Helix Coil Transition in DNA

Dyutiman Das

Loomis Laboratory of Physics,
University of Illinois at Urbana-Champaign

(Dated: May 2, 2000)

The theory of the helix-coil transition is introduced with the help of a toy model, and a simple estimate of the melting temperature is made. Modifications necessary for the DNA are indicated, along with the role of solvents, pressure etc. on the transition. Lastly the nature of the phase transition is discussed in light of the finite size scaling analysis.

I. INTRODUCTION

Proteins and nucleic acids are long chain polymers, which in an aqueous solution are expected to behave like one dimensional random coils. This is an extremely high entropy state because of the large number of possible conformations. However, in the solid phase, the free energy is minimised by taking up a helical conformation, the famous double stranded helical structure of the DNA being an example. Experimental evidence indicates that the isolated biopolymer in solution is in a state of equilibrium between helix and random coil.
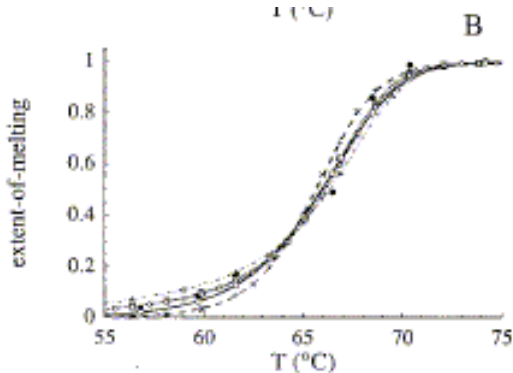

FIG. 1: Temperature variation of optical density of DNA obtained from experimental data [7]

This raises interesting questions about the nature of the equilibrium, i.e., the coexistence of both helix and coil, and the effects of temperature, pH, salt concentration, and chain length on the equilibrium. Measurements of properties such as optical density [7] show a very sharp transition with temperature, as in Fig. 1 indicating that the helix-coil transition might be a phase transition.

Different models in the limits of different chain lengths have been proposed to incorporate this transition. However since the temperature in biological systems remain constant, it is assumed that the transition is brought about by changes in the solution properties. The effects of the solution on the helix stability and coiling has been a major area of study in recent years. Though most of the theory was developed in the middle of the last century, the field has received a new boost with the development of computational techniques.

Another area of interest has been the very nature of the transition. According to the Landau theory, an infinite one dimensional system with finite range interaction do not exhibit phase transition. The method of "finite size scaling" avoids this by postulating that the system is not one dimensional and that the interaction is long range. These recent studies show that the transition is first order though the results are not conclusive.

In this paper I will first discuss the simplest model of the α-helix, to develop the theory of the amino-acids; though this is inaccurate it'll give a simple picture of how to apply the general techniques of statistical mechanics to biopolymers. Later this will be improvised to incorporate the DNA molecule.

II. A TOY MODEL

Consider a long chain molecule consisting of N linked units, each unit being able to exist only in two states, the helical state (h) and the random coil state (c). The units, called residues consist of several atoms and if the dihedral angles (angles of internal rotation about its single bonds) are those characteristic of the helix, its in a state (h); otherwise its in the state (c).

Thus there are 2^N states of the molecule generated by the expression $(h + c)^N$, and the free energy $G\{h, c\}$ is a function of a given $\{h, c\}$ sequence. Then the partition function for the system becomes

$$Z(N) = \sum_{\{h,c\}} \exp[-G\{h, c\}/RT] \qquad (1)$$

with the probability of the occurrence of a given sequence being given by the Boltzmann distribution. All the thermodynamic observables can be calculated form this partition function, and hence the problem reduces to the calculation of this quantity. Because the number of conformations is formidably large, several approximations

are required to answer any question of biological interest.

