Yerevan State University

Gor Norayr Hayrapetyan

Loop factor in conformational transitions of nucleic acids

03.00.02 – Biophysics

THESIS

for the PhD degree in Physics

Scientific Supervisor

Dr. Sci., Associate Prof.

Y. Sh. Mamasakhlisov

Yerevan – 2013

Table of Contents

Introduction

Relevance of the work. One of the main problems in the physics of macromolecules is determination of the physical laws defining the structure and biological function of the single-and double-stranded nucleic acids. It is well known that the biological function of biopolymers is determined by their spatial structure. In this regard, it is important to determine the main factors and patterns affecting conformations and conformational transitions. One of these factors is the formation of the secondary structure of nucleic acids and the occurrence of long loops in the areas free from Watson-Crick base pair formations. In addition, there remain a number of open questions related to the effect of a heterogeneous sequence of nucleotides on the structure and conformational transitions in nucleic acids. The study of these questions is also interesting in terms of bioinformatics. The algorithms for the calculation of the thermodynamic parameters and optimal prediction of the secondary structure of single-stranded RNA are now widely used in biological research. Proper account of loop entropy and sequence heterogeneity will significantly improve the existing algorithms and promote the development of new approaches to the problem. In light of the above, the relevance of this work is determined by the development of new approaches to the study of conformational entropy of loops and effects of the nucleotide sequence.

The objectives are:

1. construction of the theory of melting of double-stranded DNA, which takes into account the topological restrictions imposed on long loops
2. investigation of the excluded volume effects in the formation of long loops in the double-stranded DNA
3. study of the effect of long loops on the phase behavior of the double-stranded DNA
4. construction of analytical theory describing the thermodynamic properties

of single-stranded RNA with a random sequence of nucleotides

5. comparative analysis of the phase behavior of ssRNA with and without account of entropy of long loop formation
6. calculation of the thermodynamic characteristics of the ssRNA

The scientific novelty consists in constructing a model of melting of the DNA double helix, without resorting to any prior assumptions about the entropy of long loop formation. Analytical dependence of the loop statistical weight is calculated based on the proposed theory, and not chosen from physical considerations. Temperature dependence of such characteristics of the helix-coil transition as the free energy, helicity degree, correlation length and correlation function was obtained. For the first time in melting of double-stranded DNA the existence of an infinite order phase transition was shown. The comparison with the results obtained in the framework of the Polanda and Scheraga was performed. An analytical theory based on the method of annealing with constraints describing the secondary structure formation of single-stranded RNA was obtained. The comparison with the numerical results shows reasonable quantitative agreement. The temperature dependence of the characteristics of single-stranded RNA denaturation as the free energy, helicity, entropy and heat capacity was obtained. For the first time the presence of two structural transitions for a random sequence of nucleotides of two types was shown. Possible connection between the low-temperature structural transition and cold denaturation of ssRNA, which is observed experimentally, was demonstrated.

The main provisions to be defended

1. Impossibility of knot formation in melted DNA and the account of excluded volume effects impact on the loop entropy and, in this way, result in a value of loop factor $c = 1$.
2. For the loop factor value $c = 1$ infinite order phase transition takes place during denaturation of DNA double helix. Near the critical temperature the correlation length diverges, as it happens during the phase transition

of the second order, whereas amplitude of the fluctuations tends to zero. Thus there are small but extended fluctuations.

3. Above the transition temperature the helicity degree is zero, which differs significantly from the behavior of the system at usual helix-coil transition. Single-stranded RNA with random nucleotide sequence shows two peaks in the temperature dependence of the specific heat of the system for a certain choice of interaction parameters. Such behavior indicates the presence of two structural transitions.
4. Low-temperature peak of the specific heat corresponds to the cold melting of RNA, when the helicity degree decreases significantly with temperature decrease. This effect is due to a large number of thermodynamically unfavorable contacts for sequence consisting of two types of nucleotides.
5. The account of long loops entropy qualitatively does not affect the behavior of ssRNA. The presence of two peaks and cold melting is observed at the same values of the interaction parameters as without account of loop entropy.

The scientific and practical value of the work is due to a significance of the role that thermodynamic effects play in the functioning of biological macromolecules and their complexes. In this regard, the theoretical study of the conformational entropy of large loops, effects of interactions between different types of nucleotides, and other characteristics of biological macromolecules is important for the interpretation of experimental results and their predictions. At the same time, understanding the basic principles underlying the organization and conformational changes in biological macromolecules is of great practical importance for solving problems in biology and its medical applications. Furthermore, the obtained results, certainly, enable the improvement of the existing bioinformatics algorithms used to calculate the stability of the secondary structure of RNA.

Approbation of the work. Materials of the thesis were presented at

- Taiwan International Workshop on Biological Physics and Complex Systems,

Taipei, Taiwan, July 21-26, 2011.

- Winter School on Calculus of Variations in Physics and Materials Science at Department of Mathematics, University of Wurzburg, Wurzburg, Germany, January 8-13, 2012.
- II Gefenol Summer School on Statistical Physics of Complex and Small Systems, Centro de Ciencias de Benasque Pedro Pascual, Spain, September 3-14, 2012.
- International Young Scientists Conference "Perspectives for Development of Molecular and Cellular Biology–3", The Institute of Molecular Biology NAS RA, Yerevan, Armenia, September 26-29, 2012.

Publications. On the topic of the thesis 8 papers are published.

Structure of the thesis. The thesis consists of an introduction, four chapters and conclusions (114 pages of text). It contains 59 figures and bibliography consisting of 103 items. The objectives of the work, the scientific novelty and practical value of the results and the main provisions to be defended are stated in the introduction. The first chapter is devoted to the review of structure, thermodynamics and biological functions of nucleic acids. The main properties of classical models of DNA are represented. Also, the literature review contains a description of the basic models of RNA secondary structure and dynamic algorithms for the calculation of its thermodynamic parameters. At the end of the first chapter the statement of the problem addressed in the second chapter is formulated. The second chapter is devoted to the DNA model which takes into account the entropy of long loops formation. In the framework of this model the main thermodynamic parameters of the system are calculated. The third chapter is devoted to the thermodynamics of the secondary structure of ssRNA with random heterogeneous sequence. The calculation of thermodynamic functions is based on the constrained annealing approach. The fourth chapter of the thesis is devoted to the influence of loop entropy on the thermodynamic properties of the secondary structure of ssRNA. The work ends with the conclusions.

Chapter 1

Literature review

1.1. The structure and biological functions of nucleic acids

1.1.1. The structure of nucleic acids

The nucleic acids are linear polymers with monomers called nucleotides. A nucleotide consists of a sugar ring, phosphate group and a nitrogenous base. The backbone of the nucleic acid consists of ribose sugar rings linked by phosphate group. Each sugar has the one of the four types of nitrogenous bases linked to it as a side group. The $5'$ carbon of one ribose and the $3'$ carbon of the next are linked by phosphate group. So, the direction of chain is $5'3'$. The two ends are referred to as $5'$ and $3'$ ends, since one end has an unlinked $5'$ carbon and one has an unlinked $3'$ carbon. There are two types of sugar rings: ribose and deoxyribose. Let's refer to the chemical differences between ribonucleic acid (RNA) and deoxyribonucleic acid (DNA). The first difference is represented in the chemical names of RNA and DNA, since one of the OH groups in ribose is replaced by proton (H) in deoxyribose. The second difference is that, in contrast to RNA, DNA comprises thymine (T) bases instead of uracil (U) bases. In other words, the nitrogenous bases in the RNA are adenine, cytosine, guanine and uracil (A, C, G, and U), while DNA consist of adenine, cytosine, guanine and thymine (A, C, G, and T). The third difference is that RNA usually occurs as single strands and DNA consists of two strands. As a result, RNA and DNA have distinctive varieties of structures. The double helical structure of DNA has two strands that are perfectly complementary in sequence. In RNA base pairs are formed intra-molecularly, leading to a complex arrangement of short helices which are the basis of the secondary structure. Some tertiary structures

of RNA are well-defined. Thus, RNA structures are more similar to globular structures of proteins than to double helices of DNA. The main role of DNA is to save the genetic information. The role of proteins is to serve as biochemical catalysts. These roles have been recognized for a long time, and it was thought that RNA is an intermediary between proteins and DNA. But now we can say that RNA is coming to be seen as an important and diversified molecule in its own right. Let's present the types of RNA.

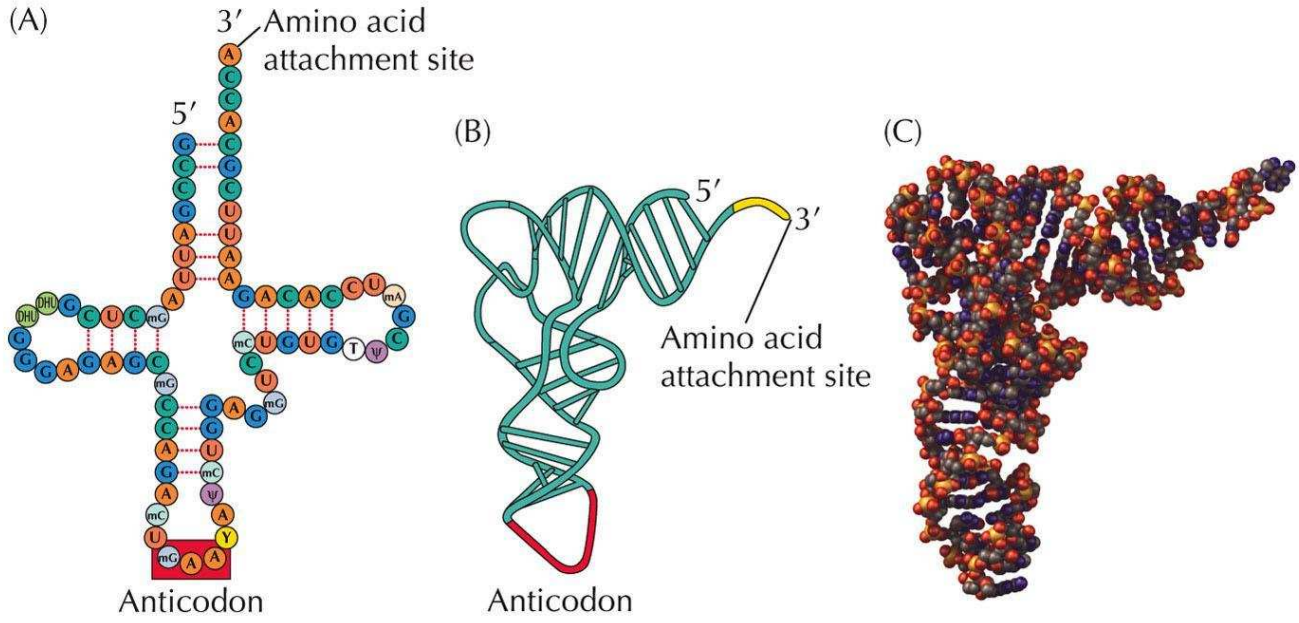

Fig 1.1. The secondary an tertiary structures of transport RNA.

1.1.2. Types of RNA

Transfer RNA (tRNA)

The common number of nucleotides in tRNA is about 76 [1, 2]. Its secondary structure is called *clover-leaf* and it is very well-defined (Fig. 1.1). Every amino acid has the own tRNA. The middle three bases of the central loop of tRNA

compose the anticodon. The codon in the mRNA and the anticodon in the appropriate tRNA are the same. The main role of tRNA is to bring the amino acid in the ribosome during protein synthesis. The shape of the tertiary structure of the tRNA has the form like letter L. Fig. 1.1 shows the clover-leaf secondary and the L-shaped tertiary structures of tRNA.

Messenger RNA (mRNA)

The mRNA has several thousand nucleotides. It is the copy of the part of one of the strands of DNA and it contains the information about a protein which has to be synthesized by ribosome. The central portion of mRNA codes the protein.

