Simple models and ideas in helix-coil transitions of DNA

What are the experimental observations? Take ***DNA*** as an *example* (actually peptides are somewhat simpler and easier to deal with, as we will see.

We can predict the **melting temperature, T_m,** of natural long DNA by the following table, which has been assembled from experimental data
Not only the *base pairing* is important, but the other *secondary* and *tertiary* structures are also crucial for predicting which structures of polynucleotides are present in solution at any particular temperature, ionic strength, pH, and presence of small molecular species and other macromolecules. It has become very apparent in the last few years, that the structure of RNA molecular species in solution controls to a great extent their *functional potentials*. Here are some of these "additional" structures, and their contributions to the overall free energy of a molecule.

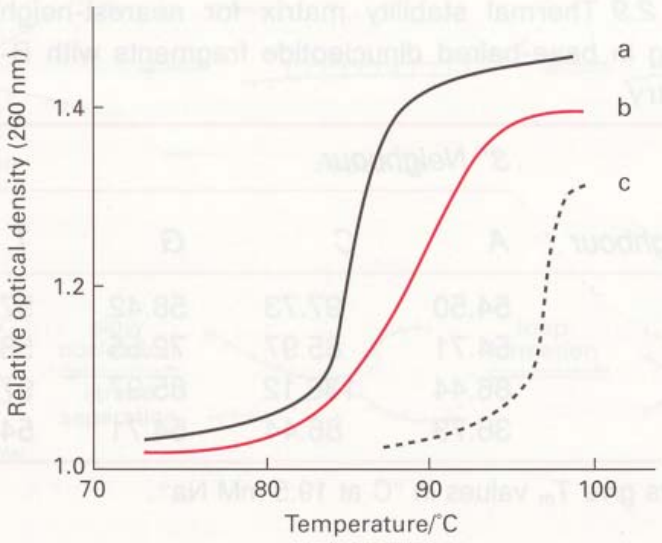

Fig. 2.35 Thermal denaturation of DNAs as a function of base composition (per cent G·C) for three species of bacteria: (**a**) *Pneumococcus* (38 per cent G·C); (**b**) *E. coli* (52 per cent G·C); (**c**) *M. phlei* (66 per cent G·C) (adapted from Marmur, J. and Doty, P. (1959). *Nature*, **183**, 1427–9. Copyright (1959) Macmillan Magazines Ltd).

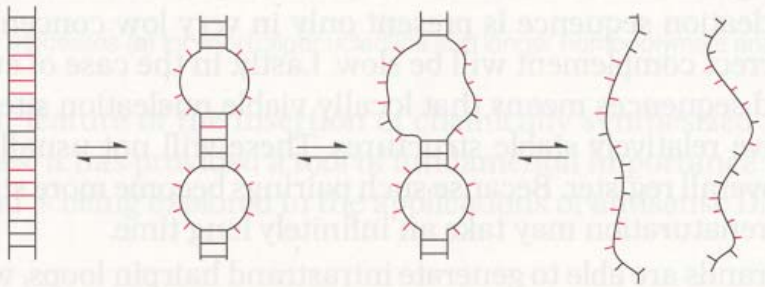

Fig. 2.36 Scheme illustrating the melting of A·T-rich regions (colour) followed by mixed regions, then by C·G-rich regions (black) with rise in temperature (left→right).

Table 2.9 Thermal stability matrix for nearest-neighbour stacking in base-paired dinucleotide fragments with B-DNA geometry

	3′ Neighbour			
5′ Neighbour	*A*	*C*	*G*	*T*
A	54.50	97.73	58.42	57.02
C	54.71	85.97	72.55	58.42
G	86.44	136.12	85.97	97.73
T	36.73	86.44	54.71	54.50

Numbers give T_m values in °C at 19.5 mM Na^+.

