Chapter 2

Macromolecules

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 macromoluces? 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.

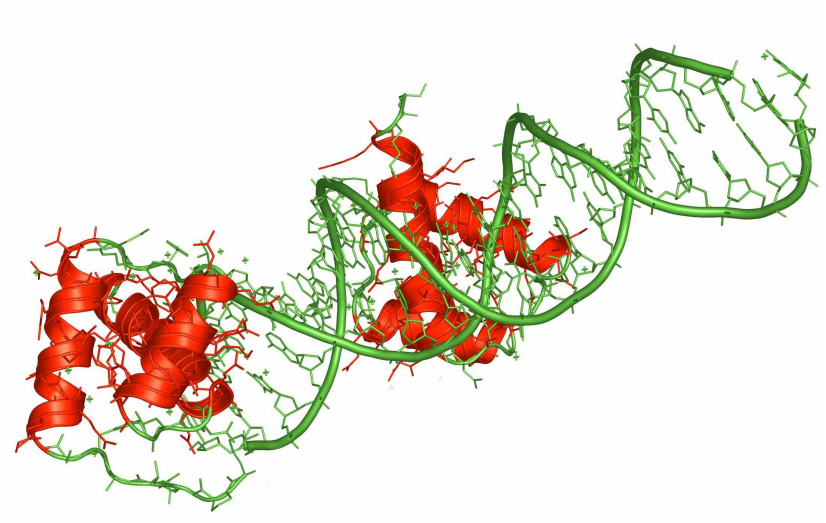

Figure 2.1: HOMEODOMAIN DNA COMPLEX DROSOPHILA MELANOGASTER

2.1 General Properties of Macromolecules

Important Concepts: degree of polymerization, chain length, contour length, bond angle, torsional angle polymer conformation

The kind of biological macromolecules we are focusing on in this chapter is shown in figure 2.1. The three-dimensional structure appears to be quite rich and complex exhibiting a double helix and other substructures. To understand this structure we need to describe the macromolecule (or polymer) using coordinates and angles between the units. Before doing so we first take a closer look at the variety of polymers that we will encounter here.
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.

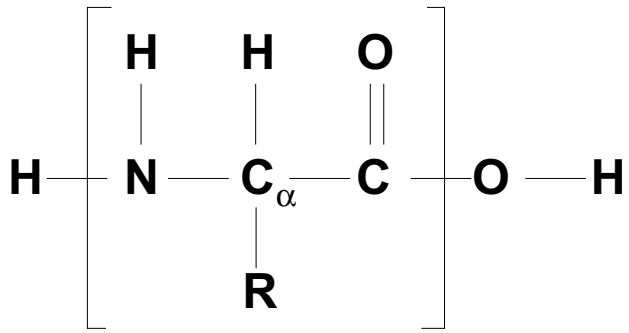

Figure 2.2: Amino acid

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.
A polymer chain of amino acid residues (see figure 2.2) is called a *polypeptide*. The proteins contain 20 different residues, with side chains having 1 to 18 atoms. The residues are abbreviated with three identifying letters of the corresponding amino acid:

```
Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile,
Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Valg
```

Almost all amino acids have the structure shown in figure 2.2. The side chain labelled R is characteristic for a specific amino acid. The only exception to the above structure is proline.
A *carbohydrate* is an organic compound that is composed of atoms of carbon, hydrogen and oxygen. Some carbohydrates are relatively small molecules. In this context the most important is glucose. Glucose is a typical *monosaccharide*. Glucose has two important types of functional groups: a carbonyl group and hydroxyl groups. Glucose exists mostly in ring structures.
Monosaccharides can polymerize by elimination of the elements of water between the anomeric hydroxyl and a hydroxyl of another sugar. This is called a *glycosidic bond*. Since most monosaccharides have more than one hydroxyl, branches are possible (see below for the possible macromolecular architectures.)
Nucleotides consist of three parts: Phosphate, Monosaccharide and a base. There are four dominant bases

- adenine (purine)
- cytosine (pyrimidine)
- guanine (purine)
- uracil (in ribonucleotides) or thymine (in deoxyribonucleotides)

Nucleotides polymerize to yield nucleic acids. These are abbreviated by

A,C,G,T.