A   Independence of Sequences

The most important assumption in calculating the partition function is that the free energy of successive sequences of c's and h's are independent. This imposes restrictions over the interaction between sequences along the chain, however this is the minimum model to make progress in calculating the partition function. This implies that the free energy of a sequence depends only on the length of the sequence and not on the position along the chain where it occurs, nor on the length of the neighbouring sequences. Hence, the statistical weight of a sequence of coil states i residues long can be written as

$$u_i = \exp[-G_i(c)/RT]$$

and similarly for a j residue long helical state

$$v_j = \exp[-G_j(h)/RT]$$

It is also seen that the helix and coil sequences alternate, hence allowing the devision of any conformation into pairs of sequences containing one helical and one coil sequence. Hence

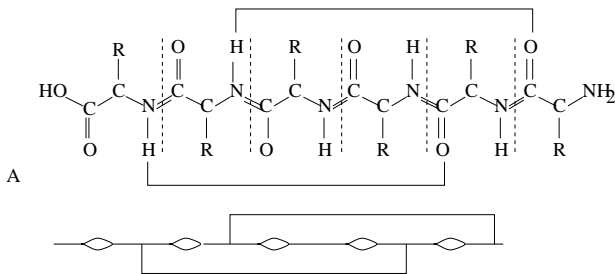$$Z(N) = \sum_{i,j} \exp[-G\{h,c\}/RT] = \sum_{i,j} \prod u_i v_j \qquad (2)$$

B   The α-Helix Model

FIG. 2: A representation of the α-helix model

The structure of the α-helix as seen in Fig 2. has the NH group of the i-th residue hydrogen bonded to the CO group of the $(i+4)$-th residue, thus spanning the covalent bonds (about which internal rotations can take place) of three residues. Thus two residues of the helical sequence are free, i.e. in a j-residue long helical sequence only $(j-2)$ hydrogen bonds are formed.

Thus, there are three types of states in the polypeptide chain: (1) a coil state; (2) a helical state with a hydrogen bond; and (3) a helical state without a hydrogen bond. The statistical weights of these states are tabulated below

Table 1

| | Statistical wight | |
State	I	II
Coil	u	u/u=1
Helix with		
hydrogen bond w		w/u=s
Helix without		
hydrogen bond v		v/u=$\sigma^{1/2}$

Here the statistical weights have been normalised with respect to the all-coil state. The physical picture of this assignment is as follows: the hydrogen bond determines the stability of the helix and hence $w \sim \exp(-E_H/RT)$, with E_H being the energy of formation of the bond; the internal rotations make up the free energy of the random coil and hence $u \sim \exp(-S/R)$. When the ratio $w/u = s$ becomes greater, less than, or equal to unity, helix is, respectively, favoured, unfavoured, and of equal probability with respect to the coil.

Since the helical sequence has only two units which contribute neither to the coil entropy, nor a hydrogen bond, $\sigma = (v/u)^2$ should be less than unity, reflecting the low probability of occurrence of the ends of the helical sequence. Thus, in this model the helix-coil transition is an order-disorder transition, as s the ratio of helix to coil is decreased (e.g. as the temperature is increased) the ordered low energy helix transforms to the disordered, high entropy coil.

From the definition, the parameter s can be expressed as

$$s = \exp[-\Delta G_{h-c}/RT] \qquad (3)$$
$$= \exp[-(\Delta H_{h-c} - T\Delta S_{h-c})/RT] \qquad (4)$$

where ΔG_{h-c}, is the difference of the free energies of the two states.

Thus we can formulate the statistical weights of the sequences, that of an i unit long coil sequence being $u_i \sim 1$ on normalisation, and for the helical sequence of $j \geq 2$ unit the weight being $v_j = v^2 w^{j-2} = \sigma s^{j-2}$. The weight of the unit long helical sequence is of course just $v = \sigma^{1/2}$.

C   Long Sequence Tendency

For a j residue long helical sequence, the negative free energy per residue is given by

$$(1/j)(G_j/RT) = (1/j)\log v_j = \log s - (1/j)\log(s^2/\sigma)$$

For $s \sim 1$, the negative free energy is maximised for large values of j, thus indicating that the most probable helical sequence will be the longest possible. This gives rise to the "all-or-none" transformation between complete helix and random coil. In this approximation the only two probable states gives rise to the simplest partition function

$$Z(N) = v^N + u^N \qquad (5)$$

Rearranged, this becomes

$$Z(N) = u^N[\sigma s^{N-2} + 1]$$

and the hydrogen bond fraction being

$$\theta(N) = \frac{\sigma s^{N-2}}{\sigma s^{N-2} + 1}$$

In the limit of long chain, this becomes the step function.

$$\theta(N \to \infty) = \Theta(s - 1)$$

The above equations do not represent the total free energy of a given length chain. In an N residue long chain with a sequence of j residues, there are a number of ways of placing the sequence on the chain, i.e. the combinatorial entropy. Thus there is a competition for the helical sequences.