Ribosomal RNA (rRNA)

The protein synthesis takes place in the ribosome. It possesses binding sites for mRNA and tRNA. It is the main role of the ribosome. Its diameter is about 250 Å. The ribosome is composed of two sub-units. Each of them consists of three rRNA and about 56 different proteins [3–5]. The main goal of ribosome is to perform one of the most important processes in the cell – the protein synthesis. It has the sites that can bind tRNA and mRNA. During the protein synthesis it moves along the mRNA. Thus we can say that tRNA molecules have the very important role in the functioning of the ribosome, and as a result, protein synthesis cannot be implemented without these molecules. Ribosomal RNAs of many organisms are sequences, and large databases are accessible giving their structural models [6–8].

1.1.3. The elements of the secondary structure of RNA

If there are two complementary parts of the sequence in the RNA molecule, those parts can form helical structures. There are possible hydrogen bonds between nitrogenous bases C–G and A–U. There may be link between G-U,

but this pair is less stable. As a rule, helices consist of at least two pairs, because isolated pairs usually are unstable. In unbroken helices there are not more than 10 pairs. There are attractive stacking interactions between base pairs. They have a great contribution in the stability of the helix. The stacking interactions are in the approximately parallel planes. To find the free energy of the helix usually nearest neighbor model is used. That is to say, there is a free energy term for every two near base pairs. Using different methods we can measure the energy and entropy changes of helix formation in the experiments, when the sequence is short [9].

In the Fig. 1.2 it is shown several structures that can occur between helices in the single-stranded RNA. Hairpin loops connect the two sides of a single helix. The loops which connect two helices are called *internal*. The loops that connect three or more helices are called *multi-branched*. Bulge loops, stems and pseudoknots are also common to single-stranded RNA (Fig. 2). The pseudoknots will be discussed later. Free energy of some loop structures have been measured experimentally, but, as a rule, the helix parameters are known with higher accuracy than the parameters of loops [10]. For instance, we don't have any thermodynamic data about multi-branched loops. So, we suppose that independence of loop free energy on nucleotide sequence. It hinge on the number of unpaired bases in the loop. The exceptions to this are tetraloops. Tetraloops are special sequences that consist of four single-stranded bases. Thanks to these structures the thermodynamic stability rises at the expense of interactions between the unpaired bases in the length-four hairpin loops, where they often occur. In the algorithm that predicts the secondary structure we have to appoint a free energy for every possible structure. After that we must compare the stabilities of all these structures. Instead of thermodynamic parameters that are not directly measured, we can take the reasonable estimates. The free energy of the secondary structure of all sequence will be determined through the free energies of different parts of chain.

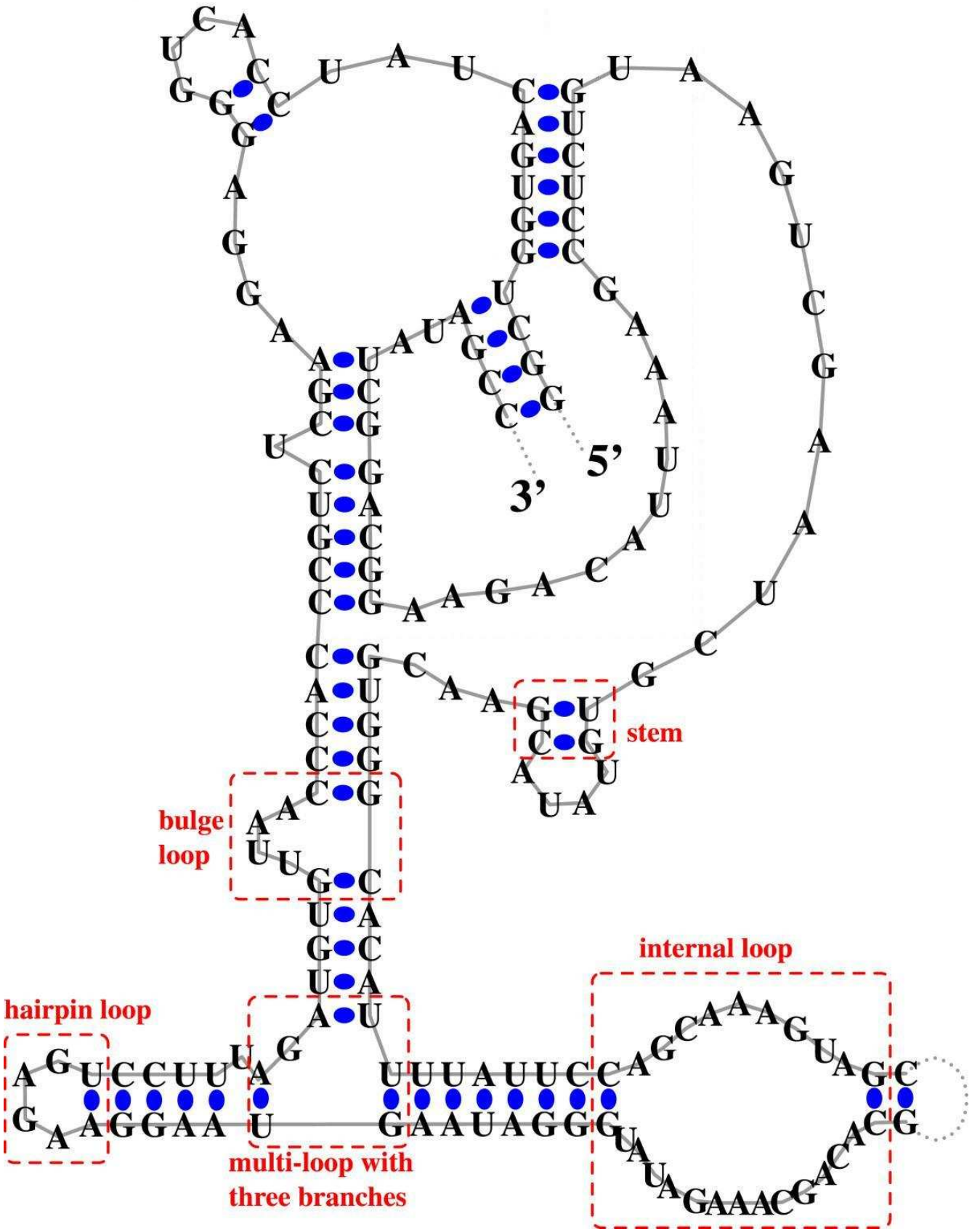

Fig 1.2. Some structures that can be formed in the single-stranded RNA

1.1.4. The tertiary structure of nucleic acids

The progress of secondary structure determination goes on faster than for tertiary structure. Until recently we had a little experimental information

about tertiary structure. In this review we will speak more about secondary structures. We will address the information that can and cannot be obtained from secondary structure alone. Although our information about tertiary structure recently rises, we assert that the information about secondary structure is very important too. The secondary structure is the figure that shows the list of base pairs that are in the structure. In the valid secondary structure base pairs have to satisfy some limitations. Let us suppose that we have the chain consists of bases that are numbered from 1 to N. Let us assume that the bases i and j are complementary. They can form a pair, if . In other words, there must be three or more unpaired bases in the hairpin loop. Let us suppose that there are formed pairs between (i, j) and (k, l) . They can be compatible if they can be in the chain simultaneously. For that they must be non-overlapping $(i < j < k < l)$ or one of them must be within other $(i < k < l < j)$. The structure where they are interlocking $(i < k < j < l)$ is called pseudoknot (Fig. 1.3). A lot of dynamic programs cannot consider the existence of pseudoknots. In the valid secondary structure all base pairs must be consistent. The secondary structure of given sequence shows the information about paired and unpaired bases and it cannot give us any information about the tertiary structure of the sequence. We can add to the diagram of secondary structures pseudoknots. If we have the information about tertiary structure, it will be more comfortable to change the secondary structure. The parts of the chain that are close in the tertiary structure we can draw near each other in the secondary structure. Thanks to this, the secondary structure of chain will contain some information about it tertiary structure. As a rule, the diagrams of secondary structures are not drawn thus to contain a lot of information about tertiary structures. Nonetheless, the secondary structure of RNA can give us enough information about its tertiary structure. We can gain the information about the domain structure of molecule and the mutual positions of the important parts. So, the secondary structure of RNA contains much more information about the shape of its molecule then

the appropriate diagram of secondary structure of proteins which is a linear polymer that consists of α helices and β sheets.

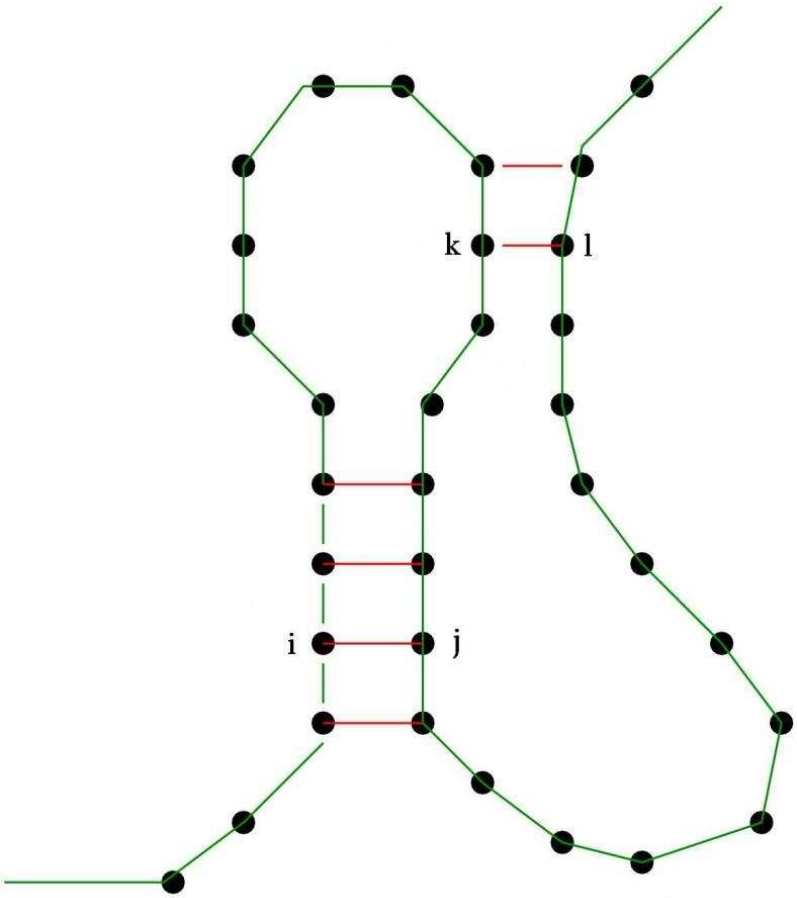

Fig 1.3. The schematic image of pseudoknot.

The main advantage of secondary structures of RNA is that the helices are thermodynamically very strongly bonded. The hierarchical folding of RNA means that first forms the stable secondary structure [11–13]. Afterwards the tertiary structure forms since a molecule can bend around some areas. The interactions in the tertiary structures can change only the weak elements of secondary structure. It is so, because their strength is too small to break the secondary structure. Those interactions can change the positions of bases in the more unstable helix. Unlike the RNAs, very often secondary structure elements in the proteins are enough unstable on their own. So, it is very difficult to separate their secondary and tertiary structures. As a rule, we ignore the existence of pseudoknots when we determine the parameters that describing the secondary structure. There are a lot of reasons for that. One of them is that the algorithm that allows us to predict the structure cannot account for

pseudoknots. For example, in the small sub-unit rRNAs the number of non-overlapping and nested helices is much more than the number of pseudoknots. So, in this case we can obtain the sufficiently accurate results without incorporating the contribution of pseudoknots. But it is obviously that some types of the pseudoknots frequently occur in the RNA and they may have functional role. Now we have a lot of data about the secondary and tertiary structures of pseudoknots [14–17]. As a result, the new dynamic programming algorithms are able to take into account pseudoknots [18]. The main problem of these algorithms is absence of information on pseudoknots thermodynamic that is needed.

1.2. Thermodynamics of nucleic acids

In this section we will discuss general mechanisms of DNA melting and relating experimental results which are represented in the review [19].