We thus consider a macromolecule to be composed of a sequence of monomers with bonds. The bond types for macromolecules can be

- covalent or
- ionic (coordinative) bonds / metallic (electron deficiency) bonds.

For the covalent bonds we shall often assume a potential of the form

$$H_{\text{bond length}} = \frac{k_b}{2} \sum_{\text{bonds}} (r-b)^2 \quad , \tag{2.1}$$

where r denotes the length of the bond and k_b is a spring constant. For convience we often set $b = 0$. Frequently, one also uses an anharmonic potential based on a finitely extensible nonlinear elastic potential (abbreviated FENE) [21, 22]

$$H_{\text{FENE}} = \begin{cases} -\frac{1}{2} r_0^2 \ln\left[1-(r/r_0)^2\right] & r \le r_0 \\ \infty & r > r_0 \end{cases} \quad , \tag{2.2}$$

where r_0 is a finite extensibility (see figure 2.3). The FENE potential is harmonic at its minimum but the bonds cannot be stretched beyond a maximum length determined by r_0.

From the theoretical side we are interested in *homopolymers*, which are composed of only one sort of monomer and *copolymers* with more than two sorts of monomers. Homopolymers consist of monomers of the same type; copolymers have different repeating units. An example of a homopolymer is Polyhydroxyalkanoate (PHA). It is a biodegradable polyester synthesized and accumulated in many bacterial species as their intracellular carbon and energy storage compound.

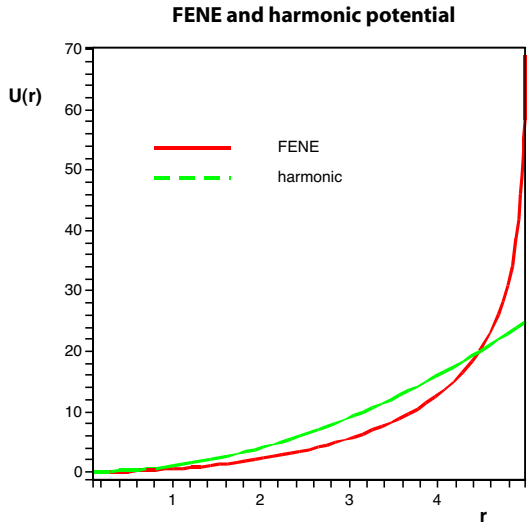

Figure 2.3: FENE (finitely extensible nonlinear elastic potential) and harmonic potential

Depending on the arrangement of the types of monomers in the polymer chain, we have the following classification:

- In *random copolymers* two or more different repeating units are distributed randomly.
- *Alternating copolymers* are made of alternating sequences of the different monomers.
- In *block copolymers* long sequences of a monomer are followed by long sequences of another monomer.
- *Graft copolymers* consist of a chain made from one type of monomers with branches of another type.

The molecular architecture can be quite complex (as we have already noticed in figure 2.1) ranging from linear, branched to crosslinked as shown in figure 2.4. A linear polymer consists of a long chain of monomers. A branched polymer has branches covalently attached to the main chain.