Short Chain	All-or-none
Medium Length	One h sequence of any size
Long Chain	Many of medium length

Though this drastic approximation has to be improved for the model to be more realistic, the simplicity of the model helps to show some general structures. This model is then generalised under different modifications and refinements for different chain lengths, and gives a better picture of the poly-amino acid.

Allowing all possible conformations the partition function becomes

$$Z(N) = (u + v)^N = u^N(1 + s)^N$$

where $u = 1$. Under this less drastic approximation, the average helical fraction is

$$\theta(s) = \frac{1}{N} \frac{\partial \ln Z_N}{\partial \ln s} = \frac{s}{1 + s}$$

This also shows the behaviour of a step function as we will see below.

D   The Transition Temperature

In spite of the simplicity of the model, an estimate of the transition temperature can be made. We see that a transition occurs at $s = 1$, or when half of the molecules are in the h and c states; we denote this temperature as the melting temperature T_m. This is the temperature where $\Delta G_{h-c} = 0$. This gives

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

Hence, s can be written as

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

where $a = \Delta H_{h-c}/RT_m$. Since $\Delta H_{h-c} \leq 0$, (since we lose heat in helix formation), a is **negative**. Rearranging

$$s = \frac{1}{1 + e^{-at}}$$
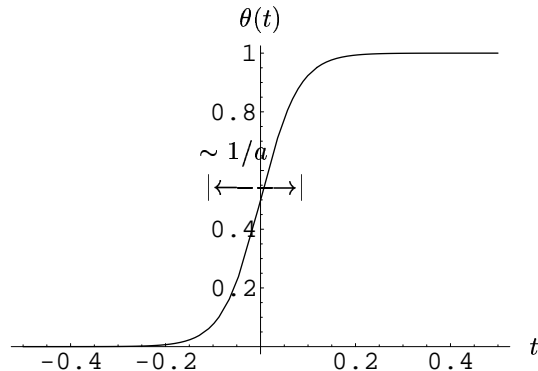
where $t = (T - T_m)/T$.

FIG. 3: The helical density under a relaxed approximation

The sharpness of this transition depends on the magnitude of a i.e. of ΔH_{h-c}. But in general $5 \leq |\Delta H_{h-c}| \leq 10 kcal/mol$, so the sharpness of the transition is limited. However from experimental values $\Delta S_{h-c} \simeq -22cal/mol/^\circ K$, its negative since the helix has less entropy, so

$$T_m = \frac{\Delta H_{h-c}}{\Delta S_{h-c}} \approx \frac{-7kcal/mol}{-0.022kcal/mol/^\circ K} \simeq 318^\circ K$$

i.e. 45° which agrees fairly well with the data from the plots of optical density, though the curve should have been sharper. In order to simulate the sharpness of the curve correctly, interaction between the different parts is essential.

Further refinement of this model is carried out by incorporating the transfer matrix method in evaluating the partition function for the helix modelled after the Ising model. For details see [1].

III.   MODIFICATIONS FOR THE DNA

In developing the theory for the DNA two major modifications are necessary; first the double stranded helix can unwind in two different ways by unwinding from the ends or by forming internal loops; and secondly the composition of the chain is heterogenous even within a given sequence. In the first part of this section I'll discuss the results obtained from incorporating the heterogeneity. In the latter part the unwinding transition will be treated in greater detail.

A   The Heteropolymer Model

The simplest refinement considers two types of units, A and B, in a specific sequence, where each unit can be in one of the two states, h or c. This involves assigning different statistical weights (four in all) for each of these units, even within a given sequence of coil or helix, and incorporating nearest neighbour interaction. The weight for all combinations are obtained by formulating four matrices, and writing the partition function in terms of these matrices. In another method the weights depend on the neighbouring states h or c, but not on the neighbour type (A or B).