According to the previous section the deoxyribonucleic acid (DNA) consists of two polynucleotide strands. They are twisted into a double helix as it is shown on Fig. 1.4. Those two strands are perfectly complementary. In DNA there are 2 hydrogen bonds between nitrogenous bases adenine and thymine and 3 hydrogen bonds between cytosine and guanine. The diameter of DNA is about 20Å. The distance of two neighboring repeating units is approximately 3.4Å. Each twist of DNA consist of ten to twelve repeating units depending on the form of DNA (A, B, Z). Dividing 360° over the number of nucleotides in the twist one will obtain twist angle for one repeating unit.

One of the most fundamental thermodynamic processes taking place in DNA is melting. This process is also called the helix-coil transition. The scheme of DNA melting is represented on Fig. 1.5. During this process the hydrogen bonds between nitrogenous bases are being destroyed, and, in the final stage, there are two separate DNA chains, which can be dealt as Gaussian coils.

The helix-coil transition is reversible process. That is to say, the decrease

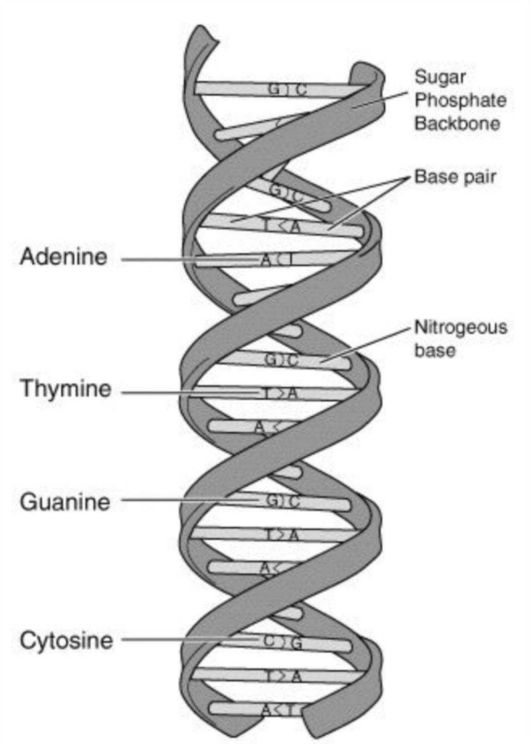

Fig 1.4. The double helical structure of DNA.

of temperature can lead to the renaturation of DNA. But if DNA is completely melted, the probability of recreation of existed helical structure tends to zero. This is result of very large influence of kinetic factors. Now let's speak about experimental data concerning DNA melting.

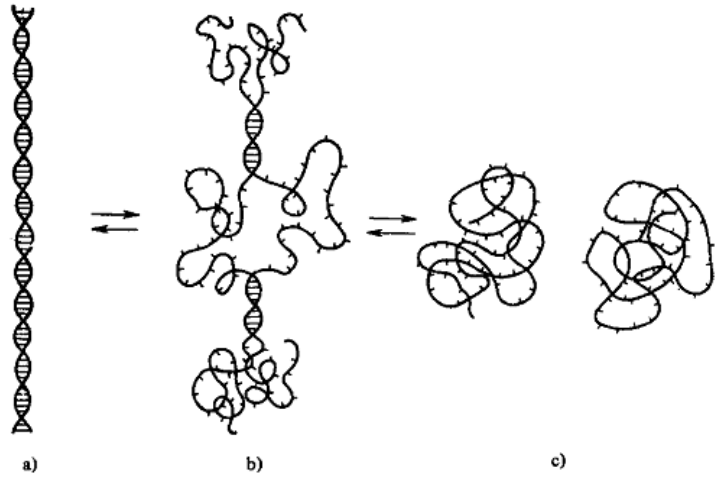

Fig 1.5. The scheme of the helix-coil transition in DNA.

There are a number of methods that allow us to study the helix-coil transition in DNA experimentally. One them is based on absorption of visible –UV radiation by DNA solution. The method is based on the structural dependence of absorption property of DNA. The absorptions of nucleotide bases is deferent for helical and coil regions [20]. It is caused by the absence of stacking interactions

in coil regions in contrast to helical. The quantity$(D - D_{min})/(D_{max} - D_{min})$, where D is the optical density of solution, and the D_{min} and D_{max} are optical densities of helical and coil structure correspondingly, relates to the degree of denaturation. Fig. 1.6 shows the temperature dependence of optical density for double stranded homopolynucleotide (melting curve). The melting curves for homopolynucleotide were studied in [21]. One can characterize the melting curve through two parameters: the melting temperature (T_m) and the width of melting interval (ΔT). The width of melting interval is determined with the formula

$$\triangle T = 1/\left|\frac{\partial\theta}{\partial T}\right|_{max}. \tag{1.1}$$

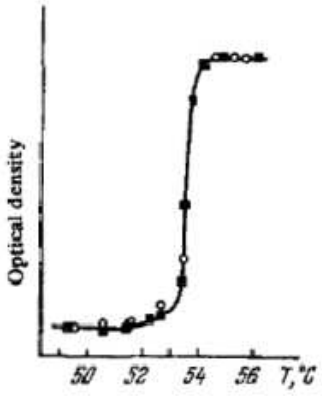

Fig 1.6. The melting curve for homopolynucleotide [22].

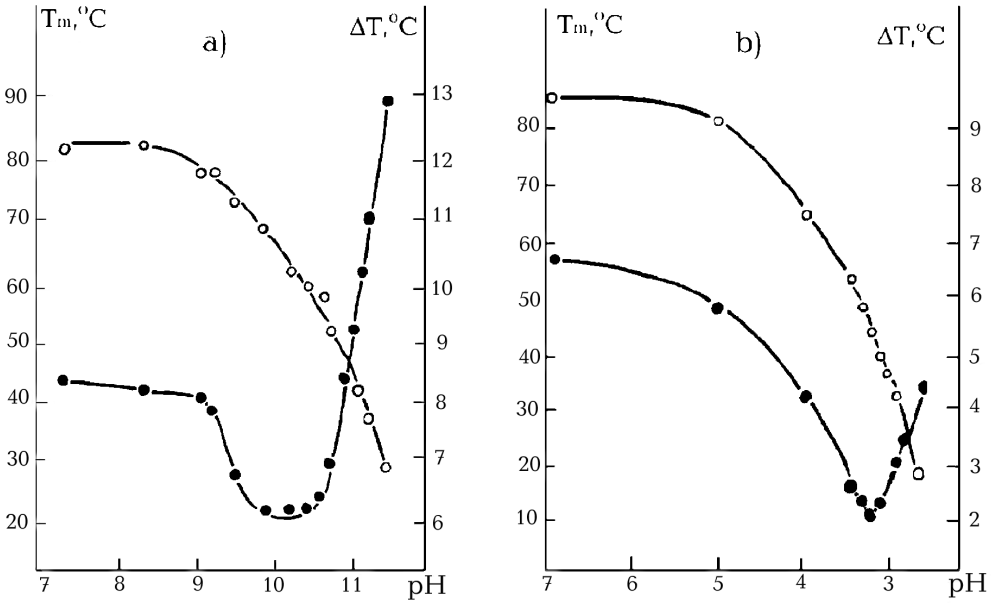

Fig 1.7. Temperature dependence of melting temperature T_m ($\circ$) and melting interval ΔT ($\bullet$) of calf thymus DNA [23, 24].

One of the main characteristics of melting curve is the GC composition of DNA. The dependence melting temperature on GC composition is shown in Fig. 1.8. The G-C composition is defined as

$$x_0 = (N_G + N_C)/(N_G + N_C + N_A + N_T), \tag{1.2}$$

where N_A, N_T, N_C and N_G are the numbers of adenine, thymine, cytosine and guanine nitrogen bases. It is seen from Fig. 1.8 that the dependence of T_m on x_o is linear.

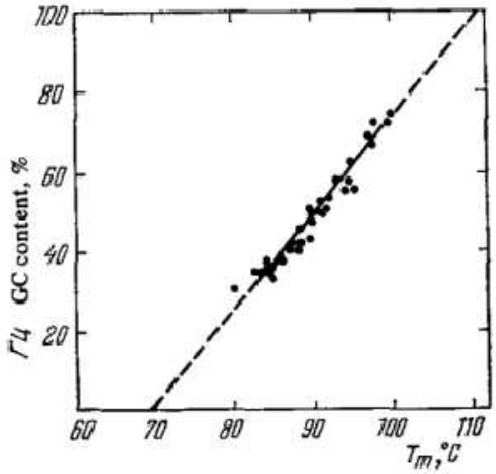

Fig 1.8. Dependence of melting temperature on the G-C pairs and the melting temperature [25].

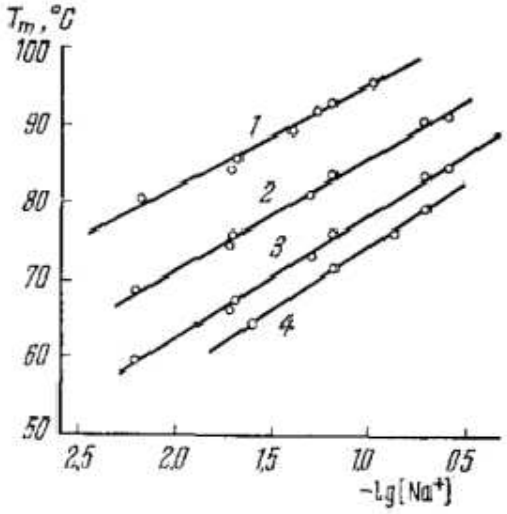

Fig 1.9. The relation between percentage of logarithm of the concentration of sodium in the solution. Line 1 was obtained for *M.* textitlysodeikticus ($x_0 = 0.72$), line 2 – *E. coli* ($x_0 = 0.5$), line 3 – *S. saprophyticus* ($x_0 = 0.33$), line 4 – *M. mycoides var capri* ($x_0 = 0.24$).

The melting temperature of DNA essentially depends on the solvent composition. The existence of double-helical structure of DNA is possible in the environment

with sufficient concentration of positive ions such as sodium and potassium ions. In case of neutral pH one can use the empirical formula for the melting temperature:

$$T_m = 176 - (2,6 - x_0)\left(36 - 7,04 \cdot \lg\left[\mathrm{Na}^+\right]\right), \tag{1.3}$$

where $\left[Na^+\right]$ is the molecular concentration of sodium ions. The dependence of melting temperature on the sodium ion concentration logarithm is shown in Fig. 1.9 [26]. The Fig. 1.9 was obtained through the formula (1.3) four different DNA. The melting temperature is much lower when $pH < 5$ or $pH > 9$. For mentioned DNA the width of melting interval is about 3°. For homopolynucleotide this parameter is nearly $0,5^\circ$. The main part of studying are done in the standard conditions ($pH = 7$, $\left[Na^+\right] = 0,196\ M$).

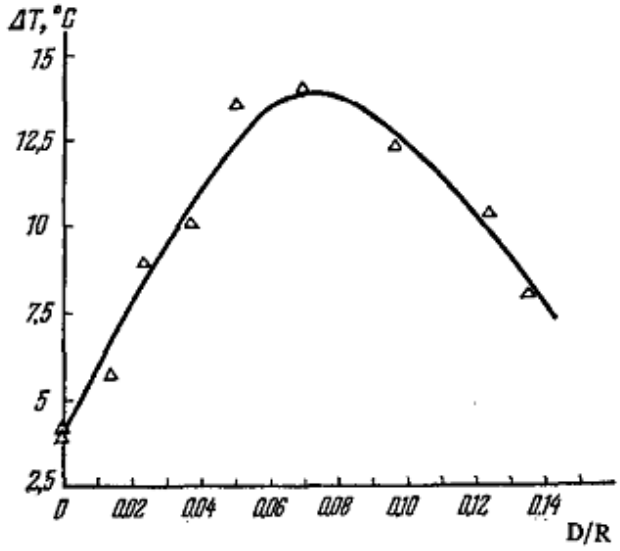

Fig 1.10. The dependence of the width of melting interval on the concentration of ribonuclease in the solution [27].