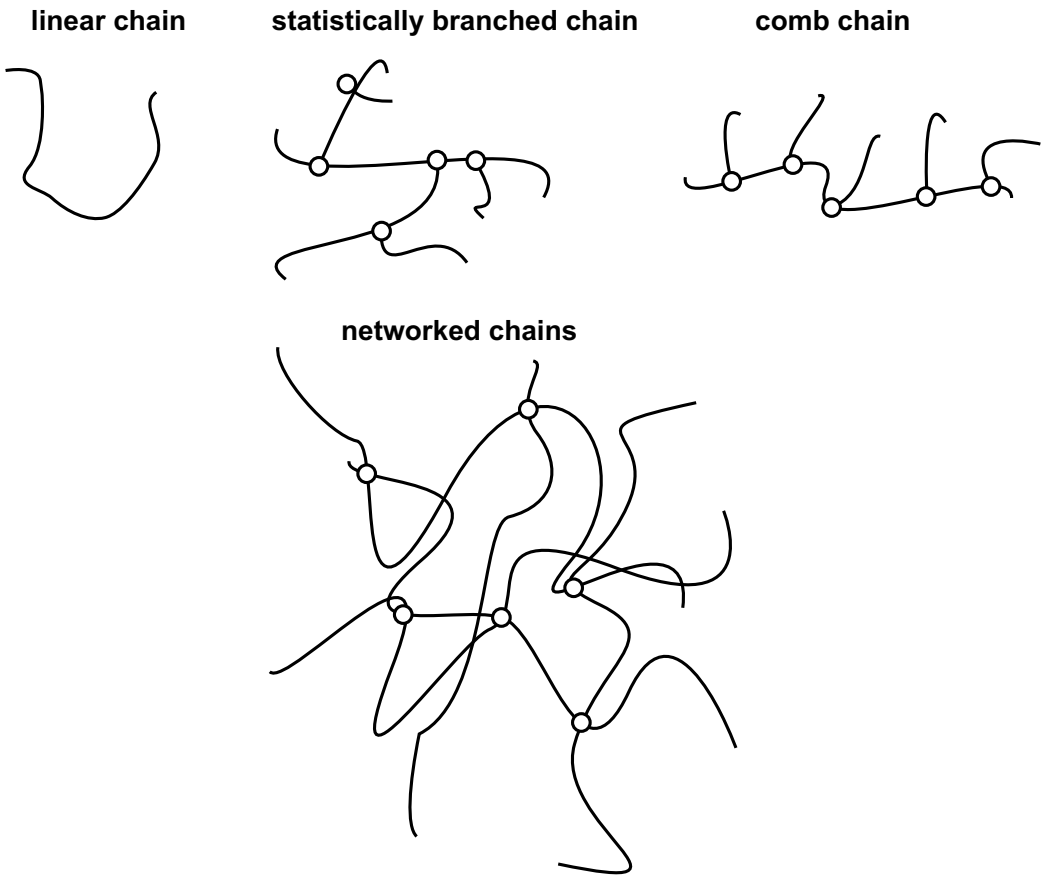

Figure 2.4: Various chain structures. The circles mark connection points between for example a main chain and side chains (comb chain).

Example 2.1.1 (Glycogen)
Glycogen is a highly branched macromolecule comprised of about 2000 glucose units which stores glucose units in animals.

Cross-linked polymers have monomers of one chain covalently bonded with monomers of another chain. Cross-linking results in a three-dimensional network.
The first measure describing the physics of 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

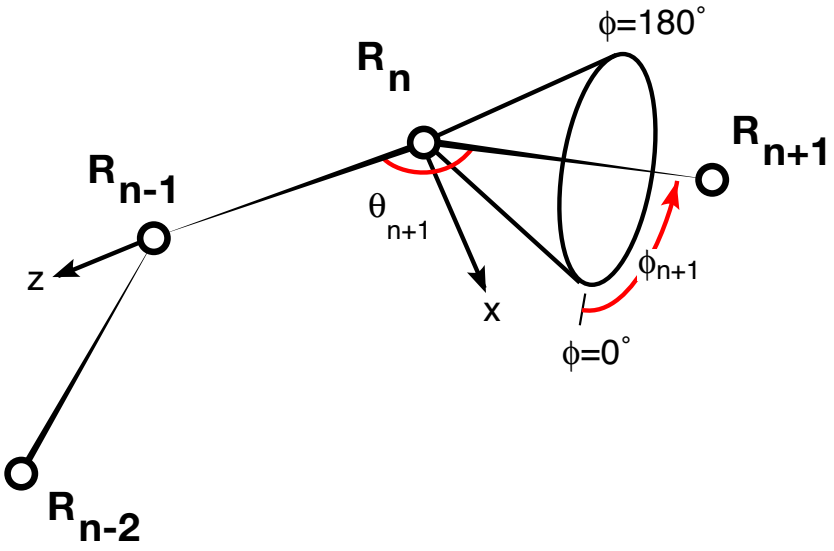

Figure 2.5: Angle definition

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 2.5. Here we have introduced the polar angle as the *bond angle*, i.e. the angle between two adjacent bonds.