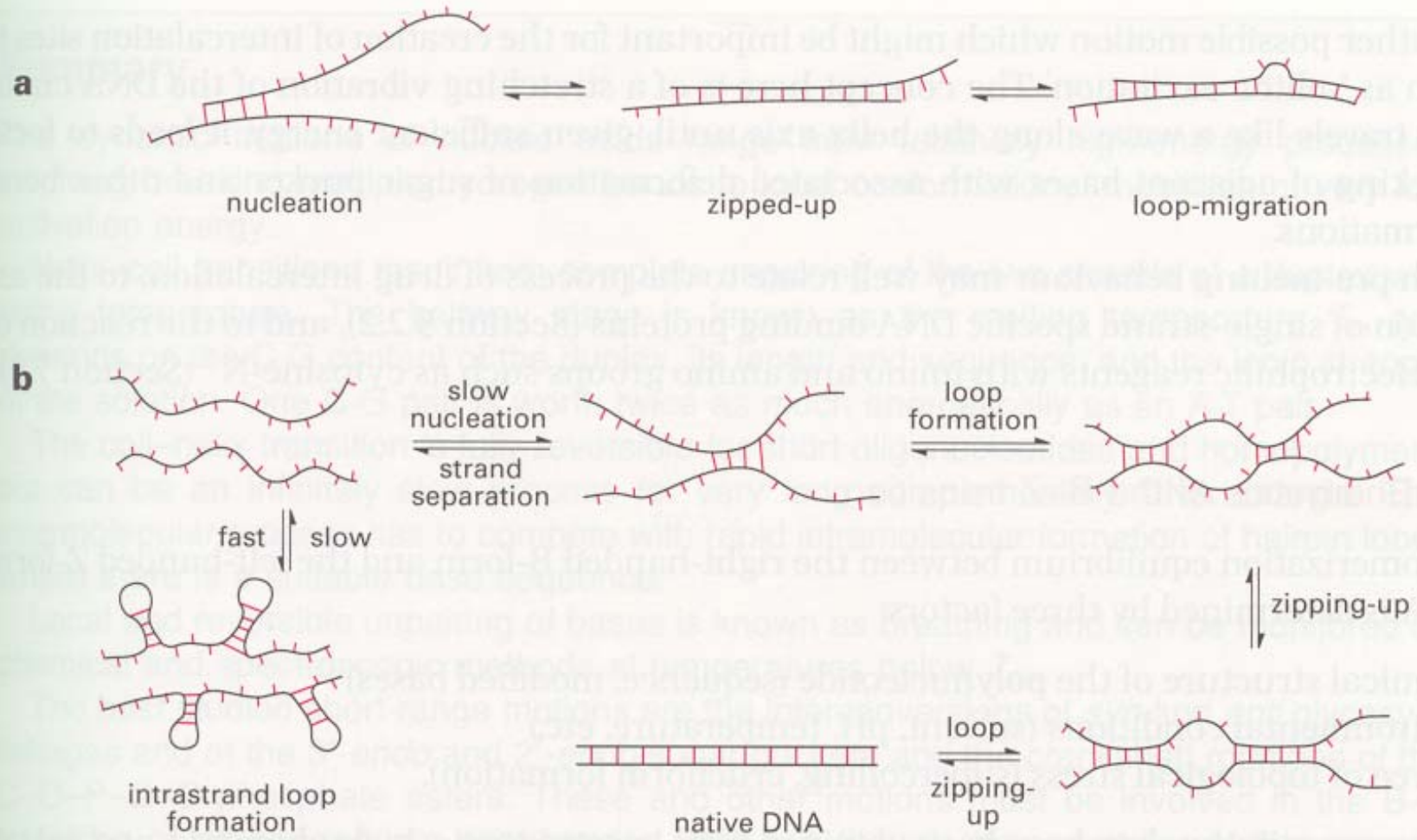

Fig. 2.37 Renaturation processes (**a**) for short oligonucleotide and longer homopolymers and (**b**) for natural DNA strands.

is now an integral feature of the insertion of chemically synthesized DNA into vectors. For RNA·DNA duplexes, it has provided a tool of fundamental importance for gene identification (Section 10.9.9) and is being explored in the applications of antisense DNA.

Table 5.1 Free energy of base pairing is determined by doublet sequences. The doublet sequences are written so that the top row represents a strand running 5'-3' from left to right; the lower strand is the complement running 3'-5' from left to right.

Doublet Sequence	ΔG
A-U doublets	
AA UU	-0.9
AU UA	-0.9
UA AU	-1.1
Mixed doublets	
CA GU	-1.8
CU GA	-1.7
GA CU	-2.3
GU CA	-2.1
G-C doublets	
CG GC	-2.0
GC CG	-3.4
GG CC	-2.9

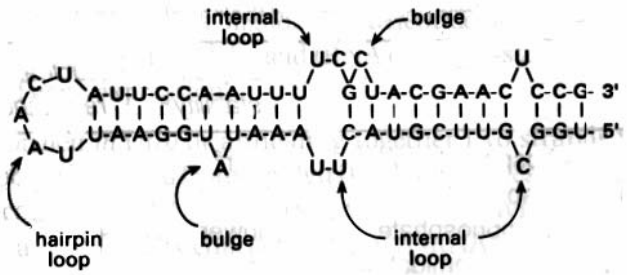

Fig. 2.34 A possible secondary structure for a 55-nucleotide fragment from R17 virus which illustrates hairpin loop, interior loop, and bulge structures. The free energy of this structure has been calculated to have a net $\Delta G°$ of -90 kJ mol^{-1} using appropriate values for base-pairs (Table 2.6) and for loops and bulges.

This is a table of the approximate free energy contributions of internal loops, bulges and hairpins. Notice they are all positive – that is they tend to destabilize the structures if the complementary base pairs are not present. But the sequences are often such that the only way the molecule can fold (or at least one of the more favorable ways) is to form such structures. Much less is known about the characteristics of these structures than for the base pairing possibilities and energetics.

Table 2.8 Free energy increments for loops (kJ mol^{-1} in 1 M NaCl, 37°C)

Loop size	*Internal loop*	*Bulge loop*	*Hairpin loop*
1	—	+14	—
2	+4	+22	—
3	+5.4	+25	+31
4	+7.1	+28	+25
5	+8.8	+31	+18.5
6	+10.5	+34.5	+18

The following experiments were made my Maxim Frank-Kamenetskii's research group in Moskow, and can be seen in his book "Unraveling DNA". All the DNA molecules in a solution have the same sequence. If you melt a solution of one type of DNA molecule (but many molecules) ***very slowly*** (these experiments can take weeks) the following data can be acquired, where the **heat absorbed** at any temperature (this experiment is done in a calorimeter) is plotted versus the **temperature**.

We see that there are regions where there are large amounts of heat absorbed, and regions where little, or no, heat is absorbed. What are these peaks due to? Remember the thermodynamic values given earlier in the lecture. The enthalpy and entropy of the different base pair permutations are very different. So the explanation seems to be that the regions corresponding to different "mean sequences" melt at different temperatures. If the temperature were raised rapidly, in comparison to the kinetics of the "melting reaction", we would see a broad curve, but each of the peaks has a width of about 0.5 C, and this is to be expected from the theory, which we will develop.

