Microscopical approach to the helix–coil transition in DNA

V.F. Morozov\textsuperscript{a,}\textsuperscript{*}, E.Sh. Mamasakhlisov\textsuperscript{a}, Shura Hayryan\textsuperscript{b}, Chin-Kun Hu\textsuperscript{b}

\textsuperscript{a}Chair of Molecular Physics, Yerevan State University, 1 Al.Manougian Str., Yerevan 375025, Armenia
\textsuperscript{b}Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan

Abstract

A model Hamiltonian for double-strand polynucleotides is suggested to describe the phenomenon of helix–coil transition. The Hamiltonian is constructed using solely the microscopical, pure physical quantities, characterizing the molecular chain, namely the energy of hydrogen-bond formation and the number of conformations of repeated unit. Realistic constraints are imposed on the conformations of chain in the case of loop formation. The advantage of the suggested approach is that the parameters of the model can be obtained from independent calculations or experiments. It is shown that with the approximation of neglecting the effect of large loops, the model of DNA is reduced to the generalized microscopical model of polypeptide chain. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: One-dimensional exactly solvable model; Cooperativity; Polynucleotides; Polypeptides

1. Introduction

The phenomena of helix–coil transitions in biopolymers have been a topic of intensive investigations for a very long time. A huge number of papers were published, in which many interesting theoretical models are suggested (for an earlier review see Ref. [1]). Almost all popular textbooks on statistical physics of polymers and molecular biophysics contain special chapters on this topic [2–5]. Experimental and theoretical investigation of helix–coil transition both in polypeptides and polynucleotides is also a topic of the day now [6–10].

Traditionally, the helix–coil transition is investigated by using the mean field approximation. In other words, the Hamiltonian of these models contain parameters, the values

\textsuperscript{*}Corresponding author.
\textit{E-mail address:} morozov@ysu.am (V.F. Morozov)
of which are averaged over all conformations of the molecule as well as over all states of solvent (e.g. cooperativity parameter and helix growth constant in Zimm–Bragg’s theory [11]). In Refs. [12–15] a different approach was suggested to construct the Hamiltonian of polypeptide chain. The model is based on the following assumptions:

1. In the framework of rotational–isomeric approximation, each repeated unit (amino acid residue) can exist in some \( Q \) different conformations.
2. Only one out of \( Q \) states corresponds to the conformation of \( \alpha \)-helix.
3. When \( \Delta \) successive repeated units are in the helical conformation, then intramolecular hydrogen bond occurs, with the energy \( U \). The model is very similar to one-dimensional Potts model with \( \Delta \)-particle interactions. From this model the main characteristics of helix–coil transition were obtained.

It is known [16,17] that one dimensional Ising model has been used to describe both the single- and double-strand biopolymers. Similarly, the purpose of this work is to extend the microscopical model of polypeptides to a double-strand DNA and also try to connect the cooperativity parameter of the double-strand macromolecule to the molecular characteristics of a single chain.

2. The model

As in the case of polypeptide chain, we construct the theory on the basis of the Hamiltonian which takes into account the essential features of the given structure. We consider the random heteropolymer which consists of (\( A–T \)) or (\( G–C \)) complementary pairs only. In this case the energy of hydrogen bond formation will be the same along the chain, while each strand itself is inhomogeneous. We can assume that the inter-chain hydrogen bonds are formed only between the bases with the same number.

We construct the model of such system as shown in Fig. 1. To each repeated unit \( i \) of one chain of the double helix a vector \( \vec{a}_i \) is assigned. Similarly, the vector \( \vec{b}_i \) is assigned to the \( i \)th unit of the opposite chain. One can consider these vectors as directed along the line connecting two adjacent sugar rings. We also connect a vector \( \vec{d} \) to each complementary pair of nitrogen bases. Formally, the magnitude of this vector is equal
to the total length of two bases and it emanates always from one chain. When the corresponding complementary pair is in the helical conformation this vector connects the ends of vectors $a_i$ and $b_i$. For simplicity, we assume that the first complementary pair is always in the helical conformation, i.e. the vector $d_0$ is always linked.

The condition for $i$th hydrogen bond to occur is

$$d_0 + \sum_{k=1}^{i} a_k - d_i - \sum_{k=1}^{i} b_k = 0 .$$

Let $\gamma_k = d_{k-1} + a_k - b_k - d_k$, then (1) may be rewritten as

$$\sum_{k=1}^{i} \gamma_k = 0 .$$

Therefore, to each complementary dinucleotide unit $k$ a vector $\gamma_k$ is connected. Note that if the hydrogen bond for the $i$th pair is formed then (2) is true, even if there are no hydrogen bonds between any of the preceding pairs.

