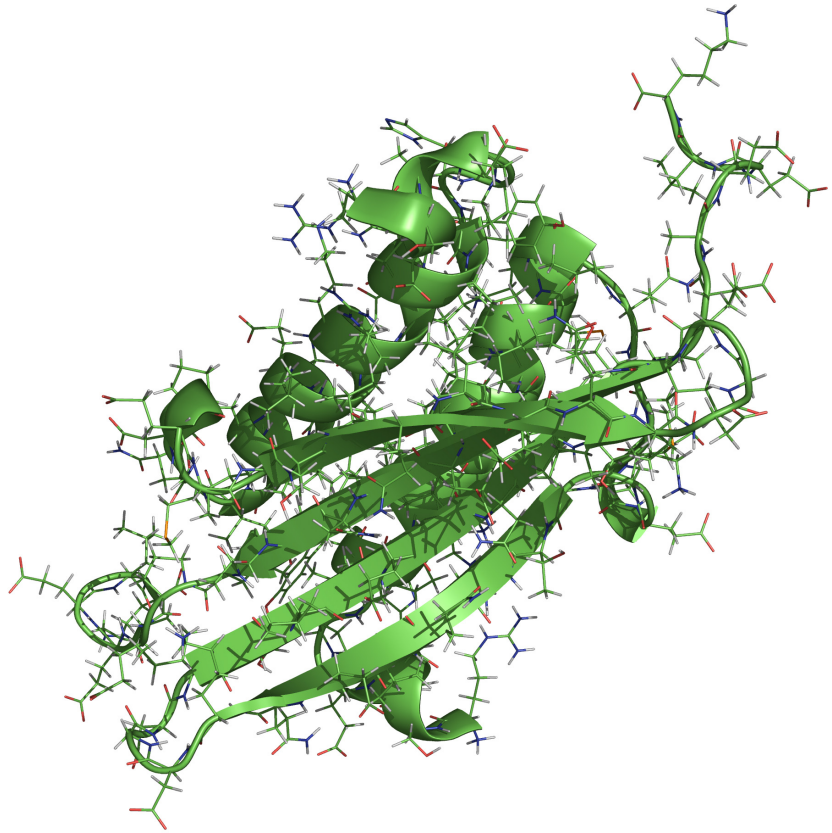


Figure: MinE protein showing α-helices and β-sheets.

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.

Protein Folding I

The long-standing question is: how do proteins fold? A protein folds due to the angles ϕ and ψ 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

Protein Folding II

Folding is an interesting problem because it involves mathematical modeling and numerical analysis. It is a 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.