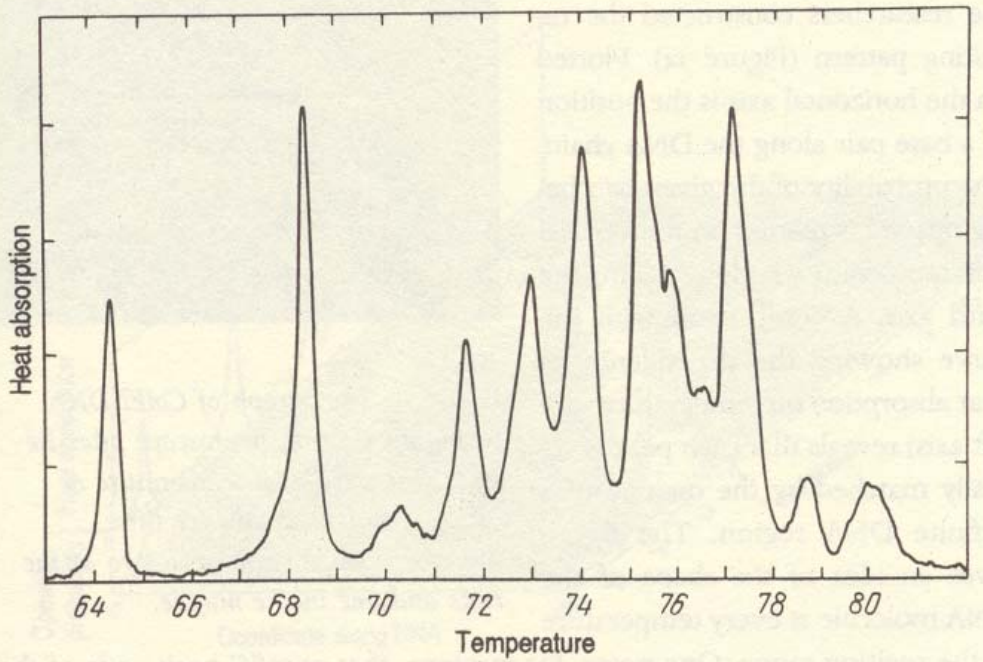

Figure 12. Dependence of DNA's heat absorption on temperature. This curve is often called the differential melting curve. The curve was obtained for DNA that has the code name ColEl and contains about 6,500 nucleotide pairs.

How can we see this phenomenon on the molecular scale – even with just single molecules, so we can be sure what is happening?