Suppose in the rotational isomeric approximation each dinucleotide can take some $Q$ different equiprobable conformations. Of these conformations only one corresponds to the helical structure and a hydrogen bond occurs. This means that the vector $\gamma_k$ can take $Q$ different values and only the value $\gamma = 0$ corresponds to hydrogen bond formation. In other words, one can introduce the parameter $Q$ as follows:

$$Q = \frac{\text{Total number of states of dinucleotide}}{\text{Total number of states for helical structure}} .$$

When the different conformations of repeated unit are not equiprobable then

$$Q = \frac{\text{Partition function of repeated unit}}{\text{Partition function of repeated unit in helical structure}} .$$

Then, a Hamiltonian of the chain may be written as

$$-\beta H = J \sum_{i=1}^{N} \delta \left( \sum_{k=1}^{i} \gamma_k, 0 \right) \equiv J \sum_{i=1}^{N} \delta^{(1)}_{i} ,$$

where $\beta = T^{-1}$, $J = U/T$, $U$ is the energy of hydrogen bonds formation in one complementary pair, $\delta$ is the Kronecker’s symbol, and $\delta^{(1)}_{i} = \delta(\sum_{k=1}^{i} \gamma_k, 0)$.

Here an important notation must be done. It is well known that hydrogen bonds are not the only factor determining the stability of two-strand molecule of DNA [18]. Stacking interactions between bases are much more important. So, we use everywhere in this text, the expression “hydrogen bonds formation”, to describe the situation when two complementary bases are the “right position” to form the helix and $U$ is the energy required to bring these bases into helical structure. Note also, that this model cannot be applied to the triple-strand helices (H-form of DNA).
3. The partition function and the secular equation

Introducing the “temperature parameter” \( V = \exp(J) - 1 \), we may rewrite the partition function as

\[
Z = \sum_{\{Y_i\}} \prod_{i=1}^{N} (1 + V Y_i^{(i)}),
\]

which can be decomposed by the degrees of \( V \). Using the relation

\[
\phi_1^{(k)} \phi_1^{(m-k)} = \phi_1^{(k)} \phi_{k+1}^{(m-k)},
\]

we may write \( V f \) as

\[
V f^{(k_1)} \phi_2^{(k_2-k_1)} \phi_{k_2+1}^{(k_3-k_2)} \cdots \phi_{k_{i-1}+1}^{(k_i-k_{i-1})}.
\]

Assuming that the first and the \( N \)th repeated units are in the same conformation (cyclic conditions) and denoting \( m_i = k_i - k_{i-1} \), we eventually obtain

\[
Z = Q^N \sum_{f} V f \sum_{m_1} \phi(m_1) \sum_{m_2} \phi(m_2) \cdots \sum_{m_f} \phi(m_f),
\]

where

\[
\phi(m) = Q^{-m} \sum_{\ell_1} \sum_{\ell_2} \cdots \sum_{\ell_m} \phi^{(m)}.
\]

The function \( \phi(m) \) may be interpreted as the ratio of the partition function of the loop of \( m \) units to that of the same chain without loops.

\[
\phi(m) = \frac{\sum_{\ell_1} \sum_{\ell_2} \cdots \sum_{\ell_m} \phi^{(m)}}{\sum_{\ell_1} \sum_{\ell_2} \cdots \sum_{\ell_m} 1}.
\]

To carry out the summation in (8) we use the condition \( \sum_{k=1}^{f} m_k = N \) and introduce the factor

\[
\phi \left( \sum_{k=1}^{f} m_k - N \right) = \frac{1}{2\pi i} \oint P \sum_{k=1}^{f} m_k - N - 1 \, dP.
\]

Substituting (11) in (8) and assuming that all \( m \)-s are independent, we obtain

\[
Z = \frac{1}{2\pi i} \oint P^{-N-1} \sum_{f=1}^{N} \left( V \sum_{m=1}^{\infty} \phi(m) P^m \right)^f \, dP.
\]

After summation over \( f \), we have

\[
Z = \frac{1}{2\pi i} \oint P^{-N-1} \frac{1 - (V \sum_{m=1}^{\infty} \phi(m) P^m)^{N+1}}{1 - V \sum_{m=1}^{\infty} \phi(m) P^m} \, dP.
\]

For large \( N \), we can write for partition function and the free energy per repeated unit

\[
Z \approx P_0^{-N-1}
\]

\[
F = T \ln P_0.
\]
Here $P_0$ is the nearest to zero pole of the expression under the integral in (13). Hence the secular equation is

$$\sum P^m \varphi(m) = \frac{1}{V}. \quad (16)$$

The secular equation (16) contains two microscopical quantities: temperature parameter $V$, which is defined by the energy of interchain hydrogen bonding and $\varphi(m)$ which is the relative statistical weight of the loop of length $m$. Both quantities can be in principle measured in experiment or calculated by other independent methods.

