

**Figure 2.20:** Helical structure found in polypeptides. To better identify the helical structure the right picture shows a cartoon of the helix.

Hydrophobic interactions have strengths of a few  $k_B T$  and are comparable in energy to hydrogen bonds.

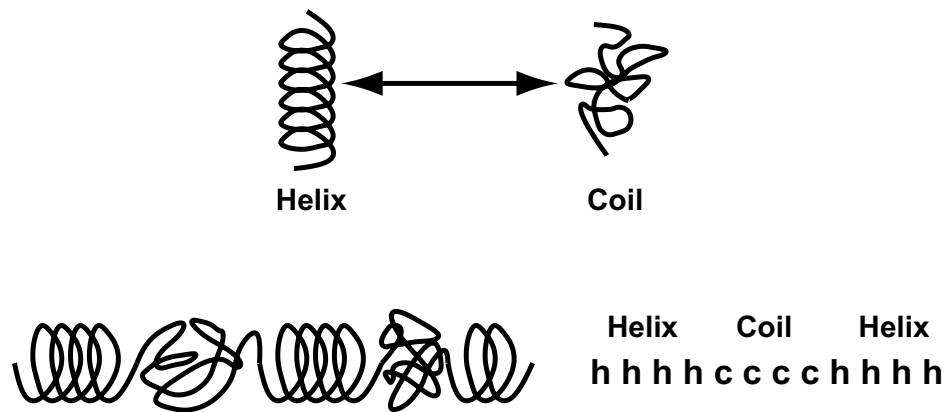
## 2.3 Helix-Coil Transition

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 stabilises 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.

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 [36, 37, 38].

To describe the macromolecule in terms of helical and non-helical parts, we de-



**Figure 2.21:** Mapping of the helix-coil transition onto a sequence of symbols

note by  $h$  a helical monomer and by  $c$  a coil monomer (see later for the analogy with the Ising model [39]). 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

$$\dots h h c c c h c \dots \quad (2.172)$$

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

$$u_i = \exp\{-\beta E_i(c)\} \quad (2.173)$$

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

$$v_i = \exp\{-\beta E_i(h)\} \quad (2.174)$$

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

$$Z_N = \sum_{\{h,c\}} e^{-\beta E\{h,c\}} \quad (2.175)$$

$$= \sum_{i,j} \prod_{i,j} u_i v_j \quad . \quad (2.176)$$

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

$$i_0, j_1, i_1, \dots, j_M, i_M, j_0 \quad , \quad (2.177)$$

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

$$M \leq \lfloor N/2 \rfloor \quad (2.178)$$

with the constraint

$$\sum_{k=0}^M (i_k + j_k) = N \quad . \quad (2.179)$$

Hence we can write

$$Z_N = \sum_{M=0}^{\lfloor N/2 \rfloor} \sum_{\{i_k, j_k\}} \prod_{k=0}^M u_{i_k} v_{j_k} \quad . \quad (2.180)$$

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

$$Z_N \approx q_{\text{eff}}^N \quad \text{for } N \gg 1 \quad , \quad (2.181)$$

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

Let us now look at the generating function

$$\Gamma(x) = \sum_{N=0}^{\infty} Z_N x^{-N} \quad . \quad (2.182)$$

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

$$\Gamma(x) < \infty \quad x > q_{\text{eff}} \quad (2.183)$$

$$1/\Gamma(x) = 0 \quad x = q_{\text{eff}} \quad . \quad (2.184)$$

Hence the partition function is the largest root of eq 2.184. So, let us look at the  $\Gamma$  in more detail

$$\Gamma(x) = \sum_{N=0}^{\infty} x^{-N} \sum_{M=0}^{\lfloor N/2 \rfloor} \sum_{\{i_k, j_k\}} \prod_{k=0}^M u_{i_k} v_{j_k} \quad (2.185)$$

$$= \sum_{M=0}^{\infty} \sum_{N=2M}^{\infty} \sum_{\{i_k, j_k\}} \prod_{k=0}^M u_{i_k} x^{-i_k} v_{j_k} x^{-j_k} \quad (2.186)$$

$$= \sum_{M=0}^{\infty} \sum_{i_0=0}^{\infty} \frac{u_{i_0}}{x^{i_0}} \sum_{j_0=0}^{\infty} \frac{v_{j_0}}{x^{j_0}} \prod_{k=1}^M \sum_{i_k=0}^{\infty} \frac{u_{i_k}}{x^{i_k}} \sum_{j_k=0}^{\infty} \frac{v_{j_k}}{x^{j_k}} \quad . \quad (2.187)$$