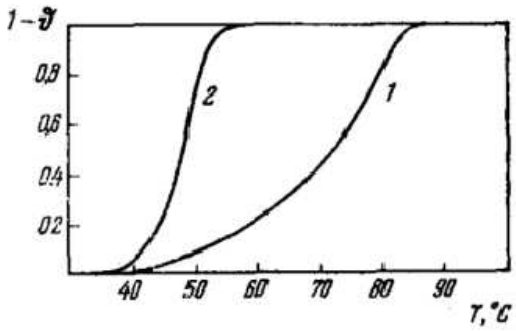

Fig 1.11. The melting curves for circular, closed polyoma DNA (1) and for the same DNA with the broken strand in 7.2 M $NaClO_4$ solution (2).

The substances that can bond to DNA, also known as ligands, have very important impact on the melting curves. For instance, such substances are heavy metal ions (Cu, Fe, etc.). As an example of influence of organic ligand, the dependence of ΔT on the concentration of the protein ribonuclease (D is the molar concentration of the protein ribonuclease in the solution, P is the molar concentration of repeated units in DNA) is shown in Fig. 1.10. Normally $D \ll P$. It is important to say that during experiment those ligands are redistributed on DNA. At a given temperature they take thermodynamically the most advantageous state. The experiments are performed for linear unclosed double stranded DNA. In the case of the circular closed DNA, then the experimental results are deferent. The characteristics of melting curve in this case were studied in [28]. The melting temperature for such DNA is higher by 20° compared to the linear unclosed DNA (Fig. 1.11). Melting chains remain twisted relative to each other in the circular DNA causing higher melting temperature. As a result, in this case the entropy of melting condition is lower than for the same condition in the linear DNA. In addition, the width of melting interval for the circular DNA is 2-3 larger times than for the linear DNA.

1.3. Dynamic programming algorithms

The most stable secondary structure of RNA molecule is specified by the minimum of the free energy. We can obtain such structures considering all possible base pairings and calculating the free energy for each secondary structure [29]. It is possible for very short sequences, since the number of possible conformations of the molecule grows exponentially with the length of the RNA molecule. However, there exist dynamic programming algorithms which allow calculation of the free energy for much longer sequences. These algorithms are based on recursive relations, which allow obtaining the thermodynamic quantities for longer sequences referring to already obtained ones for short sequences. Now we will discuss the programming algorithm for a very simple

set of energy rules. In the frameworks of this model we will suppose that each base pair contributes -1 in the energy of whole chain and penalties related with loops are neglected. So, the structure with the minimal free energy is characterized with the maximum number of base pairs. Therefore, this model is called "maximum matching model" [30]. Let us suppose that the energy of bonding between i and j bases ($\epsilon_{i,j}$) is -1 if those bases are complementary and it is ∞, if they are not. Our aim is to find the minimal energy of the subchain from i to j ($E_{i,j}$). If the last base j forms a pair with the base k, then the sequence will be divided into two subsequences: from i to $k-1$ and from $k+1$ to $j-1$. We will not discuss structures containing pseudoknots. In the other words, the bases that are in the different sections cannot form a pair. If j and k form a pair, the energy will be equal to $E_{i,k-1}+E_{k+1,j-1}+\epsilon_{i,j}$. If they do not form a pair, it will be equal to $E_{i,j-1}$. So the minimal energy of this subchain is

$$E_{i,j} = min(E_{i,j-1}, min_{i\leq k\leq j-4}(E_{i,k-1} + E_{k+1,j-1} + \epsilon_{i,j})). \qquad (1.4)$$

We will assume that $E_{i,j} = 0$, if $j-i \leq 4$. Thereafter, we will find $E_{i,j+1}, E_{i,j+2}$ and so on. As a result we can obtain the minimal energy $E_{1,N}$ of whole chain with length N. This algorithm estimates the contribution of individual base

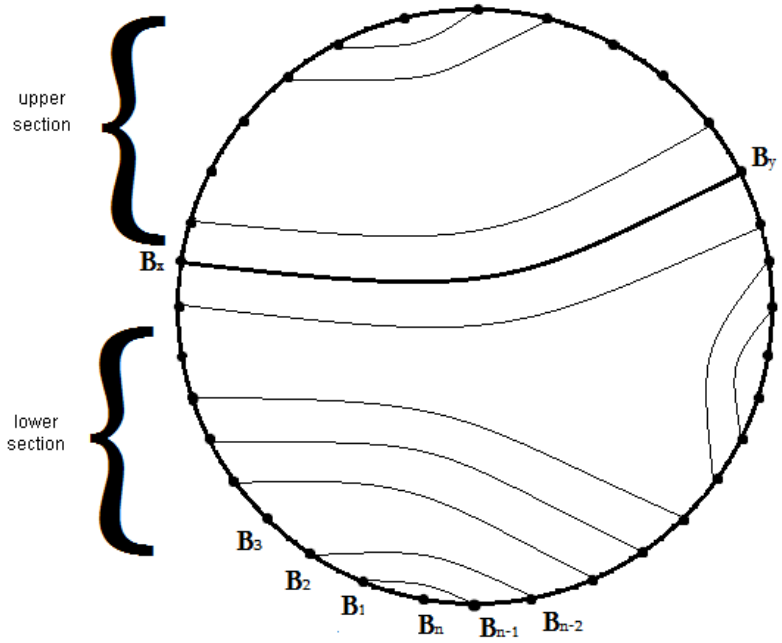

Fig 1.12. Scheme of RNA secondary structure without loops.

pairs to the energy of the secondary structure of RNA. Suppose we have a sequence of nucleotides from B_1 to B_n, and it is located on the circle (Fig. 1.12). Let's assume that B_x and B_y form a pair. Our goal is to find out whether B_x and B_y form a pair in the secondary structure that we are looking for. The arc B_xB_y divides the circle into two parts: the upper and lower sections. The exclusion of pseudoknots means that if two nucleotides form a pair, then both must be either in upper or lower section. Thus, nucleotides from different sections cannot form a base pair. So, energy of the secondary structure will be determined by the energies of the upper and lower sections and an impact of the local pair B_xB_y.

If we have real biological sequences, it is necessary to consider all the possible interactions. At the same time obtaining the recurrence relations will be more complex.

Consider the sequence, which consists only of nitrogenous bases A and U. In receipt of it, we assume that with probability P falls A, and with probability $(1-P)$ – U. Lets calculate its partition function, where the sum is taken over all possible structures, except pseudoknots. For that, we distinguish a region (i, j). Suppose that $j+1$ forms pair with k. In this case we will have two subsequences: from i to $k-1$ and from $k+1$ to j. Without pseudoknots, the partition function of any subchain of ssRNA molecule calculates recursively [31, 32] as

$$Z_{i,j} = Z_{i,j-1} + \sum_{k=1}^{j-1} Z_{i,k-1} q_{ij} Z_{k+1,j-1}, \tag{1.5}$$

where $Z_{i,j}$ is the partition function of the subchain between nucleotides i and j, $q_{ij} = \exp(-\beta\epsilon_{ij})$ being the statistical weight of the base pair formation between nucleotides i and j.

Fig. 1.13 shows secondary structure of RNA consisting of $N = 150$ nucleotides obtained with means of relation 1.5).

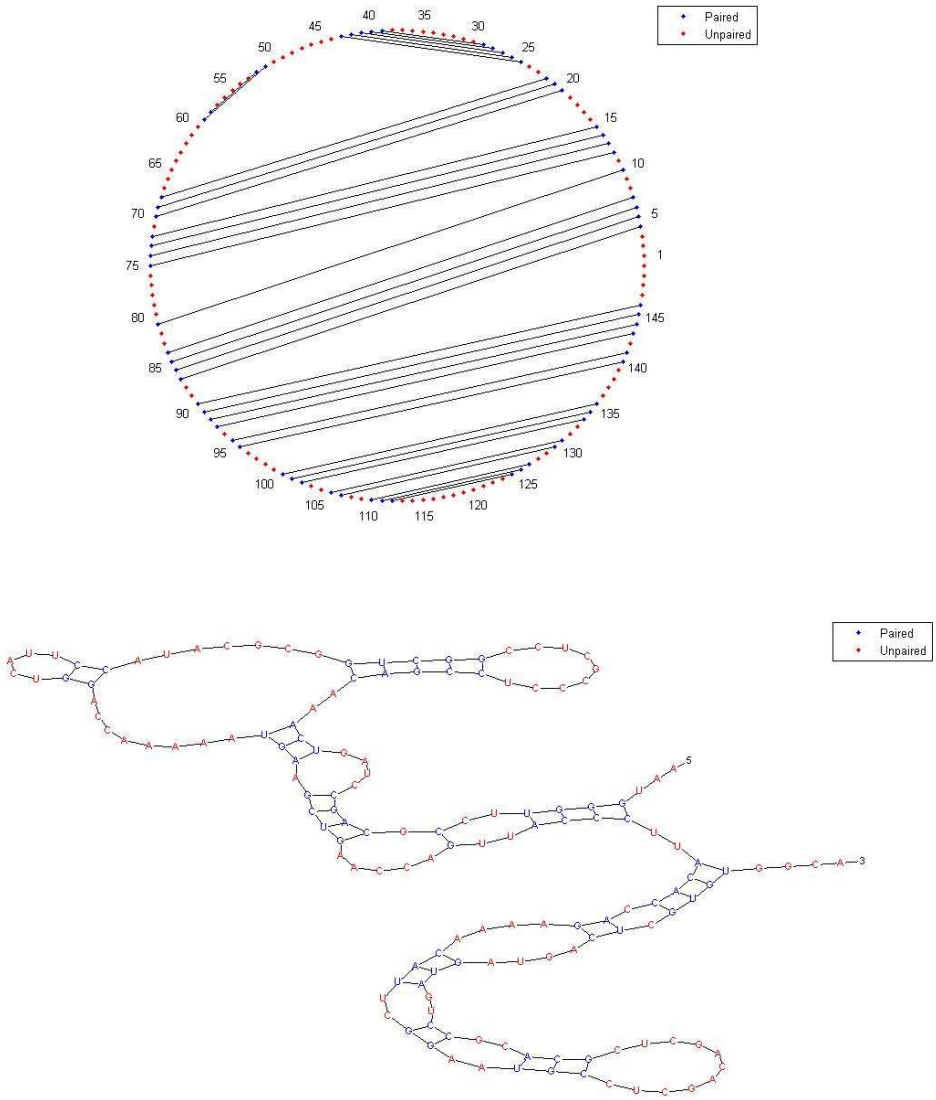

Fig 1.13. The schematic picture of the secondary structure of RNA for sequence that consist of 150 nucleotides.

1.4. The existing theories of secondary structural transitions in DNA.

1.4.1. Zimm-Bragg model

This model [33] is the first consistent and most studied statistical theory of helix-coil transitions. It is based on the one-dimensional Ising model. Let us

suppose that the number of amino acids in the chain is N. In the frameworks of this model it is assumed that the state of the repeating unit described by the state of the oxygen atom in the carboxyl group. If that atom forms a hydrogen bond between molecules, we will denote that state by number 1. Other states will be denoted by number 0. As a result, we will have the sequence of ones and zeros for each configuration. The parameter s is determined by the change of the free energy when the length of the helix increases by one monomer.

$$s = \exp(-\frac{\Delta F}{RT}), \tag{1.6}$$

where $\Delta F = F_h - F_c$. If a monomer, which follows three or more free repeating units, forms a hydrogen bond, free energy increases. The cooperativity parameter σ is associated with the increase of free energy:

$$\sigma = \exp(-\frac{F_s}{RT}), \tag{1.7}$$

where F_s is the additional free energy. When $N \gg 1$ partition function reads

$$Z = TrP^N, \tag{1.8}$$

where N is number of repeating units, P is the matrix of statistical weights. In this case

$$P = \begin{pmatrix} 1 & \sigma s \\ 1 & s \end{pmatrix}. \tag{1.9}$$

As a result, the secular equation will have the following form

$$(1-\lambda)(s-\lambda) = \sigma s. \tag{1.10}$$

This is simple equation and we can obtain exact solutions. It is shown [33] that the helix-coil transition is in the following interval

$$1 - \sqrt{\sigma} \ll s \ll 1 + \sqrt{\sigma}. \tag{1.11}$$

Considering that $\sigma \ll 1$, we will have that

$$\Delta T = 2\sqrt{\sigma}\frac{KT_m^2}{\Delta H}. \tag{1.12}$$

Thus, thermodynamic analysis of helix-coil transitions becomes possible with means of Zimm-Bragg model.