Protein Folding IV

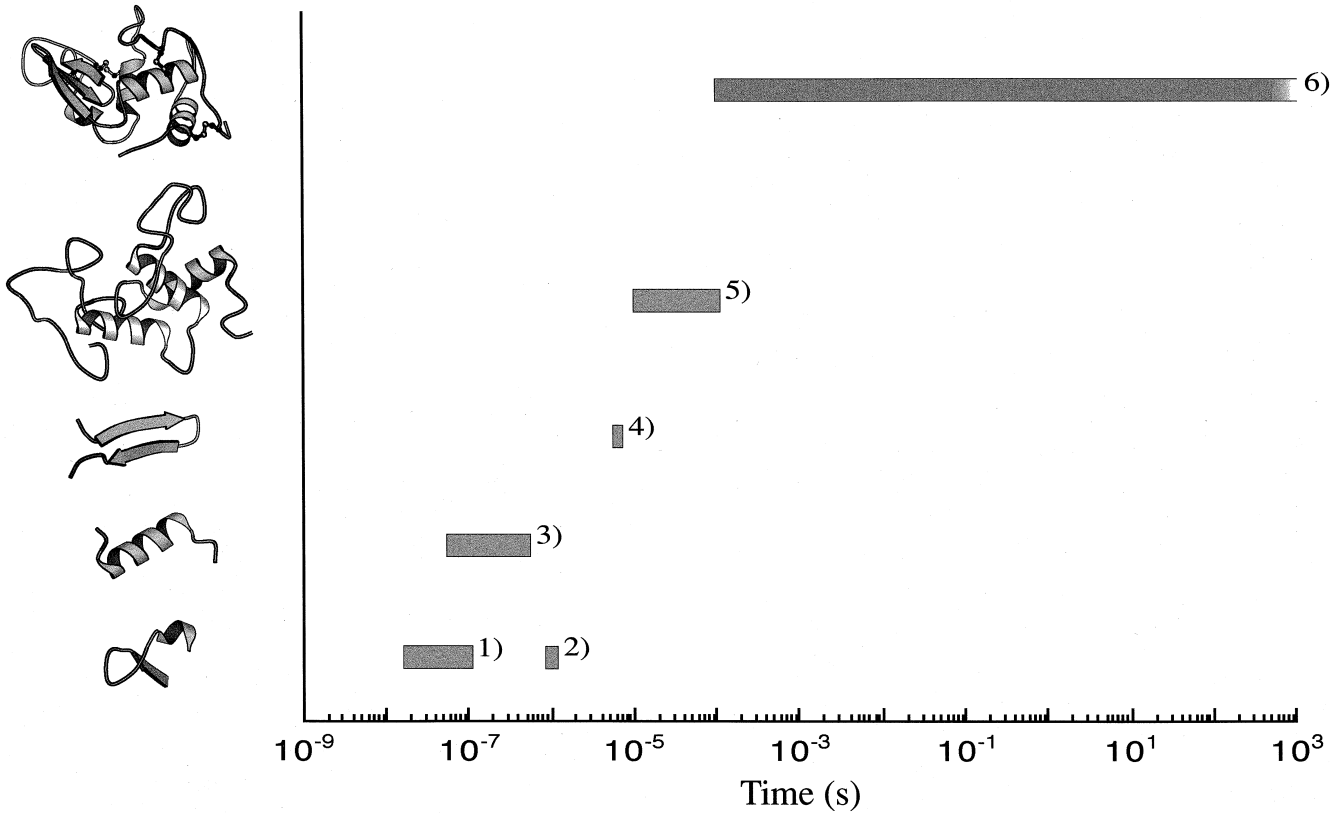


Figure: Time scales for the formation of structural elements in protein folding (Taken from O. Bieri and T. Kiefhaber, Biol. Chem. 80, 923-929 (1999)

Lattice Protein Models I

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 Λ 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 19. 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.