These models agree fairly well with the Monte Carlo simulations, and show that the breadth of the transition is broader for the heteropolymer than the homopolymer model discussed earlier. This is understood as follows: the types A and B are considered as weak and strong helix formers, respectively, i.e. A units melt at a much lower temperature than B, hence spreading the transition over a broad temperature range. Calculations show that a random sequence consists of rather short sequences of A and B, and the physical picture of the heteropolymer will be much different from the homopolymer case, for example the end-to-end distance distribution function in the transition region would be drastically affected.

B   Role of the Solution

The essential features of the DNA molecule that are different from polyamino acids, which must beincluded in the model are:
(1) The backbone is a polyelectrolyte containing charged phosphate groups.
(2) The formation of the double-stranded helix requires the association of two chains.
(3) The double stranded helix can unwind from the ends, forming free chains, or in the interior, forming loops.
(4) Natural DNA is a specific-sequence copolymer.

The uncoiling transition has recently been studied by Galindo and Sokoloff [10] by using a model incorporating these general ideas.

The dissociation of the double stranded DNA into single strands involve two steps: the uncoiling of the helix and separation of the strands in which a complete breaking of the hydrogen bond takes place. The solution is described by the solvent, concentration, and type of counterions. In early studies, the effects of the ionic concentration on the DNA melting temperature, the counterion distribution, and binding were addressed by solving the Boltzmann equation.

The computational model of the α-helix is shown in Fig 4. In this model a_0' is the radius of the imaginary cylinder over which the two helical strands are wrapped. The distance between two consecutive charges on the same chain is denoted by d_{pp}, the positions of a pair
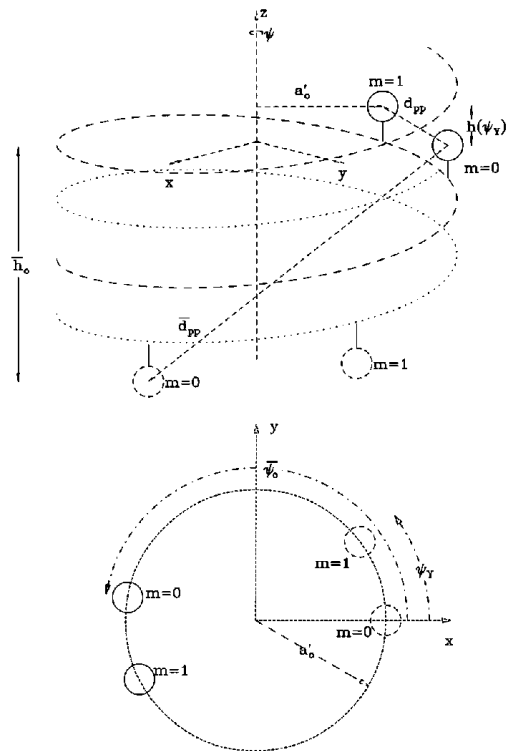

FIG. 4: The DNA modelled for computational purposes [10]

of charges associated with a neucleotide pair can be used to generate the positions of the consecutive pair by rotation about the axis, ψ_Y, and a translation along the axis, $h(\psi_Y)$. $\bar{d}_{pp}$ is the distance between a pair of charges that correspond to the phosphate groups in a pair of complementary nucleotides.

The coiling-uncoiling transition is described in terms of two conformations: the first X, in which the helix is completely uncoiled, $\psi_Y = 0$, and $h(0) = d_{pp}$ (i.e. two parallel charge chains); the second conformation Y, in which the helix is coiled with an angle ψ_Y, and $h(\psi_Y < d_{pp})$. The change in free energy per helix between the two conformations is

$$\Delta W(Y, X) \equiv \sum_{m=-\infty}^{+\infty} \delta W_m \qquad (6)$$

where

$$\delta W_m = W_m(Y) - W_m(X)$$

and the W_m's are sum of terms like

$$w = \frac{q^2}{\tilde{\epsilon}(1 + k\sigma)} \frac{\exp[-k(r - \sigma)]}{r}$$

Here $\tilde{\epsilon}$ is the dielectric constant of the solvent, q the charge of the species and

$$k = (4\pi \sum nq^2/\tilde{\epsilon}k_B T)^{1/2}$$

From these calculations its found that decreasing the dielectric constant of the solvent near the DNA, causes the change in free energy to decrease and the minimum approach zero for a critical value.