1.4.2. Loop entropy in Poland-Scheraga model

This model helps to describe the existence of loops in the DNA and it gives us reasonable results [34]. There are two main interactions in DNA: hydrogen bonding and stacking. The hydrogen bonds are formed between two bases that are in the different chains. The stacking interactions occur between neighboring nitrogenous bases. Let suppose that the statistical weight of hydrogen bonds is t and the statistical weight of stacking interactions is τ. So, if we have the ordered sequence that consists of j backbone units, the statistical weight of sequence will be written as

$$v_j = t(t\tau)^j \equiv \sigma w^j, \tag{1.13}$$

where $\sigma \equiv t$ and $w \equiv t\tau$. The sequence generating function [35, 36] reads

$$V(x) = \frac{\sigma w}{(x-w)}. \tag{1.14}$$

We will take $\sigma = 1$. In this case we can ignore the inhomogeneities in t and τ depending on the sequence. As a result

$$V(x) = \frac{w}{(x-w)}. \tag{1.15}$$

If we take

$$U(x) = \frac{1}{V(x)}, \tag{1.16}$$

it can be obtained as

$$U(x) = \frac{x}{w} - 1. \tag{1.17}$$

It is obtained in [37] that if the loop consists of N bases, the entropy of that loop will have the following form

$$S(N) = R(N\ln\Omega - [A + \frac{3}{2}\ln N\ ]). \tag{1.18}$$

The constant A cannot be exactly found. The term $RN\ln\Omega$ is the conformational entropy of the free chain.

Let us suppose that the chain is placed on the two-dimensional lattice (square lattice). If there is a loop in the chain, it means that the ends of that loop must match. As a result the number of moves to the right has to be equal to the number of moves to the left, and the number of moves to the up must be equal to the number of moves to the down. So, if the number of bases in the chain is N, the number of moves in the left-right directions is $\frac{N}{2}$ and it is equal to the number of moves in the up-down directions. So, the number of loop conformations is

$$Q = \frac{\left[\left(\frac{N}{2}\right)!\right]^2}{\left[\left(\frac{N}{4}\right)!\right]^4} \tag{1.19}$$

If we use Stirling's approximation

$$n! = e^{-n}n^n(2\pi n)^{\frac{1}{2}}, \tag{1.20}$$

we will obtain

$$Q = N\ln 2 - [\ln\left(\frac{\pi}{4}\right) + \ln N]. \tag{1.21}$$

For the three-dimensional case (cubic lattice) we will have that the numbers of moves in the $\pm x$, $\pm y$ and $\pm z$ directions is equal to $\frac{N}{3}$, when N is large. As a result, the number of loop conformations reads

$$Q = \frac{\left[\left(\frac{N}{3}\right)!\right]^3}{\left[\left(\frac{N}{6}\right)!\right]^6}. \tag{1.22}$$

Using Stirling's approximation $\ln n! = n\ln n - n$, we will get

$$\ln Q = N\ln 2 - [\ln\left(\frac{\pi}{6}\right)^{\frac{3}{2}} + \frac{3}{2}\ln N\ ]. \tag{1.23}$$

It is obvious that Eqs. (1.18) and (1.23) are similar. So, we can write that

$$\ln Q = Na' - (b' + c\ln N\ ), \tag{1.24}$$

where $c = 1$ for the square lattice and $c = \frac{3}{2}$ for the cubic lattice. As a result we can say that the entropy of the loop will be obtained through the following

equation

$$S_{loop}(N) = R[Na - (b + c\ln N)] \tag{1.25}$$

where $c = d/2$, if we ignore the excluded-volume interactions. The quantity d is the dimensionality of the space. The statistical weight of a free chain is

$$u_N = (e^a)^N, \tag{1.26}$$

and so,

$$u_N = e^{\frac{\left[S_{loop}(N)\right]}{R}} = (e^a)^N e^{-b} N^{-c}. \tag{1.27}$$

For the nucleic acids $N = 2(i+1) \approx 2i$. When $d = 3$, we have

$$u_i = (constant) u^i i^{-3/2}. \tag{1.28}$$

It is shown in the article [38] that if we consider the long-range contacts and use the series expansions, we will obtain that the quantity c has the following values

$$c \sim 1,75, \tag{1.29}$$

when dimensionality if the space is equal to three and

$$c \sim 1,46 \tag{1.30}$$

for 2D.

The model suggested by Mukamel is Poland-Scheraga type. In the framework of this model authors considered the effects of excluded-volume interactions. Although they considered those interactions approximately, obtained results allow understanding of dependence between the unbinding mechanism and the nature of the transition.

According to this model, the monomers in DNA can be found in two states: bounded and unbounded. So, the chain is represented as a sequence of these states. The binding energy is the same for all monomers. The statistical weight of a bounded pair is

$$\omega = \exp\left(-\frac{E_0}{T}\right), \tag{1.31}$$

where $E_0 < 0$ is binding energy and T is the temperature. If a segment of the chain consists of k bounded units, the statistical weight of that sequence is given by

$$\omega^k = \exp\left(-\frac{kE_0}{T}\right). \tag{1.32}$$

The statistical weight of the unbounded chain of length k will be determined by the change of entropy. For large k it has the form $\frac{As^k}{k^c}$, where s is non-universal constant. The exponent c describes properties of a loop. Authors consider case where $A = 1$. The grand canonical partition function will be determined by

$$Z = \sum_{M=0}^{\infty} G(M) z^M = \frac{V_0(z)\, U_L(z)}{1 - U(z)\, V(z)} \tag{1.33}$$

where $G(M)$ is the canonical partition function of the chain with length M, z is the fugacity,

$$U(z) = \sum_{k=1}^{\infty} \frac{s^k}{k^c} z^k, \tag{1.34}$$

$$V(z) = \sum_{k=1}^{\infty} \omega^k z^k, \tag{1.35}$$

$V_0(z) = 1 + V(z)$ and $U_L(z) = 1 + U(z)$. The quantities $V_0(z)$ and $U_L(z)$ can be found for boundaries. The average chain length can be obtained from partition function as

$$< L >= \frac{\partial \ln Z}{\partial \ln z}. \tag{1.36}$$

When $< L > \to \infty$, the order parameter θ will be function of temperature. The average number of bounded pairs will be determined by

$$< m >= \frac{\partial \ln Z}{\partial \ln \omega}. \tag{1.37}$$

So,

$$\theta = \lim_{L\to\infty} \frac{< m >}{< L >} = \frac{\partial \ln z^*}{\partial \ln \omega}, \tag{1.38}$$

where z^* is the fugacity when $< L > \to \infty$. Using

$$V(z) = \frac{\omega z}{(1 - \omega z)}, \tag{1.39}$$

we will obtain that

$$U\left(z^{*}\right)=\frac{1}{\left(\omega z^{*}\right)}-1. \tag{1.40}$$

The nature of transition will be determined through the dependence of z^{*} on ω. It is shown in [39, 40] that there are 3 cases:

1. when $c \leq 1$, there is no phase transition.

2. when $1<c \leq 2$, transition is continuous.

3. when $c>2$, we have a first order transition.

The exponent is $c=d\nu$, where d is the dimension of space. If the walks are random and ideal, $c=\frac{d}{2}$. So, when $d \leq 2$, there is no transition, when $2<d \leq 4$, the transition is continuous and the transition is first order, when $d \geq 4$.

1.4.3. Peyrard-Bishop model

The transfer integral method was used for the analysis of Peyrard-Bishop model [41–45]. In this model authors used the fact that there exists an analogy between the study of the conformational properties through statistical physics and the diffusivity equation. The DNA denaturation problem was modeled as a particle in the Morse potential, which describes the hydrogen bonding. It was introduced a pair of variables for every repeating unit. That pair describes the deviation of chain segment in the frameworks of every repeating unit in the direction, which is parallel to the axis of DNA helix. Certain deviation was considered. If the value of deviation is larger, the hydrogen bonds are destroyed. Also, the harmonic pairing, which simulates the stacking between neighboring repeating units, was studied.

The Hamiltonian of this model will have the following form:

$$H=\sum_{n} \frac{p_{n}^{2}}{2m}+W\left(y_{n}, y_{n-1}\right)+V(y_{n}), \tag{1.41}$$

where $p_{n}=m\frac{dy_{n}}{dt}$, m is the reduced mass of bases. The potential $V(y_{n})$ describes the interactions between two repeating units. In other words, it describes

the hydrogen bonding. The potential $W(y_n, y_{n-1})$ describes the interactions between two repeating units along the DNA molecule (stacking interactions). It is convenient to use the Morse potential, because it is standard for description of the chemical bonds and it has appropriate form. We have strong repulsion at short distances, the minimum in balance and it becomes flat at large distances. Through the Hamiltonian we can find the dependence of average value of deviation from equilibrium on value of the constant pairing. The average deviation characterizes the degree of denaturation.

1.4.4. The preliminary model of helix-coil transition in double-stranded DNA

Consider a double-stranded homopolynucleotide with complementary binding in the region of helix-coil transition. This is possible if we consider a real DNA with an approximation that energies of AT and GC pairs are equal.To address the order of the phase transition in double-strand DNA we need to consider homopolymeric DNA with complementary base pairing. Experimentally, it can be created using the stability inversion approach, proposed in [46]. In presence of the appropriate concentrations of alkylammonium compounds, stability of GC and AT pairs can be equalized or even inversed. In case of the same stability of the GC and AT pairs double-strand DNA behaves as a homopolynucleotide with symmetric loops. In other case, we consider the random heteropolymer which consists of AT complementary base pairs only. One chain is the random sequence of A and T nucleotides and another chain is complementary to the first one. We can say the same for the GC pairs. In this case, the energy of the hydrogen bond formation will be the constant along the chain. We can assume that the inter-chain hydrogen bonds are formed only between the bases having the same number. So all loops are symmetric. The macromolecule is schematically presented in Fig. 2.1. We study the formation of hydrogen bonds between complementary repeated units of two chains. For simplicity, let us

assume that the first repeated units are bound. As it was introduced in [47], the Hamiltonian for the macromolecule is

$$-\beta H = J\sum_{i=1}^{N}\delta_1^{(i)}, \tag{1.42}$$

where $\beta = T^{-1}$, $J = \frac{U}{T}$, U is the energy of the hydrogen bond formation in one complementary pair, $\delta_1^{(i)}$ takes value 1 if a hydrogen bond of the ith complementary pair is formed and 0 if not. Since the first pair is bound, $\delta_1^{(i)}$ is nonzero if two chains form a closed cycle between the first and ith repeated units. The presence of other cycles inside the interval $[1, i]$ is possible. Actually, this model is a Poland-Sheraga (PS) type model [34]. The partition function for Hamiltonian (1.42) is

$$\Lambda = \sum_{\{\vec{\gamma}_i\}}\exp\left(-\beta H\right) = \sum_{\{\vec{\gamma}_i\}}\prod_{i=1}^{N}\left(1+v\delta_1^{(i)}\right) \tag{1.43}$$

where $v = e^J - 1$. Let $\vec{\gamma}_i$ be a set of all possible values $1, 2, ..., Q$ which enumerate conformations of the chain. The partition function can by developed as the series in v. By using the relationship $\delta_1^{(k)}\delta_1^{(m)} = \delta_1^{(k)}\delta_{k+1}^{(m-k)}$ we can write the term corresponding to v^f as

$$v^f\delta_1^{(k_1)}\delta_{k_1+1}^{(k_2-k_1)}\delta_{k_2+1}^{(k_3-k_2)}\ldots\delta_{k_{f-1}+1}^{(k_f-k_{f-1})} \tag{1.44}$$