Lattice Protein Models II

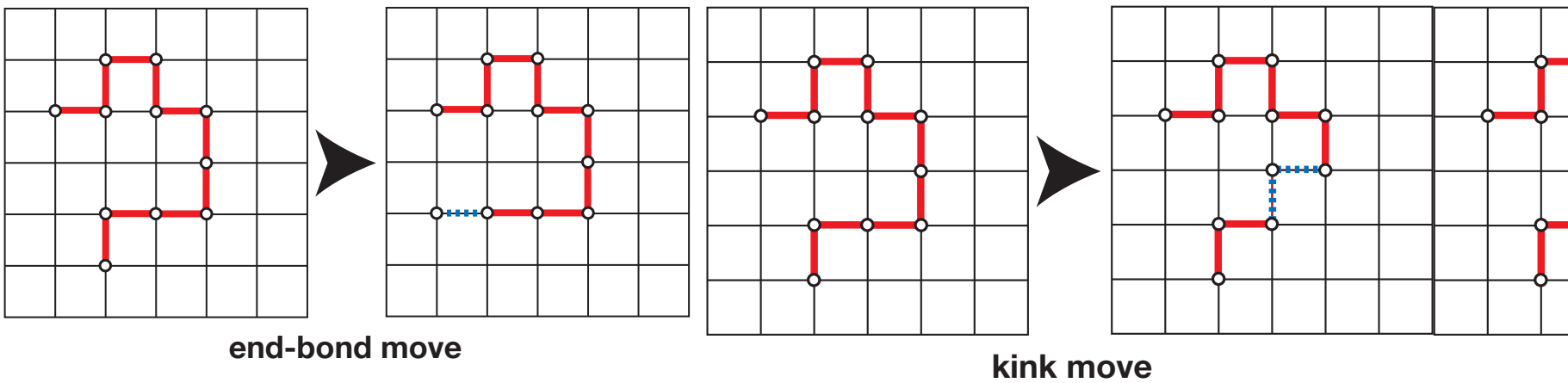


Figure: Possible move to change the conformation of a self-avoiding random walk (SAW)

Lattice Protein Models III

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 20) depends thus on the number of non-adjacent hydrophobic amino acids that occupy adjacent grid points in the lattice.

Lattice Protein Models IV

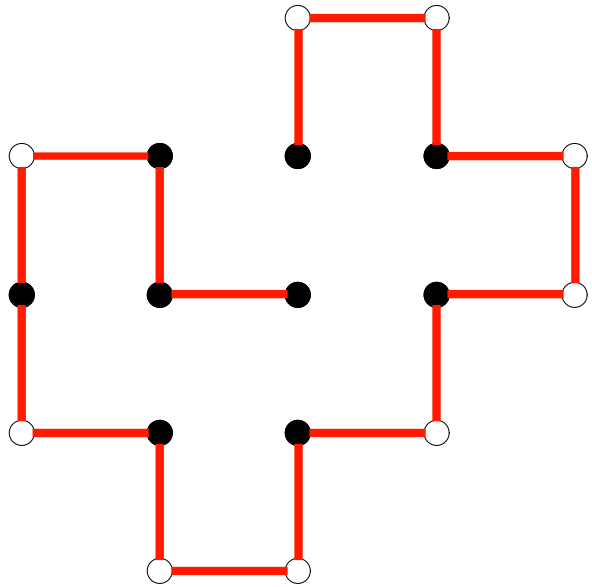


Figure: Conformation in the HP model. The black dots denote the hydrophobic acids.

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 hydrophyllic amino acid (see figure 21). The hydrophobic force is presumed to be dominant. For the interaction energy we take the values as shown in table 3.