Small radii of counterions seem to destabilise the DNA, since the change in the free energy becomes smaller with reduction in radii. It is also seen from these calculations that there is an uncoiling transition at sufficiently high concentrations.

As mentioned earlier, since the temperature of the biological system remains stable, the transition is not induced by changes in temperature. Study of hydrostatic pressure on the thermal stability shows that surface pressure creates a compressive stress in the inter base hydrogen bonds, and this might also cause the transition [13].

IV. IS IT A PHASE TRANSITION ?

Fundamental work on critical phenomenon shows that phase transition can't occur in one dimensional systems with finite range interaction, however the sharp nature of the helix-coil transition makes it a very strong candidate. Recently, it has been conjectured that this transition is due to long range interactions, and the fact that the DNA is not one-dimensional.

The theory of **finite size scaling** has also started to play a very important role in this field. As is known, the fluctuations away from the transitions being small, the system is unaware of its finite size, and the behaviour appears to be critical. However approaching the transition temperature, the fluctuations tend to diverge but are limited by the finite system size, and the behaviour deviates from that of a true phase transition. The observables remain finite, but reach a maxima at some temperature below that for the infinite system.

Now any discontinuity in the thermodynamic observables of interest should reflect itself in the non-analyticity of the partition function. For the DNA, the quantity of interest is the fraction of helical content

$$\theta(s) = \frac{1}{N} \frac{\partial ln Z_N}{\partial ln \ s}$$

Hence a discontinuity $\Delta\theta$ in θ, would imply a first order transition. A study of the partition function zeroes will hence yield information about the nature of the phase transition.

In the work of Hansmann et. al. [15] the partition function is written as

$$Z(\beta) = \sum_E n(E)u^E$$

where $n(E)$ is the density of states and $u = e^{-\beta}$. The energy E includes the Coulomb, Lennard-Jones terms as well as that arising from the hydrogen bonds and torsional deformations. The FSS relation for the zeroes

$u_j^0(N)$

$$u_j^0(N) = u_c + \left(\frac{j}{N}\right)^{1/d\nu} \qquad (7)$$

shows that the distance of the zeroes from the value for the infinite system scales with the relevant system size $L \sim N^{1/d}$.
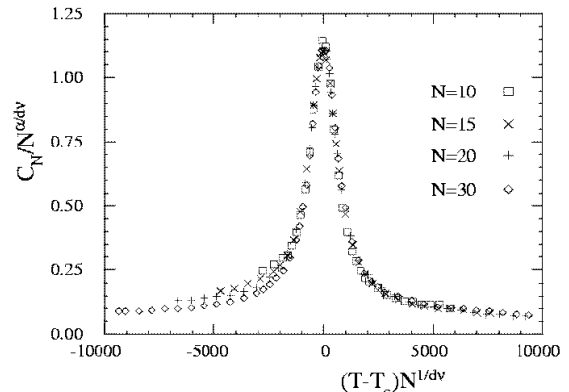

FIG. 5: Calculation of C_V for different chain lengths [15]

However, the most important point of this analysis is that there is no theoretical indication to assume d as a particular integer. The specific heat curve obtained from the energy fluctuations show the finite size effects, it reaches a maximum at some value of temperature which gives an estimate for the value of u_c. Choosing $T_1(N)$ and $T_2(N)$ such that

$$C(T_1) = 1/2 C(T_c) = C(T_2)$$

the following scaling for the width of the specific heat

$$\Gamma_C(N) = T_2(N) - T_1(N) \propto N^{-1/d\nu} \qquad (8)$$

and the specific heat exponent α is obtained from

$$C_N^{max} \propto N^{\alpha/d\nu}$$

Using these equations, they obtain $d\nu = 0.98$, and $\alpha = 0.86, \gamma = 1.06$. The scaling plot for the specific heat is shown in figure 5: curves of all length chains nicely collapse on each other, indicating the reliability of these exponents. Within the error bars these agree with the scaling relation

$$d\nu = 2 - \alpha$$

The renormalisation group values are $d\nu = 1$ and $\alpha = \gamma = 1$ for the first order transition, however the error bars are queit large.