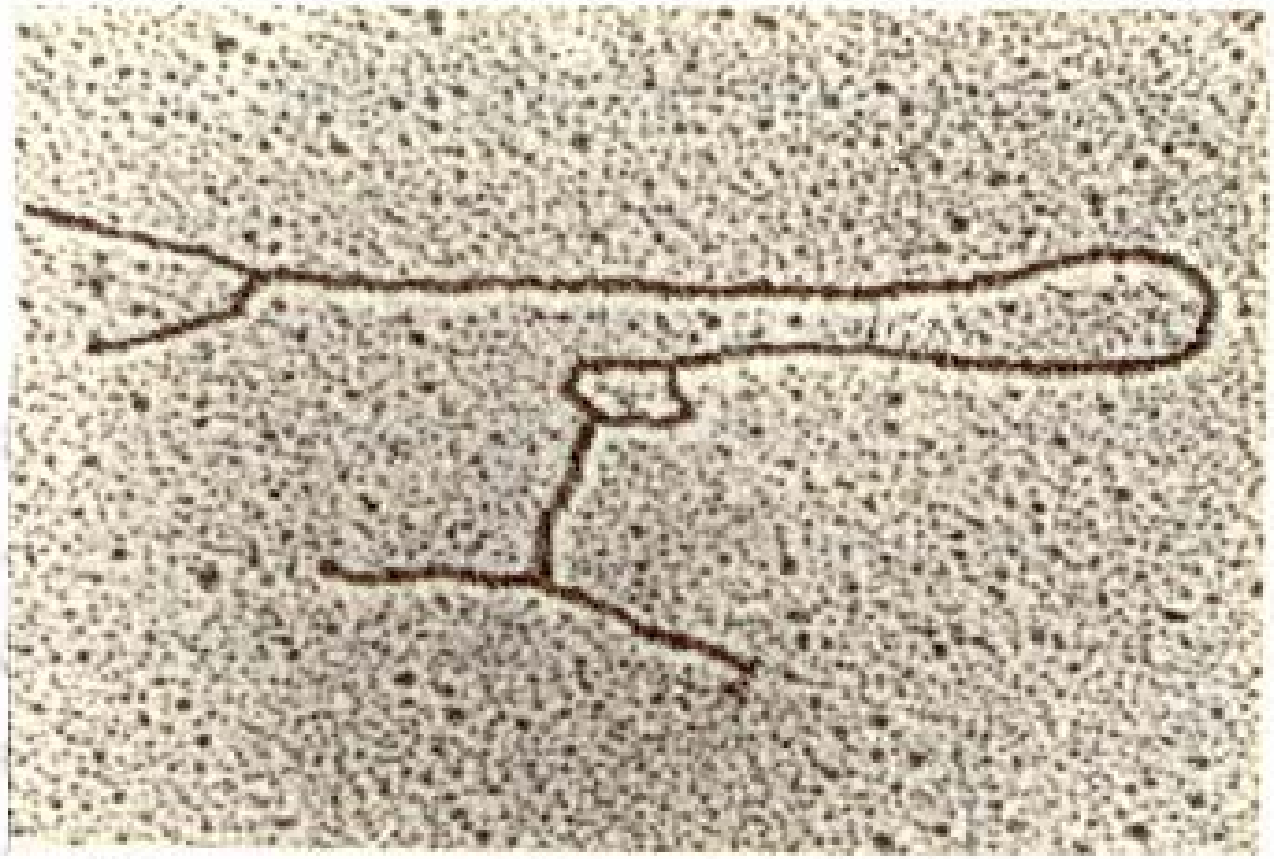

Figure 13. Photograph of ColEl DNA (using an electron microscope after its state was fixed at a temperature of 73°C). One can clearly see three

The “**melting**” of a DNA molecule can be visualized on a molecular scale by electron microscopy. A solution with DNA molecules (all the same) were heated up to a particular temperature. The strands of the molecule in the regions of the DNA molecules which were melted at any particular temperature have their nitrogenous bases in contact with the solvent. A chemical that reacts with the exposed bases was added to the solution. The buried bases (base paired) were not affected. Then the solution was cooled, the solvent was exchanged, and electron micrographs of the DNA molecules were made.

The statistics of many molecules was made, and because the position along the DNA molecules could be determined in the pictures, the following three dimensional representation of the data could be made.

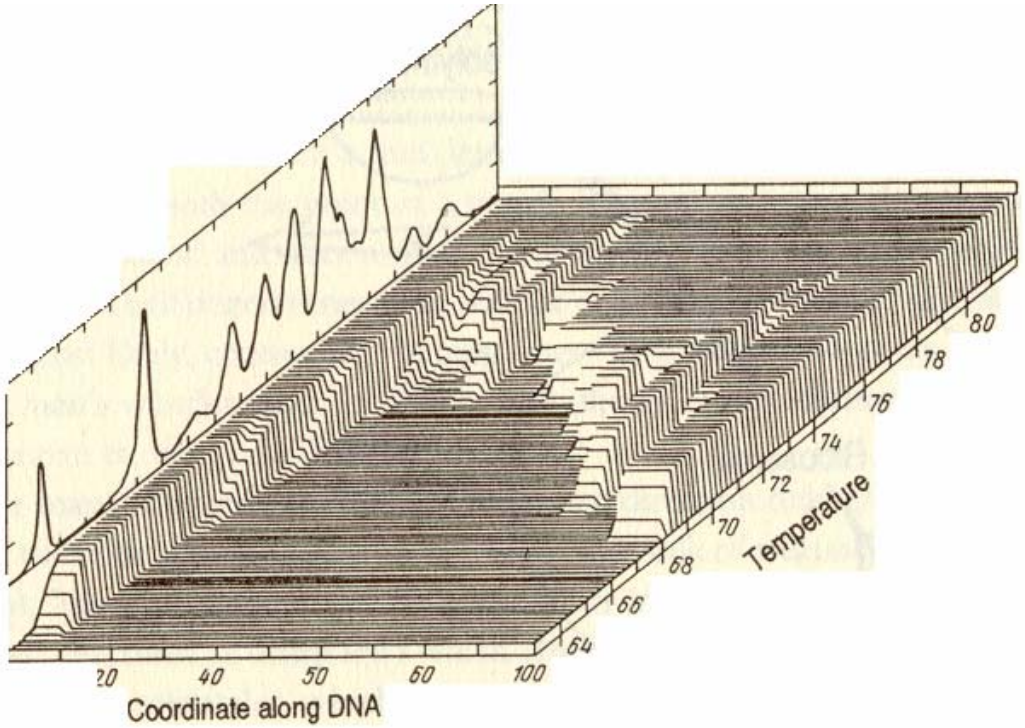

Figure 14. The full picture of ColEl DNA melting, obtained by computer processing of a large number of electron-microscope photographs of the type shown in Figure 13.

This experiment brings up some very *important considerations* regarding the stability of different regions of DNA.

1) The broader width of the melting (disorder) transition is not really due to the fact that the DNA melting reaction is happening in one dimension, where we know that no first order phase transition can happen, but it is happening because of the heterogeneity of the sequence in natural DNA.

2) It is interesting that there are organisms that live at temperatures above many of the temperatures where the sequences melt, so they must have a way to control this.

3) In general, for the correct functioning of a cell, the temperature does not vary, and it is far below the temperatures where these sequences are melting (usually at 37 C).

So what is the role of this large sequence variance in the stability of different regions of the DNA in a chromosome?

During transcription, RNA polymerase binds to very special sequences on the chromosome, and these are AT-rich sequences. To function, the RNA molecule has to separate the strands and “read” the sequence information, and this will take place most easily at sequences that are easily melted. Therefore the promoters region have high AT content, and the proteins act as local “heaters”. How they can do this we will discuss later, but it is a fact that these proteins (many more than just the polymerases) essentially bind to sequences that are easily melted (or sometimes the reverse). So this phenomenon is an important control mechanism in the life of a cell. Even during the elongation phase of transcription the melting propensity of the

DNA (and the newly synthesized RNA molecules) is used as a control mechanism to control when, and how fast certain sequences on the chromosome are synthesized.