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

$$U(x) \equiv \sum_{i=1}^{\infty} u_i x^{-i} \quad (2.188)$$

$$V(x) \equiv \sum_{j=1}^{\infty} v_j x^{-j} \quad (2.189)$$

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

$$\Gamma(x) = U_0 V_0 \sum_{k=0}^{\infty} (UV)^k \quad (2.190)$$

$$= U_0 V_0 \frac{1}{1 - UV} \quad . \quad (2.191)$$

$\Gamma(x)$ ,  $U(x)$  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

$$U_0 V_0 |_{x=q_{\text{eff}}} \neq 0 \quad (2.192)$$

we have

$$U(q_{\text{eff}}) V(q_{\text{eff}}) = 1 \quad (2.193)$$

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

Corresponding to these three states we need statistical weights

$$\text{coil} - u \quad (2.194)$$

$$\text{h with h - bond} - w \quad (2.195)$$

$$\text{h without h - bond} - v \quad (2.196)$$

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

$$\text{coil} - u/u = 1 \quad (2.197)$$

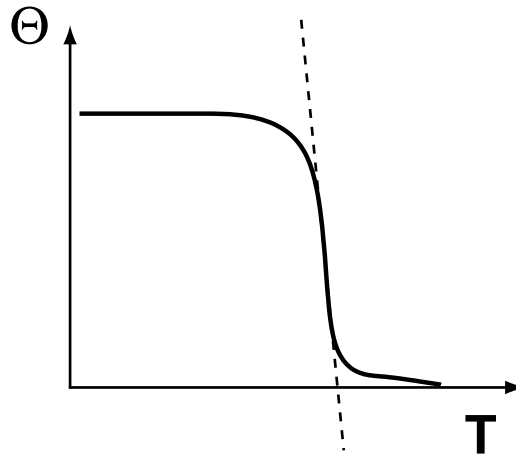
$$\text{helix with h - bond} - w/u = s \quad (2.198)$$

$$\text{helix without h - bond} - v/u = \sigma^{1/2} \quad (2.199)$$

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

$$\begin{array}{lll} c - \text{sequence} & u_i = u^i & 1 \\ h - \text{with h - bond} & v_1 = v & v_1 = \sigma^{1/2} \\ h - \text{without h - bond} & v_j = v^2 w^{j-2} & v_j = \sigma s^{j-2} \end{array} \quad (2.200)$$

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 2.22:** Dependence of the order parameter on the temperature for the helix-coil transition

With the above defined statistical weights and using eq 2.176 we have

$$Z_N = \sum_{ij} \prod u_i v_j = \sum_{ij} \prod \sigma s^{j-2} \quad (2.201)$$

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

$$\theta = \frac{1}{N-2} \frac{\partial \ln Z_N}{\partial \ln s} \quad (2.202)$$

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

$$\theta = \frac{1}{N} \frac{\partial \ln q_{\text{eff}}^N}{\partial \ln s} = \frac{s}{q_{\text{eff}}} \frac{\partial q_{\text{eff}}}{\partial s} \quad (2.203)$$