$$H_{\text{bond angle}} = \frac{k_\theta}{2} \sum_{\text{angles}} \left(\cos\theta_{\text{angle}} - \cos\theta_0\right)^2 \quad . \tag{2.3}$$

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 [35].

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 2.6 is shown the possible potential energy with the corresponding labeling (the angle and labelling is also listed in table 2.1).

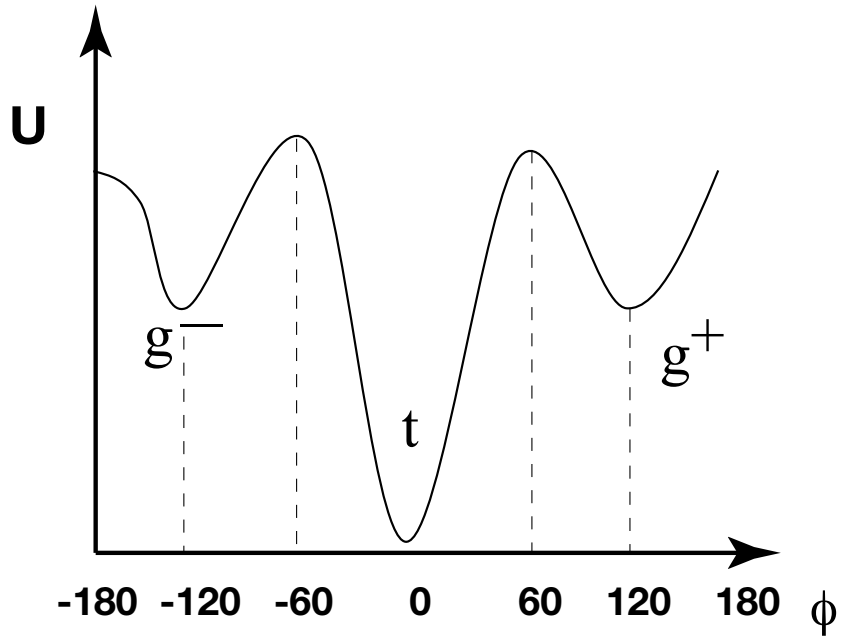

Figure 2.6: A typical torsion potential

$$H_{\text{torsion}} = \sum_{\text{dihedral angle}} \left[\frac{k_1}{2}(1 - \cos\phi) + \frac{k_2}{2}(1 - \cos 2\phi) + \frac{k_3}{2}(1 - \cos 3\phi) \right] \quad . \tag{2.4}$$

Table 2.1: Name convention for specific angles and their property

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

Further Reading 2.1.1 (Special Conformations)

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.

2.1.1 Freely-jointed chain

Important Concepts: end-to-end distance, radius of gyration, Kuhn length, persistence length, scaling

The simplest model of polymer conformation treats the molecule as a chain of rigid subunits, joined by perfectly flexible hinges [20]. In this *freely jointed chain model* the chain is made up of N links, each of length b with no excluded volume. Thus it corresponds to a random walk where each step has length b. 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_BT. 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.
The joints of the chain are at positions $\mathbf{R}_n$ and are joined by the link vectors

$$\mathbf{r}_n = \mathbf{R}_n - \mathbf{R}_{n-1} \quad . \tag{2.5}$$

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

$$\mathbf{R}_e = \mathbf{R}_0 - \mathbf{R}_N = \sum_{n=1}^{N} \mathbf{r}_n \tag{2.6}$$