4) Note that the melting of such long heterogeneous DNA does not always melt from the ends; internal loops are first made depending on the sequence.

5) The melting of DNA is very important for biotechnology, as we will now show.

We are going to investigate the thermodynamic description of linear chain polymers with monomer units that can exist in different states. Most of the important polymers in biology (biopolymers) are one dimensional polymers, where the monomers are connected to each other in an ordered linear array (for instance all the polymers which we have considered up to now, such as the proteins and nucleic acid polymers).

By “thermodynamic description” we mean we want to know the details of the “conformation” of the biopolymer at a given temperature, pressure, ionic strength, pH and molecular environment. Obviously it is important to predict the polymer conformation because all of the biopolymers in general carry out some biological function, and the ability, or the capacity, with which the molecule can operate, depends on the “state” that the bio-macromolecule finds itself in.

Biological molecules are special adaptable machines with many micro-states. We have seen

this when we discussed the conformational space of proteins. If the molecule were a normal “machine” that has to have all “parts” working in a certain way, or that has to be assembled in a definite particular way, then we would know how the macromolecule works just by knowing if all the small parts are in order.

Think of a car motor. In general all parts of a finely tuned motor have to be assembled in a certain way with very rigorous requirements. An efficient machine that is realized by good “man-made” engineering design has very strict performance characteristics of the different parts. If any of the crucial parts does not work exactly right, we know the machine could not be used, and that it has to be sent back to the shop for repairs. But biological systems are quite different.

Good man-made machines (or electronic circuits) don’t have “slop”.

But molecules have to deal with a very complex ever-changing situation. They have to operate in a very sloppy environment. In a way biological molecular machines have developed in such a way that they are flexible, and in a way sloppy, and this is an inherent part of their design. Well - we don’t want to exaggerate.

Molecules, even large ones, cannot escape from a very hectic environment. They are designed to operate in rather wildly frenzied surroundings, where the molecules are battered with objects of their same size, and are “pushed” over distances comparable to their size. This has major consequences when we consider how these molecules function, or even

exist.

Biological systems do work very precisely in many respects; however, the parts are not always to be found in only one state. There are many different states that the smaller parts of a macromolecule can be in, and still the overall biological system - this can be a single macromolecule, a supramolecular assemblage of molecules, or a whole organism - works well.

The *details* of the conformational distribution within the statistical mechanical “state-space” in which a macromolecule finds itself can be *quite complex*. The allowed conformations, most importantly the most prevalent conformations, depend upon the details of the possible intermolecular and intramolecular interactions, as well as the random conformations that are possible if there are no such interactions (e.g. as we have discussed when covering the theory of polymer physics).

Perhaps we are not interested in the intricacies of how the functional parts of a macromolecular system works, and how the complex interactions between the molecular components lead to some functional biological usage.

In this case a purely phenomenological point of view of the system (as is sometimes the case, even for some biophysical studies, and more often for a higher level description of some biological process) will suffice. In such a case we can just ask how the system acts under whatever circumstances exist normally, and we can just describe the actions and behavior of

the system from a purely descriptive point of view (sort of a “black box” level, as far as the molecular characteristics are concerned, and many times using models that correspond to the realm of large objects and classical mechanics).

But if we want to ask how the system works on a molecular level, the level of difficulties rises tremendously. One of the reasons for this, besides the large-scale complexities that can, and do, arise, is that the biomolecules are made of many small parts, and these parts can individually exist in many states, and the parts can interact with each other on a very small scale.

The number of multiple states for any biopolymer, as we have pointed out earlier, are unbelievably large (numbers like $(20)^{100}$ or so different arrangements, and each “part” can exist in 3-10 different states, not counting the proximal and distal intermolecular interactions along a folded chain).

We need a way to investigate this molecular level of action, without getting lost in all the complexities, which would lead us nowhere. The molecules seem to have found a way (through evolution) to avoid getting lost in their conformational and functional space. They can limit themselves to a relatively very tiny subset of the possible state-space, even if this “selected subspace” still seems large to us from an experimental (or even a computer theoretical) point of view. And they do this in a way so that they often retain an impressive flexibility. But we do have ways to observe this scale of workings, and we will discuss such things later.

However, before we can even begin figuring out how these macromolecules can “find” their “working distribution” in this large conformational space, we have to know the approximate energies and forces that act between the parts of the macromolecules on the molecular and atomic scale. In general this is not an easy undertaking, because we can only “disassemble” the macromolecule to a certain extent without losing the whole “function” that we are trying to understand. We need to have ways to observe the system so that we are learning about its inner molecular workings in as gentle a way as possible. This can be done on several levels.

One very elegant way (which we will discuss later) is to just watch the ***fluctuations*** of the molecular system; then if we are clever, we can observe the important characteristics without disturbing the system at all. This is sometimes possible, but as yet, it only works for certain systems, and for certain questions. In general we have to **perturb** the system in order to get this information. The levels of perturbation, and the methods of observing the results of a perturbation, vary widely. We may discuss this when we discuss “relaxation” methods.

How do we analyze how a macromolecule is built?

1) The most fundamental thing we have to know first is **the identity of the building blocks** (after we have selected the biological system and gotten some significant molecular components – this is also a prerequisite!). This took decades of work previously, but either it has been done to a large extent, or the methods have been developed so that we can fairly

rapidly get this information nowadays. It is the province of the chemist.