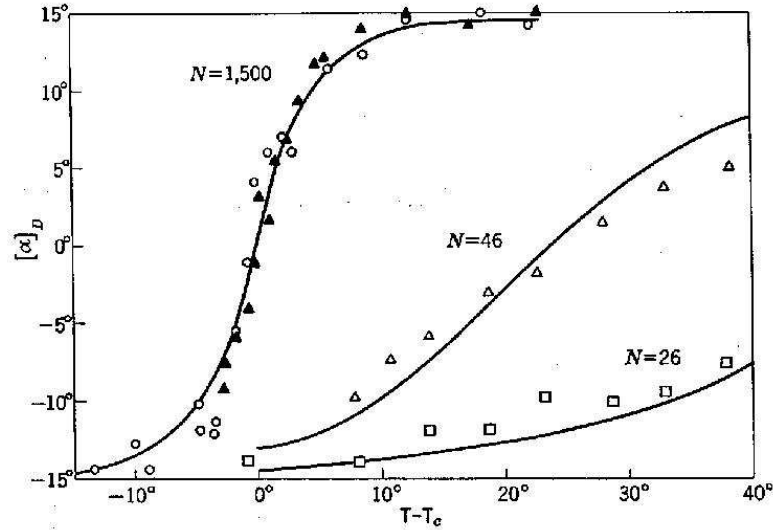
We determine  $q_{\text{eff}}$  from eq 2.193

$$q_{\text{eff}}^3 - q_{\text{eff}}^2(u+w) + q_{\text{eff}}(wu-uv) + uvw + v^2u = 0 \quad (2.204)$$

with the solution

$$q_{\text{eff}} = \frac{1}{2} \left\{ w + u + \sqrt{(w-u)^2 + 4uv^2/w} \right\} \quad (2.205)$$

and for  $\theta$



**Figure 2.23:** Experimental results for the helix-coil transition

$$\theta = \frac{1}{2} \left\{ 1 + \frac{s-1}{\sqrt{(s-1)^2 + 4\sigma s}} \right\} . \quad (2.206)$$

In figure 2.22 is shown the qualitative result for the order parameter  $\theta$ . A comparison to the experimental findings is shown in figure 2.23.

We can make contact with the Ising model by setting

$$\sigma = e^{-4J/k_B T} \quad (2.207)$$

$$s = e^{2H/k_B T} \quad (2.208)$$

to find

$$\theta = \frac{1}{2} \left\{ 1 + \frac{\sinh(H/k_B T)}{\sqrt{\sinh^2(H/k_B T) + e^{-4J/k_B T}}} \right\} . \quad (2.209)$$

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



**Figure 2.24:** Poland-Scheraga-model of the DNA-melting transition

$$\langle m \rangle = 2\theta - 1 = \frac{\sinh(H/k_B T)}{\sqrt{\sinh^2(H/k_B T) + e^{-4J/k_B T}}} . \quad (2.210)$$

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

## 2.4 DNA Melting

DNA melting refers to the dissociation of the two strands of the double helix by an increase of temperature. It is a reversible phase transition. Dissociation can occur also through a change of pH.

The melting or denaturation of DNA is a thermodynamic phase transition. The order of the transition is still debated due to the effect of 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 [40] and the zipper-model [41].

The Poland-Scheraga-model considers the DNA molecule as composed of an alternating sequence of bound and denatured states as depicted in figure 2.24. 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 denatured segment or loop is entropically favored over a bound one.

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



$$\begin{array}{lll}
\text{coil sequence in a loop} & \delta(i)\sigma & \\
\text{coil sequence at the end of the chain} & 1 & (2.211) \\
\text{helix sequence} & s^j &
\end{array}$$

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

$$s^j = \exp(-jE/k_B T) \quad . \quad (2.212)$$

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  $S(i)$

$$S(i)/k_B = \ln \delta(i) = ai - c \ln i + b \quad , \quad (2.213)$$

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

$$\delta(i) = e^{S(i)/k_B} \propto i^{-c} \approx \kappa^i / i^c \quad , \quad (2.214)$$

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

$$Z = \sum_{N=0}^{\infty} G(N)x^N = \frac{V_0(x)U_N(x)}{1 - U(x)V(x)} \quad , \quad (2.215)$$

with

$$U(x) = \sum_{i=1}^{\infty} \frac{\kappa^i}{i^c} x^i, \quad V(x) = \sum_{j=1}^{\infty} s^j x^j \quad (2.216)$$