4. The generalized model of polypeptide chain

In this section, we will use Eq. (16) to discuss the different behaviour of $\varphi(m)$ for large and small loops. For this purpose, we introduce some characteristic scale $\Lambda$ for rigidity of the chain. In other words, we consider each chain as freely jointed and consisting of rigid segments of length $\Lambda$, and then consider a loop of $m$ complementary pairs.

When $m \leq \Lambda$, the loop can be formed in only one fixed conformation. Then $\varphi(m) = Q^{-m}$, i.e. for small values of $m$ the function $\varphi(m)$ is exponentially decreasing. For $m > \Lambda$, a loop can be realized by many conformations of the chain and the behaviour of $\varphi(m)$ is changed from exponential decreasing to $\varphi(m) = Q^{-\Lambda} - \psi(m)$, where $Q^{-\Lambda}$ is the last term of exponential decreasing of $\varphi(m)$, and the function $\psi(m)$ describes the behaviour of large loops. Therefore,

$$\varphi = \begin{cases} Q^{-m}, & m \leq \Lambda, \\ Q^{-\Lambda} - \psi(m), & m > \Lambda. \end{cases} \quad (17)$$

Using the notation $\lambda = Q/P$, we can write the secular equation (16) in the form

$$\lambda^{\Lambda-1} [\lambda - (V + 1) + L(\lambda)(\lambda - 1)](\lambda - Q) = V(Q - 1), \quad (18)$$

where

$$L(\lambda) = \sum_{m=\Lambda+1}^{\infty} \left( \frac{Q}{\lambda} \right)^m \psi(m). \quad (19)$$

This term is present in many theories of the helix–coil transition in double-strand biopolymers [1,16,17]. It represents the so-called “loop factor” and it is usually being added to the single chain approximation, to proceed to the double chain.

Let us analyse Eq. (18) and first neglect the constraints imposed on the chain conformation due to loop formation. In other words, let $\psi(m) = 0$. The secular equation (16) can be written as

$$\lambda^{\Lambda-1} [(\lambda - (V + 1))(\lambda - Q)] = V(Q - 1). \quad (20)$$
We see that this secular equation is reduced to the secular equation of the generalized model of polypeptide chain [13], defined by the Hamiltonian

$$-\beta H = J \sum_{i=1}^{N} \prod_{k=0}^{A-1} \delta(\gamma_{i+k}, 1),$$  \hspace{1cm} (21)$$

where the sum is over the set \( \{\gamma_i\} \) which defines the conformational space of the chain. For real polypeptides the value of \( A \) is taken as equal to 3. Other values of \( A \) have been used to analyse in detail the influence of the geometry of hydrogen bonds on the process of helix-coil transition. There is an essential difference between Hamiltonian (5) of double strand chains and Hamiltonian (21) of the polypeptide chain. Eq. (21) is written in conformational space, while (5) in the ordinary three-dimensional space.

Earlier it was shown [13,14] that the Zimm-Bragg parameters for polypeptide chain can be represented in the following form:

$$S = V + \frac{1}{Q}, \quad \sigma = Q^{1-A}$$ \hspace{1cm} (22)$$
in some approximation. In the Zimm-Bragg theory, the helix-coil transition occurs at the value \( S = 1 \). Here it corresponds to \( V = Q - 1 \), whence

$$T_m = \frac{U}{\ln Q}.$$ \hspace{1cm} (23)$$

In the case of double-strand chains the parameter \( Q \) should be interpreted as the ratio of the partition function of hydrogen-bonded dinucleotide to the partition function of dinucleotide, which is free of hydrogen bonds. The parameter \( A \) determines the scale on which the hydrogen bond in the loop of \( A \) units can occur in a unique way. So, with good approximation, \( A \) could be identified with the length of statistical segment of isolated polynucleotide chain. Unlike the polypeptides, where the region of helical state is very small part of accessible conformational space [19] and \( Q \) is of the order of tens, in polynucleotide chain the region of helical state is comparable to the whole conformational space. It means that \( Q \) is of the order of unity. Estimations from the conformational analysis [20] give \( Q \approx 3 \) and experimental value of the length of statistical segment for single chain DNA is about 10–15 [21]. Using these values in (22), we get \( \sigma = 10^{-5} - 10^{-7} \). Hence, the transition in polynucleotides is much sharper than that in polypeptides, which is in agreement with experimental data. This value for \( \sigma \) is obtained also by other authors using mean-field approach [17]. Thus, we can state that the sharpness of helix-coil transition in homopolynucleotides is mainly due to the rigidity of single chain.