2) Then we have to know **how the building blocks are assembled** in a “primary” way (that is the primary sequence) and this methodology has also been highly developed (and is still being developed) so that we can fairly rapidly get this information today. Previously this also took years. This is sequence analysis of proteins and nucleic acids (DNA/RNA) – for instance, the human genome project.

3) Next we want to know how the different parts of the sequence fit together to form a **tertiary structure**, and this is where things get interesting for the discussion now. The techniques of X-ray analysis and NMR especially have been improved and developed so much in the last years, that we have a good idea of the “average” 3-D structure of many macromolecules. There is a lot of room for improvement and there are always new and exciting improvements, but there has been tremendous progress in this area. We have seen many examples in the lectures, and discussed the “rules” for assemblage and arrangement of the macromolecular structures.

4) But now we come to the topic we want to discuss – what are the **energies** responsible for the structures that have been found, and how do these interactions explain the way the molecules “fold” and how they function. And how do we get this information? This information must come from experiments. These questions are essentially asking us to understand the thermodynamics of the molecules in their environment, and to understand the statistical mechanics on the molecular scale. The *statistical mechanics is especially important* because

all the action takes place on the molecular scale, and on this scale things fluctuate wildly unless they are confined by tight overall constraints. The overall thermodynamics (better thermo*statics*) will not tell us anything about the distributions of conformations and fluctuations of the macromolecules. BUT we also need to have the thermodynamic experiments, descriptions and numbers.

Remember that the constraints are expected to be just the level necessary in order to “hold” the macromolecules in particular conformations so they can function “good enough” for their job. The molecules function on the molecular scale individually, so that we are not just interested in a fancy statistical mechanical way to describe some overall macromolecular state that does not fluctuate on the level of interest (like a crystal of a piece of metal). The statistical mechanical description is necessary to describe the molecular system realistically at all.

It may be that the distribution within the state-space of a macromolecule is limited to a narrow extent due to strong interactions. Maybe certain characteristics can be fairly well understood using models that are purely static; if so this is convenient. If this is the case, then we want to understand how this is accomplished; that is, how to constrain a molecularly sized system so strongly. However, the real system is fluctuating, some much more than others and we have to understand the interactions between the parts to understand these fluctuations.

This is the reason for carrying out the type of experiments we will discuss under the topic of “helix-coil” transitions. It could be described well as “order-disorder” transitions. We will

consider simple cases; the complex cases are then not hard to imagine, and it is not hard to imagine how difficult and complex they can be!

The theories of *helix-coil transitions* are clad with many different outward appearances. There are numerous advocates of the different models, each purporting to have advantages for solving the problem of linking the statistical probabilities of a large number of molecular states with experimental evidence that is based on only a few macroscopic parameters. The central problem in understanding the statistical mechanics, and eventually the thermodynamics, of any conformational space of a macromolecule, be it a protein, DNA, RNA or lipid, is to find a *convenient representation for the partition* function.

Partition functions are a weighted sum over an enormous number of macromolecular conformations, where the weighting factor for each "state" representing any particular molecular conformation depends on the energy (enthalpy) and entropy corresponding to that state. If we know the partition function, and all the components that go into constructing it, we can then calculate the relative probability of finding a particular conformation, or sub-ensemble of states, within the conformational space of the molecule. This depends on the external thermodynamic conditions of the molecules, such as the temperature, pressure, ionic strength, pH or other solution or environmental conditions.

The extent of interactions with other macromolecular components is also very important for the distribution in the space of all possible molecular conformations.

*Once we know the **partition function**, we can construct all the thermodynamic quantities possible from the system, and can calculate all the experimental observables*. *This is our link between the experiments and the description on the molecular scale*. We have seen how this works with the polymer chains, where we calculated the macroscopic characteristics from the microscopic description.

Experimentally we do things like perturb the macromolecular system with temperature or pressure changes, change the solvent (different ions or pH), or add "ligands" which are molecules that bind (or otherwise interact) with the macromolecules. We then see how the system reacts, and try to understand the energetics of how the molecule responds to the perturbation.

A new dimension has entered the experimental field in the last years, and this is the realm of single-molecule experiments – we will discuss these experiments later. This gives us a direct way to look at the microscopic scale of things – but we still need the statistical description.

We will use the same tools that we use in physics to describe statistical systems, but on a molecular scale; that is the molecule itself becomes the system. Remember what a "melting" curve looks like (where we change the temperature). Take this as an example.

The partition function

Assume we are dealing with a linear molecular system with N units, p_i, where "i" stands for

the position of the unit in the chain from the left end. Write the primary structure:

p_1-p_2-p_3-p_4-...p_{N-1}-p_N

Say that each state can exist in a certain state ("h" or "c"). So we have N^2 possibilities for the "conformation of this bio-"chain".

For instance, one single conformation can be represented by:

$\overset{c}{p_1}$- $\overset{h}{p_2}$- $\overset{h}{p_3}$- $\overset{c}{p_4}$-... $\overset{h}{p_{N-1}}$- $\overset{c}{p_N}$

where unit 1 is in the state "c" ($\overset{c}{p_1}$) , unit 2 in the state "h" ($\overset{h}{p_2}$), ..., unit N is in the state "c" ($\overset{c}{p_N}$).

You can have any combination of these states (2^N of them – the positions of the units within the chain cannot permute, of course).