Imposing cyclic conditions and defining m_i as $k_i - k_{i-1}$ we obtain

$$\Lambda = Q^N\sum_f v^f\sum_{m_1}\varphi\left(m_1\right)\ldots\sum_f\varphi\left(m_f\right) \tag{1.45}$$

where

$$\varphi\left(m\right) = Q^{-m}\sum_{\gamma_1}\sum_{\gamma_2}\ldots\sum_{\gamma_m}\delta_1^{(m)}. \tag{1.46}$$

According to (1.46), $\varphi(m)$ is the ratio of a number of states corresponding to the formation of a loop of length m and all states of the chain of length m. So the function $\varphi(m)$ may be interpreted as the probability of the loop formation

of the length m. Using the condition $\sum_{i=1}^{f} m_i = N$ and multiplying (1.45) by factor $\delta\left(\sum_{k=1}^{f} m_k - N\right)$, we obtain

$$\Lambda = \frac{1}{2\pi i}\oint z^{-N-1}\sum_{f=1}^{N}\left(v\sum_{m=1}^{\infty}\varphi\left(m\right)z^{m}\right)^{f}dz \tag{1.47}$$

In [47] the function $\varphi(m)$ was chosen approximately as

$$\varphi(m) = \begin{cases} Q^{-m}, m \leq \Delta \\ Q^{-\Delta}, m > \Delta \end{cases} \tag{1.48}$$

Using the saddle-point approach, one can show that the characteristic equation for the free energy in the thermodynamic limit is the same as in the GMPC model, which is a Potts like one-dimensional model. This representation of $\varphi(m)$ is empiric and ignores the loop formation with length less than Δ which characterizes the single-chain rigidity.

In the present study, we generalize the model to the case of loops of an arbitrary length. To this end, the problem of loop formation will be represented in terms of random walks.

Chapter 2

The random walk model of helix-coil transitions in double-stranded homopolynucleotides

2.1. Description of the basic random walk model (Model A)

The structure of a homopolynucleotide is considered as a sequence of alternating helical and coil regions. Helical regions, which are essentially one-dimensional, are stabilized by hydrogen bonds and stacking interactions. Coil regions are apparently d-dimensional, where d is the topological dimension of the space where the DNA chain is embedded. We will focus on the three-dimensional case, $d = 3$.

The main concept of the random walk description is quite simple. We consider the molecule of DNA as two random chains which are initiated from the same point. As it was mentioned above, the complementary pairs of nitrogenous bases are able to create hydrogen bonds, and each binding corresponds to the intersection of two random chains. We label the pairs of complementary polymer units (which can be potentially bonded) by integers $1, 2, ..., N$, and construct N planes perpendicular to the polymer axis in such a way that both units of the i-th pair lie on the i-th plane with the coordinates x_i and y_i. If the complementary units are bounded, they are represented by a single point ($x_i = y_i$) on the corresponding plane. The projection of all N planes onto a single plane gives the collection of points x_i, y_i, $i = 1, 2, ..., N$ which can be considered as the position of $2D$ random walks at the moments of discrete times $i = 1, 2, ..., N$ (Fig. 2.1). This construction admits arbitrary conformations of polymer chains with a single but important exclusion: all planes $1, 2, ..., N$ are crossed by polymer chains sequentially from the first to the last one and any

return from the i-th to the $(i-1)$-th plane is forbidden.

In the absence of meanders this guarantees the exclusion of three-dimensional knots and additional base pairs inside the loops. Thus, our approach [48, 49] describes three-dimensional loop statistics more adequately than the traditional one [34].

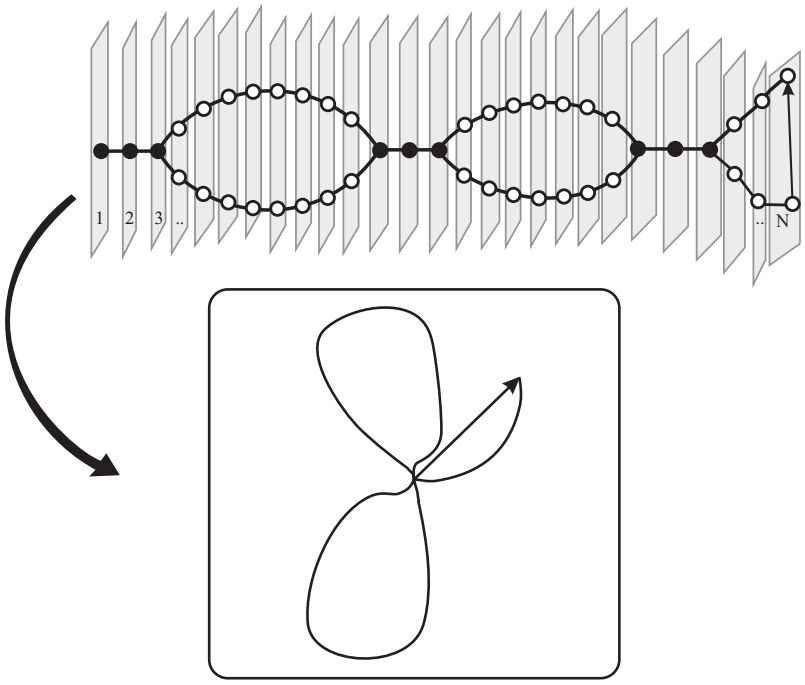

Fig 2.1. Scheme of the model

For the sake of convenience, we consider a simple random walk on the quadratic lattice. The $2D$ simple random walk jumps one lattice left, right, up or down at each discrete time step. Later on we will extend this model to the case when a stay at the origin during several stops is allowed.

To write the partition function (1.43) in terms of a random walk, we refer to the well-known generating functions [50]. The generating function for the first return is

$$F(z) = \sum_{m=1}^{\infty} f_m z^m, \tag{2.1}$$

where f_m is the probability of the first return at the m-th step. The generating

function for any return is

$$P(z) = \sum_{m=1}^{\infty} p_m z^m, \tag{2.2}$$

where p_m is the probability of any return at the m-th step. $P(z)$ can also be represented as an integral which, in the case of two-dimensional random walk on the quadratic lattice, is [50]

$$P(z) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{d\varphi d\psi}{1 - \frac{z}{2}(\cos\varphi + \cos\psi)} = \frac{2}{\pi} K(z^2), \tag{2.3}$$

where $K(z)$ is the complete elliptic integral of the first kind. Using the known relation between $F(z)$ and $P(z)$

$$F(z) = 1 - \frac{1}{P(z)}, \tag{2.4}$$

we obtain the analytical expression for $F(z)$. Taking into account that $\varphi(m) = p_m$ is the return probability on the m-th step and N is the whole number of steps, we can rewrite (1.47) in terms of the generating function for $P(z)$ as

$$\Lambda = \frac{1}{2\pi i} \oint z^{-N-1} \sum_{f=1}^{N} v^f (P(z) - 1)^f dz. \tag{2.5}$$

Now let us consider the partition function in terms of the generating function of the first return $F(z)$. Each time the particle returns to the origin, we add a weight $k = e^{\frac{-U}{T}}$ to the random walk which is the statistical weight of the base pair formation. The probability of the final return of the particle to the origin after N steps, i.e., the partition function of the double chain with connected first and last monomers reads:

$$\Lambda = \sum_{j=0}^{\infty} k^j F(z)^j|_{z^N} = \frac{1}{2\pi i} \oint_{C_0} \frac{1}{1 - kF(z)} \frac{dz}{z^{N+1}}, \tag{2.6}$$

where the contour C_0 encloses the origin in a clockwise manner. We discuss two cases for the value of hydrogen bond energy U. When $U < 0$, $k > 1$ we have attraction of a particle at the origin.

2.1.1. Calculation of partition function ($k > 1$)

To estimate the integral Λ in (2.6) for the case $k > 1$ around a contour C_0 enclosing the origin, we consider another one around C_1, consisting of the circular part with the radius $1 + \delta$ and indentation around branch points at $z = \pm 1$ (Fig. 2.2) [51, 52]. Further, we will choose a positive δ small enough to use an asymptotic expression of (2.4) near the points ± 1 on the indentation part of C_1.

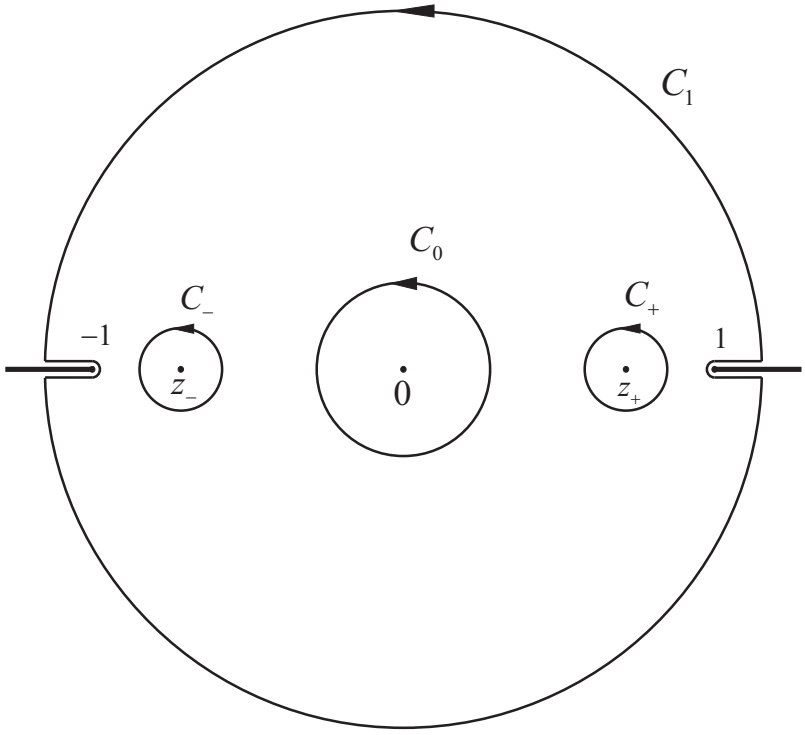

Fig 2.2. The choice of the contour in the complex-z for the case of attractive origin

Notice that there are two simple poles z_+ and z_- inside the contour C_1, which can be found by solving the equation $F(z) = \frac{1}{k}$. To compute the contour integral, we subtract integrals around C_+ and C_- enclosing the poles z_+ and z_- from the integral around C_1

$$\Lambda = \oint_{C_0} = \oint_{C_1} - \oint_{C_+} - \oint_{C_-} . \tag{2.7}$$

For the last two integrals we obtain

$$
\begin{aligned}
-\oint\limits_{C_+} &= \frac{1}{z_+^{N+1}} \frac{1}{kF'(z_+)} \\
-\oint\limits_{C_-} &= \frac{1}{z_-^{N+1}} \frac{1}{kF'(z_-)}.
\end{aligned} \tag{2.8}
$$

As the function $F[z]$ is even on the interval $(-1, 1)$, we get $F'(z_+) = -F'(z_-)$. Using the fact that N is even, it can be shown that

$$
\oint\limits_{C_+} + \oint\limits_{C_-} = 2 \oint\limits_{C_+}. \tag{2.9}
$$

To estimate $\oint\limits_{C_1}$, we notice that the integral around the circular part of the contour is proportional to $\frac{1}{(1+\delta)^{-N}}$, which is negligible compared to $\oint\limits_{C_+}$ and $\oint\limits_{C_-}$ for large N as $z_+ = |z_-| < 1$. The integral around indentation of the points 1 and -1 can also be ignored because it is bounded in magnitude by a number independent of N. We will evaluate the last one more explicitly in 2.1.2.