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

$$\ln Z \simeq N \ln x_1. \quad (2.217)$$

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

$$U(x_1)V(x_1) = 1 \quad . \quad (2.218)$$

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

$$\theta = \frac{1}{1 + \frac{\sigma s}{x_1} \sum_{i=1}^{\infty} x_1^{-i} i^{1-c}} \quad . \quad (2.219)$$

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

$$\sum_{i=1}^{\infty} x_1^{-i} i^{-c} = \frac{x_1 - s}{\sigma s} \quad . \quad (2.220)$$

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

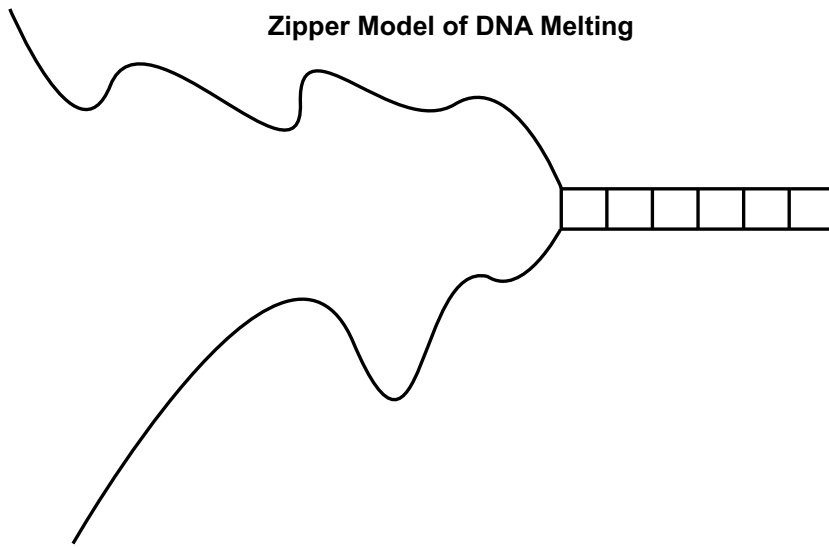
$$s_c = \frac{1}{1 + \sigma \zeta(c)} \quad , \quad (2.221)$$

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  $d > 4$ .



**Figure 2.25:** Kittel zipper model of the DNA-melting transition

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.

Kittels zipper-model [41] 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 2.25). 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

$$G^p e^{-p\epsilon/k_B T} \quad (2.222)$$

to the partition function. Thus

$$Z_N = \sum_{p=0}^N G^p e^{-p\epsilon/k_B T} = \frac{1 - x^N}{1 - x} \quad , \quad (2.223)$$

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

$$\langle \theta \rangle = \frac{\sum_p p x^p}{\sum_p x^p} = x \frac{d}{dx} \ln Z_N \quad (2.224)$$

$$= \frac{N x^N}{x^N - 1} - \frac{x}{x - 1} \quad (2.225)$$

which is shown in figure 2.26. We can expand the order parameter in the neighbourhood of the critical point  $x_c = 1$  using

$$\epsilon \equiv |x - x_c| \propto |T - T_c| \ll 1 \quad . \quad (2.226)$$

With this

$$\langle \theta \rangle = G \frac{d\epsilon}{dG} \frac{d}{d\epsilon} \ln Z_N \quad (2.227)$$

$$= \frac{1}{2} N \left( 1 + \frac{1}{6} N \epsilon - \frac{1}{360} N^3 \epsilon^3 + \dots \right) \quad . \quad (2.228)$$

For  $T = T_c$  we have

$$\frac{1}{N} \frac{d\langle \theta \rangle}{d\epsilon} = \frac{1}{12} N - \frac{1}{240} N^3 \epsilon^2 \quad (2.229)$$