Table: Energy in the HP Model

	H	P
H	-1	0
P	0	0

Lattice Protein Models VI

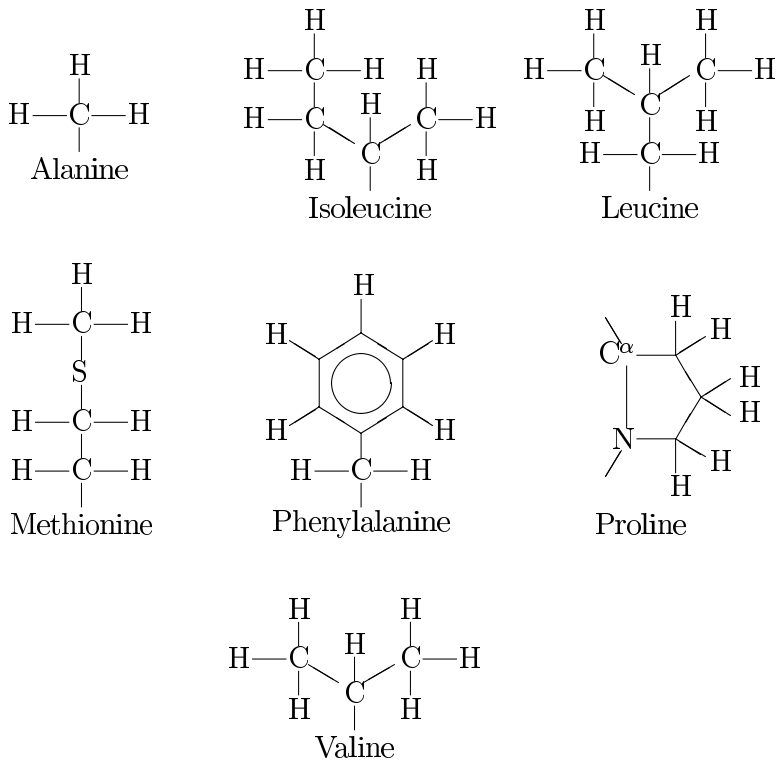


Figure: The hydrophobic amino acids

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

$$w : (1...|s|) \rightarrow Z^d \tag{94}$$

and the energy by

$$E = \sum_{1 \leq i < j \leq N} B_{i,j} \delta(\mathbf{R}_i, \mathbf{R}_j) \quad , \tag{95}$$

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

$$H = \sum_{i<j} \epsilon_{i,j} \left[\delta(|\mathbf{R}_i - \mathbf{R}_j| - \sigma) - \delta_{j-\mathbf{1},i} \right] \tag{96}$$

shows that we are dealing with a model that falls into the class of the random heteroployer models (see **??**). Here σ 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 22 shows a conformation which is very highly designable.

Lattice Protein Models IX

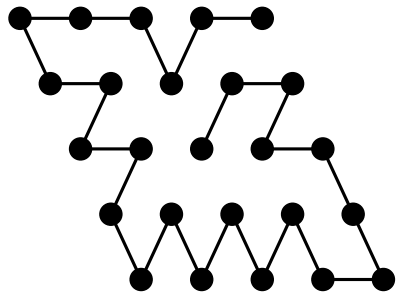


Figure: A conformation which is highly designable within the HP-model

DNA Models I

- DePablo Model
- Martini Model

RNA Models I

Chromatin I

- Two-Angel Model
- Solonoid

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 [24]. Whereas the structure of the core particle has been resolved up to high atomic resolution [25], 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$100mM), 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 [26–28]. Increasing the salt concentration decreases the entry-exit angle α of the stem as it reduces the electrostatic repulsion between two strands.