Further studies [14] have been done on minimal models without any system specifics show that there is universality in the scaling behaviour within the finite-size scaling analysis.

V. CONCLUSIONS

The field of biopolyers, and especially the DNA is so vast that it is not possible to cover all the aspects of the helix-coil transition in this short report. Some major phenomena like fluctuations, and loop formation has been left out. Extensive experimental studies have been carried out in this feild, but that too could not be covered. This is a relatively old field which has received a great boost in recent years. Many new avenues are opening up especially in the role of different parameters such as solution properties or pressure in bringing about the transformation. This is becoming more important since its is assumed that the helix-coil transition plays a major role in the early stages of protein formation, and hence a more detailed understanding of the process is necessary. The study of the scaling behaviours is very recent, and there is much work to be done in this sub-field, since most of the calculations done so far are not very conclusive. It has not been yet convincingly proved that the transition is indeed first order, the Hansmann [14] calculations have large error bars, which leaves room for the transition to be second order.

REFERENCES

[1] D. Poland and H. A. Scheraga. "Theory of Helix-Coil Transition in Biopolymers", Academic Press *New York and London.*

[2] S. Lifson and B. H. Zimm. Simplified Theory of the Helix-Coil Transition in DNA Based on a Grand Partition Function. *Biopolymers* **1**, 15 (1963).

[3] B. H. Zimm and J. K. Bragg. Theory of Phase Transition between Helix and Random Coil in Polypeptide Chains. *J. Chem. Phys.* **31**, 526 (1959).

[4] S. Lifson. theory of the Helix-Coil Transition in DNA Considered as a Copolymer. *Biopolymers* **1**, 25 (1963).

[5] D. Poland and H. A. Scheraga. Phase Transitions in One Dimension and the Helix-Coil Transition in Polyamino Acids. *J. Chem. Phys.* **45**, 1456 (1966).

[6] M. E. Baur and L. H. Nosanow. Simple Model for the Helix-Coil Transition in Polypeptides. *J. Chem. Phys.* **38**, 578 (1963).

[7] G. Vámosi and R. Clegg. The Helix-Coil Transition of DNA Duplexes and Hairpins Observed by Multiple Fluoroscence Parameters. *Biochemistry* **37**, 14300 (1998).

[8] R. Clegg. Lecture Notes for PHY450 at the University of Illinois.

[9] B. Åkerman. Barriers against DNA-loop formation in porous matrix. Phys.Rev.E. **54**, 6685-6696 (1996).

[10] C. E. Galindo and J. B. Sokoloff. Uncoiling Transition for DNA in solution. Phys.Rev.E. **54**, 691-705 (1996)

[11] J. F. Marko and E. D. Siggia. Statistical Mechanics of Supercoiled DNA. Phys.Rev.E. **52**, 2912-2938 (1995).

[12] Y.Z.Chen and E.W.Prohofsky. Near Neighbour effects in coöperative modified self-consistent phonon approximation melting in DNA. Phys.Rev.E **49**, 873-881 (1994).

[13] Y.Z.Chen and E.W.Prohofsky. Theory of Pressure-dependent Melting of the DNA Double Helix. Phys.Rev.E. **49**, 2100-2108 (1993).

[14] J.P.Kemp, U.H.E.Hansmann and Z.Y.Chen. Is there a universality of the helix-coil transition in protein models? xxx.lanl.gov physics/0003017 (2000)

[15] N.A.Alves and U.H.E.Hansmann. Partition Function Zeros andFinite Size-Scaling of Helix-Coil Transitions in a Polypeptide. xxx.lanl.gov cond-matt/0002098 (2000).