For the asymptotic expression of the integral Λ for large N we get

$$
\Lambda = \frac{2}{z_+^{N+1} kF'(z_+)}, \tag{2.10}
$$

where z_+ is the positive pole of integrand in (2.6) defined from the transcendental equation

$$
F(z) = \frac{1}{k}. \tag{2.11}
$$

2.1.2. The asymptotic analysis of the partition function ($k < 1$)

In this section, we give asymptotic analysis of the integral Λ in (2.6) for the case $k < 1$. This case is when $U > 0$, and it corresponds to a repulsive origin. In this case we have no poles z_+ and z_-; therefore, we must estimate the value of the integrals Λ_{MP} and $\Lambda_{M'P'}$ on the indentation parts MP and $M'P'$ of the points ± 1 of the contour C_1 (Fig. 2.3). As the number N is even, we have $\Lambda_{M'P'} = \Lambda_{MP}$. For the $2D$ random walks, the generating function $F(z)$

is expressed by the complete elliptic integral of the first kind $K(z)$ and has an asymptotic limit near point 1

$$F(z) = 1 - \frac{1}{\frac{2}{\pi}K\left(z^2\right)} \xrightarrow[z\to 1]{} 1 - \frac{1}{-\frac{1}{\pi}\log(1-z)}. \tag{2.12}$$

Substituting $F(z)$ in the formula for Λ we obtain

$$\Lambda = \frac{1}{2\pi i}\int\limits_{M}^{P} \frac{1}{z^{N+1}} \frac{dz}{1 - k\left(1 - \frac{1}{-\frac{1}{\pi}\log(1-z)}\right)} \tag{2.13}$$

Let us divide integration (2.13) into two parts MR and RP, where R is a point of intersection between contour C_1 and the real axis. Considering the branches of the logarithmic function on MR and RP separately we get

$$\Lambda = \frac{1}{2\pi i}\int\limits_{M}^{R} \frac{1}{z^{N+1}} \frac{dz}{1 - k\left(1 - \frac{1}{-\frac{1}{\pi}(\log(z-1)+i\pi)}\right)} + \frac{1}{2\pi i}\int\limits_{R}^{P} \frac{1}{z^{N+1}} \frac{dz}{1 - k\left(1 - \frac{1}{-\frac{1}{\pi}\log(z-1)-i\pi}\right)} \tag{2.14}$$

After joining these two integrals and turning to a new variable $z' = z - 1$ we obtain

$$\Lambda = \int\limits_{0}^{\delta} \frac{1}{(z'+1)^{N+1}} \frac{k/\pi\, dz'}{\left[k - \frac{1-k}{\pi}\log(z')\right]^2 + (1-k)^2}. \tag{2.15}$$

For a fixed small δ, formula (2.15) can be rewritten as

$$\Lambda = \frac{k\pi}{(1-k)^2}\int\limits_{0}^{\delta} \frac{1}{(z'+1)^{N+1}} \frac{dz'}{\log^2(z')}. \tag{2.16}$$

To proceed with the asymptotic analysis for large N, we denote by $y = Nz$ and get for the integral part of (2.16)

$$\Lambda_1 = \frac{1}{N}\int\limits_{0}^{N\delta} \frac{1}{(\frac{y}{N}+1)^{N+1}} \frac{dy}{(\log(y) - \log(N))^2} \sim \frac{1}{N\log^2(N)}\int\limits_{0}^{N\delta} \frac{e^{-y}dy}{\left(1 - \frac{\log(y)}{\log(N)}\right)^2}. \tag{2.17}$$

Further, we divide the integral expression in (2.17) into two parts, integrating

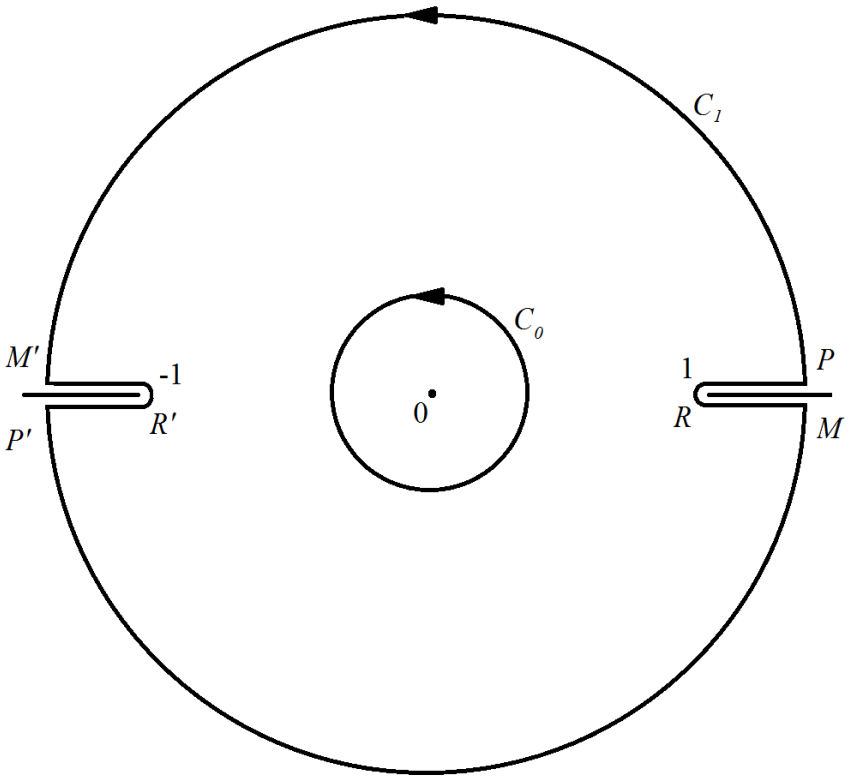

Fig 2.3. The choice of the contour in the complex-z for the case of repulsive origin

from 0 to $1/N$ and from $1/N$ to $N\delta$. For the first part, we have

$$\frac{1}{N\log^2(N)}\int_0^{1/N}\frac{e^{-y}dy}{\left(1-\frac{\log(y)}{\log(N)}\right)^2}\lesssim\frac{1}{N}\int_0^{1/N}\frac{dy}{\log^2(y)}\sim\frac{1}{N^2\log^2(N)}. \tag{2.18}$$

For the second part of integration, we derive

$$\frac{1}{N\log^2(N)}\int_{1/N}^{N\delta}\frac{e^{-y}dy}{\left(1-\frac{\log(y)}{\log(N)}\right)^2}=\frac{1}{N\log^2(N)}(1+O(\frac{1}{\log(N)})). \tag{2.19}$$

We see that (2.18) is negligible compared to (2.19) and finally we obtain an asymptotic expression of the partition function

$$\Lambda=\frac{k\pi}{(1-k)^2}\frac{1}{N\log^2(N)} \tag{2.20}$$

and the average energy vanishes in the limit $N\to\infty$.

For the average energy per step and helicity we obtain

$$\begin{aligned} E &= -\frac{(1+k)\log k}{(1-k)}\frac{1}{N} \\ \theta &= \frac{1+k}{1-k}\frac{1}{N} \end{aligned} \tag{2.21}$$

which tend to 0 when N goes to infinity.

2.2. The thermodynamic characteristics of the Model A

The internal energy per step, in units T, is given by

$$\overline{E} = -\frac{k\log k}{N}\frac{\partial \log \Lambda}{\partial k} = -\frac{k\log k}{2\pi i N\Lambda}\oint \frac{F(z)}{(1-kF(z))^2}\frac{dz}{z^{N+1}} \tag{2.22}$$

The calculations similar to those for derivation of (2.10) from (2.6) lead in the limit of large N to:

$$\overline{E} = -\frac{\log k}{kz_+F'(z_+)}. \tag{2.23}$$

The helicity degree θ is defined as an average fraction of hydrogen bonds in the biopolymer, i.e. is the ratio of the average and maximal numbers of the hydrogen bonds. For the simple random walk model, the maximal number of returns to the origin and, therefore, the maximal number of bonds is $\frac{N}{2}$. Using (2.23) we can write the helicity degree as

$$\theta = \frac{2}{kz_+F'(z_+)}. \tag{2.24}$$

The thermal dependence of the helicity degree is shown in Fig. 2.4. The slow decay of θ demonstrates the gradual helix-coil transition.

2.3. Modified model with account of entropy of base pair formation (Model B)

Each nucleotide is a group of atoms described by internal degrees of freedom, the dihedral angles. The base pair formation gains the energy but results in the

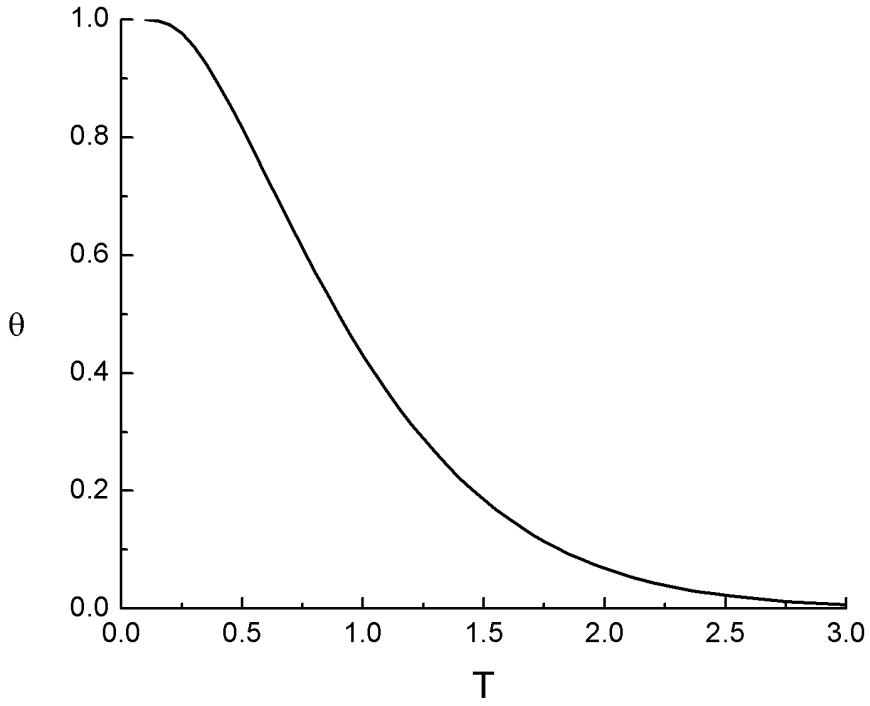

Fig 2.4. Helicity degree (θ) of the Model A in dependence on temperature (T)

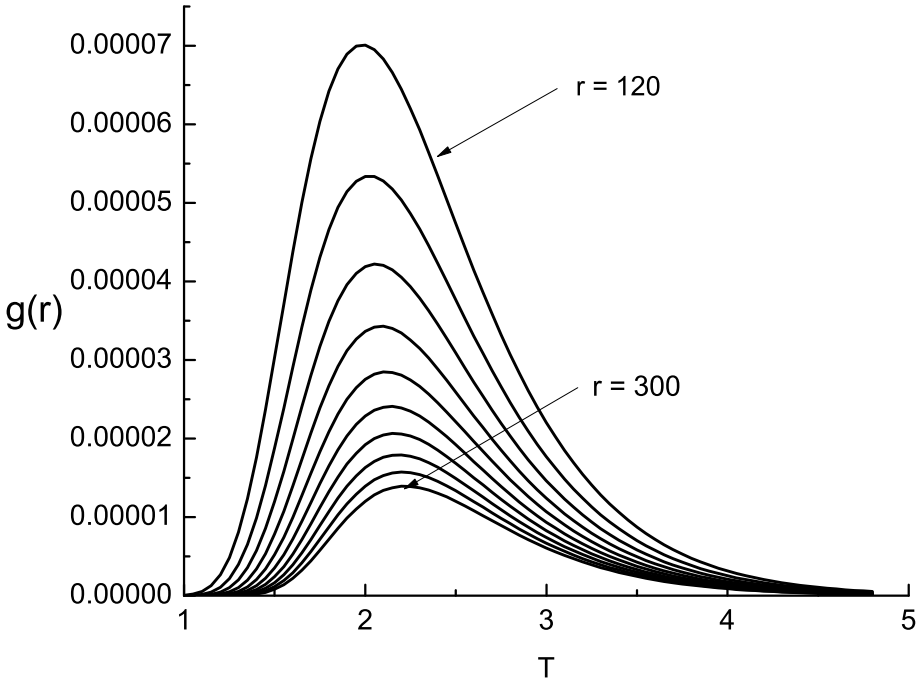

Fig 2.5. Correlation function of Model A

entropy loss. To address the issue of the internal structure of nucleotides, we modify the statistical weight of single base pair, so that, it is $w = \exp(-\frac{\Delta U - T\Delta S}{T})$, where ΔU is the energy and ΔS is the entropy of the base pair formation.