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

$$< \mathbf{r}_n >= 0 \tag{2.7}$$

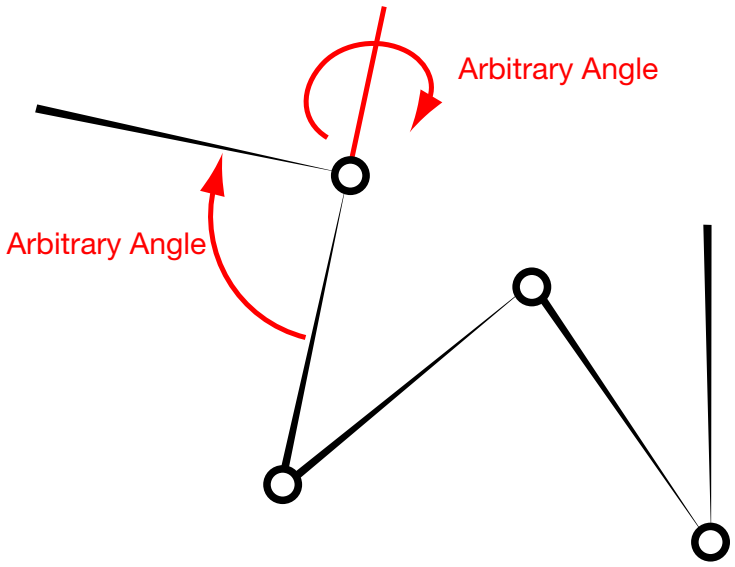

Figure 2.7: Freely-jointed chain

and

$$< \mathbf{r}_n \mathbf{r}_m >= \delta_{nm} b^2 \quad . \tag{2.8}$$

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

$$< \mathbf{R}_e >=< \sum_{n=1}^{N} \mathbf{r}_n >= \sum_{n=1}^{N} < \mathbf{r}_n >= 0 \tag{2.9}$$

and

$$< \mathbf{R}_e \mathbf{R}_e > = < \sum_{n=1}^{N} \sum_{m=1}^{N} \mathbf{r}_n \mathbf{r}_m > \quad (2.10)$$

$$= \sum_{n=1}^{N} \sum_{m=1}^{N} < \mathbf{r}_n \mathbf{r}_m > \quad (2.11)$$

$$= \sum_{i=1}^{N} b^2 \quad (2.12)$$

$$= Nb^2 \quad . \quad (2.13)$$

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

$$< R_e^2 > \propto N \quad (2.14)$$

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

$$< R_e^2 > \propto N^{2\nu} \quad (2.15)$$

introducing the exponent ν. Thus $\nu = 1/2$ here.
DNA is much stiffer than an alkane chain. Hence DNA has a much larger $< R_e^2 >$ 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

$$< R_e^2 > = N_K l_K^2 \quad (2.16)$$

$$L_c = N_K l_k \quad , \quad (2.17)$$

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.

$$\xi_p = l_k/2 \quad L_c >> l_k \tag{2.18}$$

and hence

$$< R_e^2 > \;=\; 2N_p\xi_p^2 \quad L_c >> l_k \tag{2.19}$$

$$L_c \;=\; N_p\xi_p \tag{2.20}$$

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 $\xi_p \approx 50nm$. 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) and we must have

$$P(\mathbf{R}_e) = \left(\frac{3}{2\pi N b^2}\right)^{-3/2} \exp\left(-\frac{3\mathbf{R}_e^2}{2Nb^2}\right) \tag{2.21}$$

which is properly normalized.

Note that the conformations of the polymer are random coils. A typical conformation of the chain is shown in figure 2.8.