For large \( U \) (2 or 3 hydrogen bonds per unit) and small \( Q \), the expression (23) gives very large values for \( T_m \). The energy of hydrogen bond [22] is about 38 hj/mole. Therefore, \( T_m \approx 10^4 \). We believe that this unrealistically large value for the melting temperature does not mean necessarily that the model is incorrect. As in the case with polypeptides, it may be related to the fact that the solvent effect is not included in the model. This value may also be improved by the exact calculation of the \( Q \) parameter from the conformational maps of oligonucleotides and using the experimental melting...
curves of heteropolymers, consisting of A–T or G–C pairs only. Right now, it is difficult to discuss this question more precisely since the model with the solvent included, is not yet available.

In the case of polypeptide model with the solvent, competitive for hydrogen bond formation [12], the equation 
\[ V = Q - 1 \]
for the melting temperature still holds, with \( V \) replaced by
\[ \tilde{V} = \frac{q^2 e^{U/T}}{e^{E/T} + q - 1}. \]  
(24)

Here \( U \) is the energy of intramolecular hydrogen bond, \( E \) is the energy of intermolecular hydrogen bond between molecules of the solvent and polypeptide chain, \( q \) is the number of orientations of the solvent molecule with respect to the N–H or C==O groups. Using reasonable values for \( Q \) and \( q \) [14,15], one obtains for the melting temperature the value \( T_m \approx 10^2 \), which is close to realistic one.

So, by neglecting the constraints on the large scale loops, we can describe the phenomenon of helix–coil transition in polynucleotides in terms of the generalized model of polypeptide chain. These results were obtained for homogeneous polymers.

Let us try to consider now the more interesting case of heteropolymers. According to Refs. [12,13], the transfer matrix for Hamiltonian (21) can be written in two forms:

\[
\hat{G} = \begin{pmatrix} V & V & \cdots & V & V \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} = \begin{pmatrix} V & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \]  
(25)

\[
\hat{M} = \begin{pmatrix} e^J & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & Q - 1 \\ 1 & 1 & 1 & \cdots & 1 & Q - 1 \end{pmatrix} = \begin{pmatrix} e^J & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \]  
(26)

For simplicity, we use the notations
\[ \hat{G} = \tilde{V} \cdot \hat{G}_1, \]  
(27)

\[ \hat{M} = \tilde{Q} \cdot \hat{M}_1. \]  
(28)

Here \( \tilde{V} \) and \( \tilde{Q} \) are the diagonal matrices. Both in DNA and in polypeptides the backbone chain is homogeneous and the difference between repeated units is due to differences in side groups.
As the hydrogen bonds in polypeptides are formed by backbone groups, the polypeptide is homogeneous for hydrogen bond energy and hence for temperature parameter \( V \). And the polypeptide chain is heterogeneous for the parameter \( Q \) because the side groups are very different. In DNA the situation is different. The helical structure is stabilized by hydrogen bonds between side groups (nitrogen bases) and the energy depends on the type of complementary pair (2 bonds for \( A-T \) pairs and 3 bonds for \( G-C \) pairs). Hence the DNA molecule is homogeneous for \( Q \) parameter and heterogeneous for \( V \). This means that for polypeptide chain it is more suitable to use Eq. (26).

Using (26), Hayryan et al. have shown in Ref. [13] that because the polypeptide chain is homogeneous in \( V \) and heterogeneous in \( Q \), the Zimm–Bragg parameters of helix–coil transition \( s \) and \( \sigma \) may be written as

\[
S_i = \frac{e^{J_i}}{Q_i}, \quad \sigma_i = \prod_{k=1}^{A-1} Q^{-1}_{i-k}.
\]

(29)

Here the product is over \( A - 1 \) amino acids. It is obvious that in the case of homopolypeptide, (29) is reduced to (22). In (29), it is shown that the cooperativity parameter \( \sigma \) depends on the types of \((A - 1)\) amino acids while \( S \) depends on the type of only one amino acid.

This is the reason that one can consider the cooperativity parameter \( \sigma \) of heteropolypeptide chain as non-local parameter while \( S \) as local.

Using (25) for DNA, we can similarly obtain

\[
S_i = \frac{e^{J_i}}{Q}, \quad \sigma_i = Q^{1-A}.
\]

(30)

Eq. (30) means that the cooperativity parameter is constant in the approximation of independence of conformational states on the type of nitrogen bases.

5. Conclusion

We suggest a Hamiltonian for the DNA molecule to describe the helix–coil transition. The conformational feature of double-strand DNA to form loops through the process of transition is taken into account. Unlike mean-field theories of helix–coil transition, there are no empirical parameters in our model. All parameters can be calculated or measured in experiments, independently from melting experiments.

More detail comparison with experimental data may be done after calculation of parameter \( Q \) from conformational analysis. The parameter \( A \) can be obtained from viscosimetric measurements.

Acknowledgements

This work was supported in part by the National Science Council of the Republic of China (Taiwan) under grant number NSC 89-2112-M-001-005.
References