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

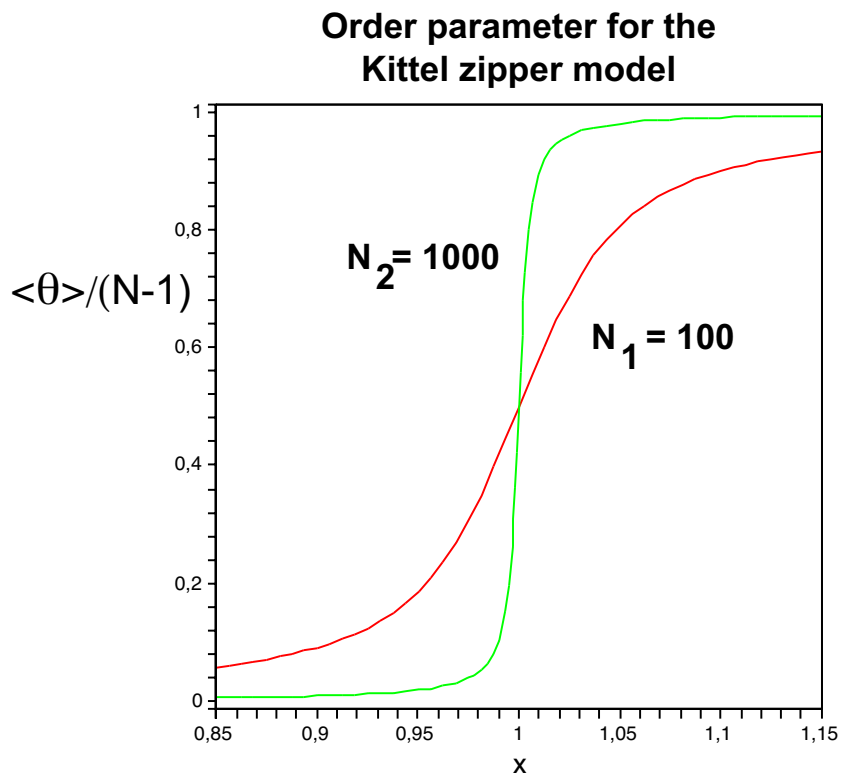
$$\frac{\langle \theta \rangle}{N} = \frac{1}{2} \quad . \quad (2.230)$$

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

$$T_c = \frac{\epsilon}{k_B} \ln G \quad . \quad (2.231)$$

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

Let us then look at the entropy



**Figure 2.26:** Kittel zipper-model of the DNA-melting transition

$$S = T \frac{\partial}{\partial T} \ln Z + \ln Z \quad (2.232)$$

$$= \left( x \ln \frac{G}{x} \right) \frac{\partial}{\partial x} Z + \ln Z \quad (2.233)$$

$$= \langle \theta \rangle \ln \frac{G}{x} + \ln(x^N - 1) - \ln(x - 1) \quad (2.234)$$

Hence

$$S \approx \langle \theta \rangle \ln G \quad (2.235)$$

for  $N\epsilon \ll 1$ . The entropy is proportional to the order parameter.

For the specific heat we find

$$C = T \frac{dS}{dT} \quad (2.236)$$

in the neighbourhood of  $T_c$

$$C \approx k_B (\ln G)^2 \frac{d\langle \theta \rangle}{d\epsilon} \quad (2.237)$$

$$\approx N k_B (\ln G)^2 \left[ \frac{1}{12} N - \frac{1}{240} N^3 \epsilon^2 + \dots \right] \quad (2.238)$$

Thus, the specific heat per bond diverges in the thermodynamic limit.

## 2.5 Polyelectrolyte

Polyelectrolytes are one of the least understood states of condensed matter, in contrast to neutral polymer solutions. Recall that a chemical compound composed of ions, in a solid, liquid or dissolved state is called a *electrolyte*. Such a system exhibits electrolytic conductivity and interionic interaction. The ions typically have charges of magnitude, equal to their *valency*  $z$  multiplied by the *electronic charge*  $e$ .

A polymer having sufficient ionic substituents along the chain to be water-soluble is called a *polyelectrolyte*. Thus, polyelectrolytes are polymers with ionizable groups that can dissociate in solution, leaving ions of one sign bound to the chain