Further Reading 2.1.2 (On length scales)

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

$$R_g = \sqrt{\frac{1}{N+1}\sum_{n=0}^{N} < (\mathbf{R}_{cm} - \mathbf{R}_n)^2 >} \tag{2.22}$$

with the center of mass

$$\mathbf{R}_{cm} = \frac{1}{N+1}\sum_{n=0}^{N}\mathbf{R}_n \quad . \tag{2.23}$$

For the freely jointed chain model we obtain

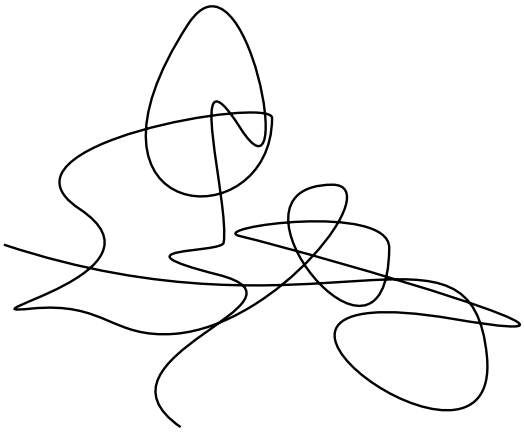

Figure 2.8: A typical random coil conformation

$$R_g = \left(\frac{N}{6}\right)^{1/2} b \quad . \tag{2.24}$$

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

$$\frac{< R_g^2 >}{< R_e^2 >} = 6 \quad . \tag{2.25}$$

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

$$\Delta S(\mathbf{R}_e) = k_B \log \frac{P(\mathbf{R}_e)}{P(0)} \tag{2.26}$$

from which we obtain the free energy difference

$$\Delta F_e = -T\Delta S = \frac{3}{2}\frac{k_B T}{Nb^2} R^2 \quad . \tag{2.27}$$

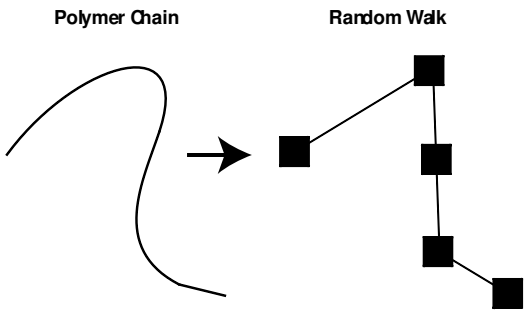

Figure 2.9: From a continuous to a lattice description

Further Reading 2.1.3 (Lattice Analog)

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.

For simplicity we assume a simple square lattice Λ with a coordination number q. Then the partition function is given by

$$Z_N = q^N \quad . \tag{2.28}$$

If we constrain ourselves to a simple square lattice then this imposes the restriction of a single bond length and the angles ± 90 and 180 degrees. The random walk starts at the origin $(0, 0)$ making steps of length 1. Assume the random walk arrives at a lattice point $p = (k_1, k_2)$ at the step n, then one of the four neighbors $(k_1 \pm 1, k_2), (k_1, k_2 \pm 1)$ of p is selected with equal probability $1/4$.

We can compute how much the mean squared displacement has grown after a step

$$< R_e^2 >_{N+1} - < R_e^2 >_N =$$
$$\frac{1}{4}\left[(k_1+1)^2 + (k_1-1)^2 - 2k_1^2 + (k_2+1)^2 + (k_2-1)^2 - 2k_2^2\right] = 1 \quad (2.29)$$

Hence

$$< R_e^2 >_N = N \quad . \quad (2.30)$$

To calculate the radius of gyration assuming random interaction between the monomers we note that the probability of having a radius R_g must be gaussian. Let c_N denote the number of lattice polymers of length N. Then the number of random walks with radius R_g ($c_N(R)$) is given by

$$c_N(R_g) = c_N P_N(R_g) \quad (2.31)$$

and hence for the entropy

$$S_N(R_g) = \ln c_N(R_g) = -\frac{a}{N} R_g^2 + c \quad (2.32)$$

with constants a and c.
The energy can be considered as arising from random contacts

$$U_N(R_g) = b\frac{N^2}{R_g^2} \quad . \quad (2.33)$$

Then the free energy is given by

$$F_N(R_g) = U_N(R_g) - TS_N(R_g) \quad (2.34)$$

which must be minimized. This yields

$$R_g \propto N^{3/4} \quad (2.35)$$

and thus $\nu = 3/4$ in this approximation.