Chromatin II

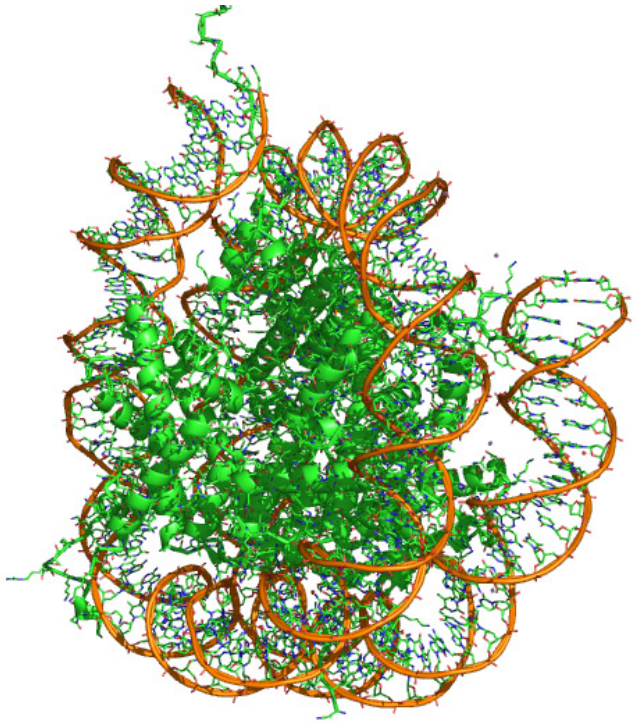


Figure: Nucleosome

Chromatin III

Figure: Histone H1

Chromosomes I

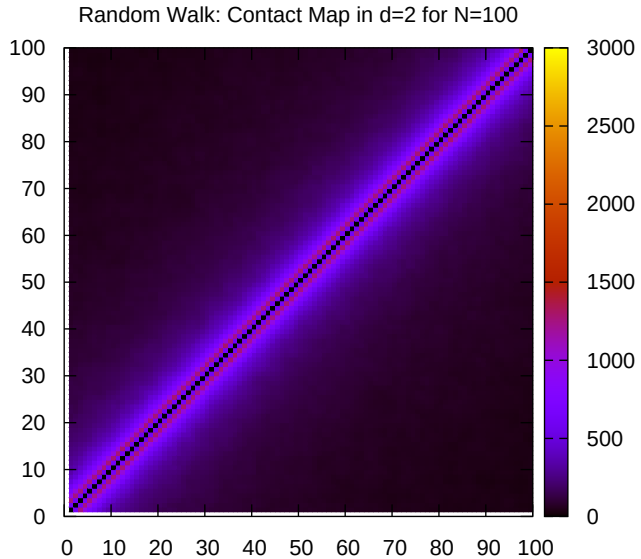


Figure: Contact map of a random walk of length $N = 100$ in $d = 2$.

Chromosomes II

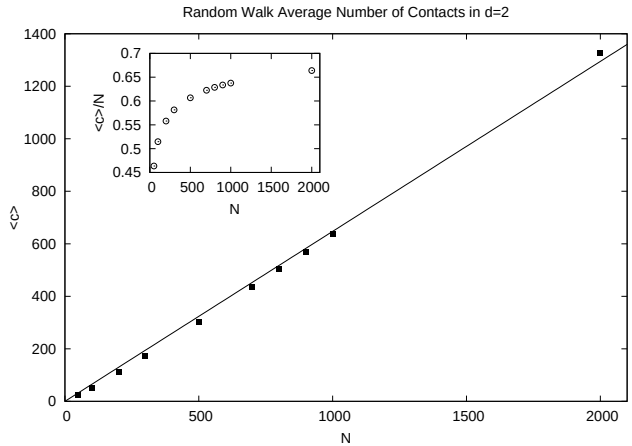


Figure: Average number of contacts $< c >$ as a function of length of a random walk in $d = 2$

Excercises I

Exercise 1: Find all possible random walks without self-intersections on the square lattice for length N=1,2,3, . . . 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 canditate is drawn independent of the current state. Repeat the above excercise and compare the auto-correlation.

Exercise 5: *Rosenbluth-Rosenbluth*

Exercise 6: *Configurational bias Monte Carlo*

Exercise 7: *Reweighting Monte Carlo*

Excercises II

Exercise 8: 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 9: **Peyrard-Bishop model of DNA** The melting of DNA can be approached from different point of view. We start from the Hamiltonian [**?**]

$$H = \sum_{i=1}^{N} \{ \frac{J}{2}(x_i + 1 - x_i)^2 + V(x_i)) \} \tag{97}$$

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.

Bibliography I

[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. Bellemanns. *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] P. 2003. J. Flory, Statistical Mechanics of Chain Molecules, Interscience, New York, 1969.

[7] 12 W. J. Orr, Trans. Faraday Soc. **43**. 1947. 1947.

[8] J. M. Hammersley and 23 K. W. Morton, J. Roy. Stat. Sot. B16. 1954. 1954.

[9] J. M. Hammersley and 23 K. W. Morton, J. Roy. Stat. Sot. 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.

Bibliography II

[14] pp. 44-55 Levinthal J. Chem. Phys. Vol. 65. A. pages 271–277, 1968. D. Mac Kerell 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, pagesJohn Wiley and Sons: Chichester, 1998.

[15] N. Madras and A.D Sokal. The pivot algorithm: a highly ecient 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 polypetide 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.

Bibliography III

[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] C. J. Geyer and 909-920 E. A. Thomposon, J. Am. Statis. Ass **90**. Khrapunov, s. pages 213–222, 1995. N., A.I. Dragan, A.V. Sivolob and A.M. Zagariya. 1997. Biochim. Biophys. Acta. 1351:.

[25] C. J. Geyer and 909-920 E. A. Thomposon, J. Am. Statis. Ass **90**. Luger et al. pages 251–260, 1995. 1997. Nature. 389:.

[26] 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.

[27] 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:.

[28] 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.

Index I