The probability that the unit "p_i" will be in a particular state with the energy ε_i will be proportional to the Boltzmann factor,

$$\exp(-\varepsilon_i / kT)$$

We are at equilibrium with constant temperature and pressure. ε_i is an energy, actually the "free energy", of the i^{th} unit in the particular state.

The polymer chain will have a certain probability of being in *each of its particular states*, and this probability is specified by the energy that each and every unit has in each state:

$$probability\ of\ a\ state = \prod_i \exp[-\varepsilon_i / kT]$$

The partition function (Z_N) is just the sum of these products over all the states possible for the polymer chain:

$$Z_N = \sum_{\mu_{1,2,3,\ldots N}} \prod_i \exp[-\varepsilon_i / kT],$$

where the sum is over all the different combinations of how the energies can be distributed over the different units. This is a very general formula, and just as useless as it is general.

We have to specify the number of unit states possible, and their corresponding relative free energies.

Now say that each state has only two possible energies, depending on whether it is in state "h" or "c", – that is, ε_c is the energy of state "c" and ε_h is the energy of state "h".

The Boltzmann factor for a state that has "m" helix units, and "n" coil units (where h = helix and c = coil) is:

$$\left\{\prod_m \exp[-\varepsilon_h/kT]\prod_n \exp[-\varepsilon_c/kT]\right\}$$

and there are

$$N!/n!m! = N!/m!(N-m)!$$

ways to have m helix states, and n coil states; so the total sum of those Boltzmann factors which have m helix, and n coil states are:

$$\sum_{m=h,n=c}\left\{\prod_m \exp[-\varepsilon_h/kT]\prod_n \exp[-\varepsilon_c/kT]\right\}$$

$$=\frac{N!}{m!(N-m)!}\left\{\prod_m \exp[-\varepsilon_h/kT]\prod_n \exp[-\varepsilon_c/kT]\right\}$$

$$=\frac{N!}{m!(N-m)!}\left\{\left(\exp[-\varepsilon_h/kT]\right)^m\left(\exp[-\varepsilon_c/kT]\right)^{(N-m)}\right\}$$

and then the <u>total sum for the **partition function**</u>, over **all the possible states** is just:

$$Z_N=\sum_{m=0}^{m=N}\frac{N!}{m!(N-m)!}\left\{\left(\exp[-\varepsilon_h/kT]\right)^m\left(\exp[-\varepsilon_c/kT]\right)^{(N-m)}\right\}$$

and this sum can be written as:

$$Z_N=\left(\left(\exp[-\varepsilon_h/kT]\right)+\left(\exp[-\varepsilon_c/kT]\right)\right)^N=\left(\exp[-\varepsilon_c/kT]\right)^N\left\{\left(\exp[-(\varepsilon_h-\varepsilon_c)/kT]+1\right)^N\right\}$$

Without loosing generality, we can define the free energy of the “c” state ε_c to be our base value, and set ε_c to zero.

This we can do because we will always use Z_N to find the probability of a certain collection of states, and this means we will always be dividing by Z_N anyway. That is, the probability of a certain selected sub-ensemble of states is

$$\text{probability of subset of states} = \left\{ \sum_{selected\ set} \prod_{\{\mu_i\}} \exp[-\varepsilon_i / kT] \right\} \Big/ Z_N$$

Then we have our **partition functio**n in a ***convenient*** **representation**:

$$Z_N = \left(\exp\left[-\left(\varepsilon_h - \varepsilon_c\right)/kT\right] + 1\right)^N = \left(s+1\right)^N$$

Note with the last notation, the partition function is expressed as:

$$Z_N = \sum_{m=0}^{m=N} \frac{N!}{m!\left(N-m\right)!}\left(s\right)^m = \left(s+1\right)^N \quad ;$$

we define "s" below.

Remember, the ε_i in the last equations, should be the "**free energies**" (G);

that is,

$$(\varepsilon_h - \varepsilon_c) = \Delta G_{h-c} = \Delta H_{h-c} - T\Delta S_{h-c}$$

And thus “s”, defined above, can be written as:

$$s = \exp[-(\varepsilon_h - \varepsilon_c)/kT] \Rightarrow \exp[-(G_h - G_c)/kT]$$
$$= \exp[-(\Delta G_{h-c})/kT] = \exp[-(\Delta H_{h-c} - T\Delta S_{h-c})/kT]$$

This means that “**s**” is just the **equilibrium constant** of one unit being in the “h” or “c” state

Remember $\Delta G = -kT\ell n(K_{eq})$ that we are writing $\Delta G \equiv \Delta G^0$

That is:

$$K_{eq,h-c} \equiv s = \frac{[h]}{[c]} = \exp[-(\Delta G_{h-c})/kT]$$

The ***average fraction*** **of "h" units** (where <m> is the average **number** of "h" units out of a total of "N" units) can be written easily by using our partition function, Z_N as:

$$\theta = \frac{1}{N}\langle m \rangle = \frac{1}{N} \cdot \left\{ \sum_{m=0}^{m=N} \frac{N!}{m!(N-m)!} [m](s)^m \Big/ Z_N \right\} = \frac{1}{N} \cdot s \left[\partial Z_N \Big/ \partial s \right] \Big/ Z_N = \frac{1}{N} \cdot \frac{\partial \ell n Z_N}{\partial \ell n s}$$

The last two expressions are simply mathematical conveniences to calculate *m* when we have a convenient expression for Z_N.

This gives:

$$\frac{\langle m \rangle}{N} = \frac{1}{N} \cdot \frac{\partial \left[\ell n \left\{ (s+1)^N \right\} \right]}{\partial \ell n s} \frac{1}{N} \frac{N(s+1)^{N-1} \cdot s}{(s+1)^N} = \frac{s}{(s+1)}$$

But what does this mean? Say that we have an **ensemble** (concentration in a solution) of a **single *individual* molecularly independent molecules**, being in the "h" of "c" state (with the same equilibrium constant, s; that is, c↔ h. Then the fraction of units in the "h" state θ would be:

$$\theta = \frac{[h]}{[h]+[c]} = \frac{[h]/[c]}{[h]/[c]+1} = \frac{s}{s+1} \quad ;$$

this is the same expression as above for the polymer. So, the fact that the units are contained in a polymer *does not make any difference for the probability of being in h or c state*. Why? This is because there is no interaction between the individual units on the polymer chain. There is no cooperativity – no interactions between the neighboring units on a chain.

Let's look at this situation just a little more:

1) Define T_m as the temperature where "s"=1, or where ½ of the molecules are in the "h" and "c" states; the "m" subscript stands for "melting" temperature. That is, this is the temperature where

$$\Delta G_{h-c} = 0$$

Remember,

$$s = \exp\left[-\left(\Delta G_{h-c}\right)/kT\right] = \exp\left[-\left(\Delta H_{h-c} - T\Delta S_{h-c}\right)/kT\right]$$

If at T=Tm, we have

$$\Delta G_{h-c} = 0$$

$$\Delta H_{h-c} = T_m \Delta S_{h-c}.$$

So,

$$\Delta S_{h-c} = \Delta H_{h-c} / T_m$$

And we can rewrite the above equation as:

$$s = \exp\left[\frac{\Delta H_{h-c}(T - T_m)}{kTT_m}\right] = \exp\left[a \cdot \frac{T - T_m}{T}\right],$$

where

$$a = \frac{\Delta H_{h-c}}{kT_m}$$

Remember ; we lose heat when we form a helix, so ***a* is negative**!.

$$\Delta H_{h-c} \leq 0$$

Use this expression to write the ***fraction helix state***:

$$\theta_h = \frac{s}{(1+s)} = \left(1+\frac{1}{s}\right)^{-1} = \left(1+\exp\left[a\cdot\frac{T_m - T}{T}\right]\right)^{-1}$$

If T<Tm, then $\left[a\cdot\frac{T_m - T}{T}\right]$ the exp becomes small, and $\theta_h \to 1$

if T>Tm, then $\left[a\cdot\frac{T_m - T}{T}\right]$ the exp becomes large, and $\theta_h \to 0$

The sharpness of the "transition from h->c depends on the magnitude of ΔH_{h-c}

But, in general

$$5 \le \left|\Delta H_{h-c}\right| \le 10 \text{ kcal/mol.}$$

so the sharpness is limited.

However there is one good thing,

$$\Delta S_{h-c} \cong -22 cal/mol/\deg$$

(that is, the helix has less entropy than the coil) , so

$$T_m \equiv \frac{\Delta H_{h-c}}{\Delta S_{h-c}} \approx \frac{-7 kcal/mol}{-0.022 kcal/mol/\deg} \cong 318\ K = 45°C$$;

this is about right! But the curve should be much sharper. We have to introduce *cooperativity* to simulate the curve shape correctly, and to uncover the mechanistic and molecular information hidden in the statistics of the "melting curve".

2) Consider a reaction where we have, instead of two states, h and c, the protonation reaction of a COO^- group at every unit position of the polymer. How would you write the acid-base equations for this polymer with N acid dissociating units? We can assume that each dissociation is independent, and it is the same. Later we can also bring in cooperativity in this "dissociation" or "binding" problem. The binding reaction could be the binding of any ligand to the individual units of the polymer chain

3) If each unit can be in "j" different states with statistical weights q_i, we just used two states, h and c, in the above example, and the *units are independent*, we can write the partition function of a polymer with N units as:

$$Z_N = \left(\sum_i (q_i) \right)^N ;$$

this is a simple equation. You can always normalize by setting one of the statistical weights to 1,as we did in the above example (because you are always dividing by Z_N anyway).

4) This system with independent units on the polymer acts just like a normal reaction in solution. The temperature curve (melting curve) is just

$$\theta \; vs \; T \; , or \quad \frac{s}{s+1} \; vs \; T \; .$$

When $T=T_m$, then s=1 and θ=1/2, as we would expect.

Because

$$s = \exp\left[\Delta H_{h-c} \left(T - T_m\right) / kTT_m\right],$$

s is always >0.

5) The free energy of the whole polymer can be written as

$$G = -kT\ell nZ_N = -kTN\ell n(s+1);$$

$$H = \frac{\partial \frac{G}{T}}{\partial \frac{1}{T}} = -N \cdot k \cdot \frac{\partial \ell n(s+1)}{\partial \frac{1}{T}} = ;$$

$$H = \frac{\partial \frac{G}{T}}{\partial \frac{1}{T}} = -N \cdot k \cdot \frac{\partial \ell n(s+1)}{\partial \frac{1}{T}} = \frac{-N \cdot k}{(s+1)} \frac{\partial s}{\partial \frac{1}{T}} = \frac{-N \cdot k \cdot s}{(s+1)} \cdot aT_m = \frac{N \cdot s}{(s+1)} \cdot \Delta H$$

What does this equation mean? How does this relate to an experiment to determine the heat of reaction of the c<->h reaction?