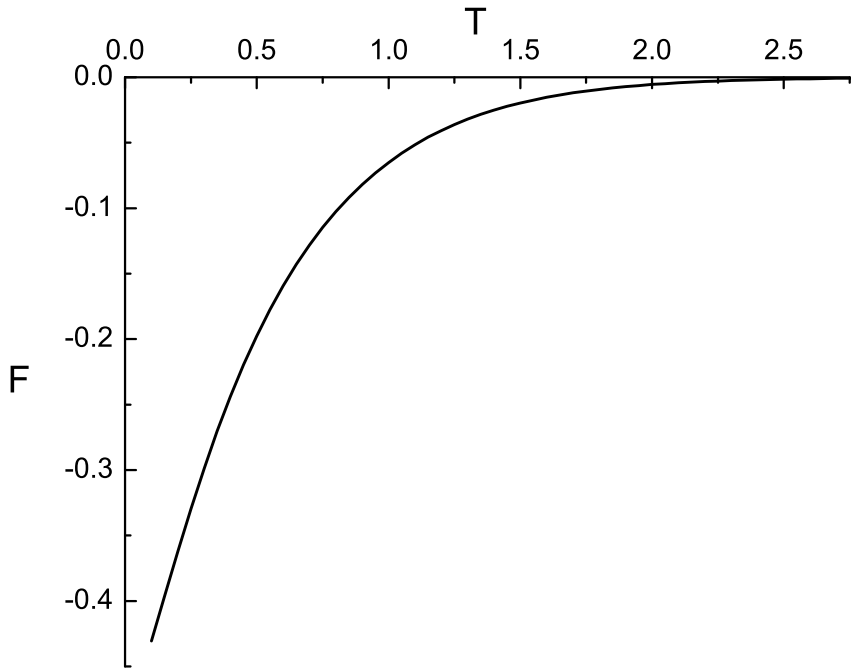

Fig 2.6. Free energy of Model A

Each nucleotide is a group of atoms having internal degrees of freedom, the dihedral angles. The base pair formation gains the energy $\Delta U < 0$, but results in the entropy loss $\Delta S < 0$, because the formation of each base pair requires appropriate relative orientation of the nitrogen bases. Thus, each time the particle returns to the origin, we add the statistical weight w.

The final return of a particle to the origin after N steps corresponds to the partition function of the double chain with the connected first and last monomers:

$$\Lambda_N = \sum_{j=0}^{\infty} w^j F(z)^j|_{z^N} = \frac{1}{2\pi i} \oint\limits_{C_0} \frac{1}{1 - wF(z)} \frac{dz}{z^{N+1}}, \tag{2.25}$$

where the contour C_0 encloses the origin in a clockwise manner.

The temperature behavior of the system is encoded in the singularities of the integrand of the partition function (2.25). Notice that there are two simple poles z_+ and z_- inside the contour C_1 (Fig. 2.7) which can be found by solving the transcendent equation:

$$F(z) = \frac{1}{w}. \tag{2.26}$$

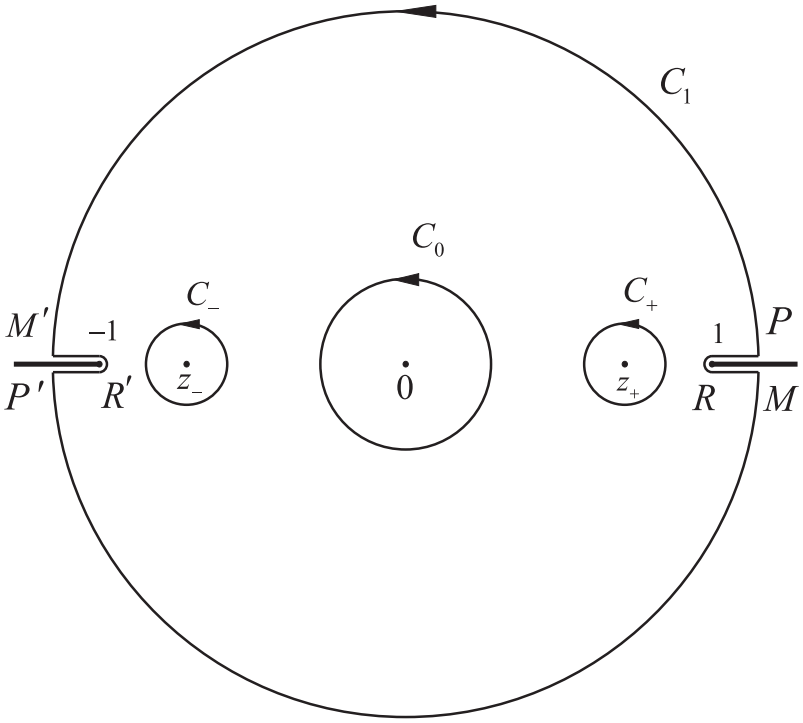

Fig 2.7. The choice of the contour in the complex-z for the case of attractive origin

The critical temperature T_c exists, which is defined by the equation

$$w_c = 1. \tag{2.27}$$

We have to consider two cases. The case $w > 1$ ($T < T_c$) corresponds to the attraction of the walk to the origin. The opposite case $w < 1$ ($T > T_c$) corresponds to the repulsive origin, and we have no poles z_+ and z_- inside the contour C_1. To estimate the integral Λ_N in (2.25) for the case $w > 1$ around the contour C_0 enclosing the origin, we consider another one around C_1, consisting of the circular part with the radius $1+\varepsilon$ and an indentation around the branch points at $z = \pm 1$ (Fig. 2.7) [51, 52]. Further, we will choose a positive ε small enough to use an asymptotic expression of (2.3) and (2.4) near the points ± 1 on the indentation part of C_1.

As number N is even, we have

$$\oint\limits_{C_+} + \oint\limits_{C_-} = 2 \oint\limits_{C_+} = \frac{2}{z_+^{N+1} w F'(z_+)}. \tag{2.28}$$

The contribution of the integral on the indentation parts MRP and $M'R'P'$

of the contour C_1 (Fig. 2.7) is $\delta\Lambda_N$. For large N we get

$$\delta\Lambda_N \simeq \frac{w}{(1-w)^2}\frac{\pi}{N\log^2(N)}. \tag{2.29}$$

The contribution from the rest of the contour C_1 is proportional to $(1+\varepsilon)^{-N}$, which is negligible compared to both values of (2.28) and (2.29). Then, for large but finite N ($T < T_c$) we obtain

$$\Lambda_N = \frac{2}{z_+^{N+1}wF'(z_+)} + \frac{w}{(1-w)^2}\frac{\pi}{N\log^2(N)}. \tag{2.30}$$

In the case $w < 1$ ($T > T_c$), Λ_N vanishes as $\delta\Lambda_N$ for large N.

To estimate the value of the parameter c, responsible for the order of transition, we address the probability f_m of the first return at the m-th step. Asymptotics of f_m can be derived from the probability of the first return to the origin after time t: $Prob(t > m) \sim \pi/\log m$ [53]. Differentiating, we obtain:

$$f_m \sim \frac{\pi}{m\log^2(m)}. \tag{2.31}$$

Of course, one can get the same result using the method similar to the derivation of (2.29) where the contour integration is applied to the function $F(z)$ [54]. Thus, the long loops asymptotics of the proposed model corresponds to $c = 1$ in the expression $\delta S(m) = -c\log m$ mentioned in Introduction.

2.4. The thermodynamic characteristics of the model B

According to the formulation of the model, Eq. (2.25) can be interpreted as a partition function of the chain starting from the first and ending by the L-th base pair. In the limit $N \to \infty$ below the critical temperature T_c, we get from Eq. (2.30)

$$\Lambda_{1,L} = \frac{2}{z_+^{N+1}wF'(z_+)}, \tag{2.32}$$

where $L = \frac{N}{2}$ is the total number of base pairs in the chain. For $T < T_c$, the density of the free energy $\mathfrak{F}$ in the thermodynamic limit is:

$$\mathfrak{F} = T\log z_+. \tag{2.33}$$

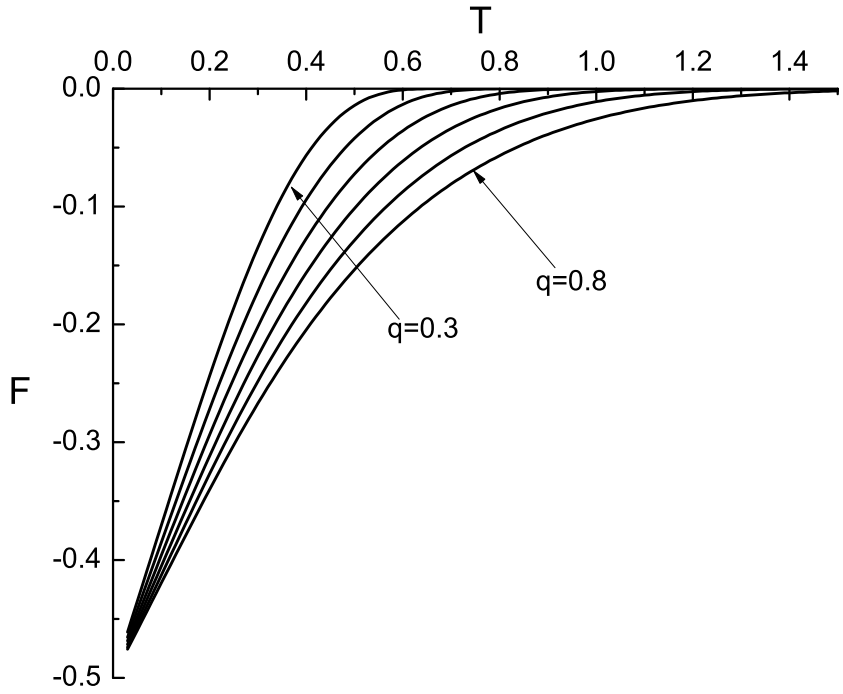

Fig 2.8. Free energy of Model B

To describe the critical behavior of $\mathfrak{F}$ near T_c, one should notice that the pole z_+ tends to 1 when $T \to T_c - 0$. Solving Eq. (2.26) by using the asymptotic expression of $P(z)$ and $F(z)$ near point 1, we find

$$z_+ \to 1 - e^{-\frac{\pi w}{w-1}}, \tag{2.34}$$

where $w = \exp(-\frac{\Delta U - T\Delta S}{T})$. Then, using Eqs. (2.33) and (2.55), we obtain the asymptotics of the free energy density in the case of $T \to T_c - 0$:

$$\mathfrak{F} \simeq T_c \exp\left(-\frac{T_c^2}{|\Delta U|(T_c - T)}\right). \tag{2.35}$$

The main observable quantity of the melting DNA is the helicity degree θ defined as an average fraction of the base pairs N_{bp} in the biopolymer,

$$\theta = \frac{1}{L}\langle N_{bp}\rangle = \frac{w}{L}\frac{\partial \log \Lambda_{1,L}}{\partial w} \tag{2.36}$$

or the degree of denaturation, $\eta = 1 - \theta$, which is defined as an average fraction of the non-bounded base pairs. Using (2.32) we can write the helicity degree for $T < T_c$ as

$$\theta = \frac{2}{w z_+ F'(z_+)}. \tag{2.37}$$

The helicity degree and the density of free energy completely vanish above the critical temperature T_c in the limit $L \to \infty$.

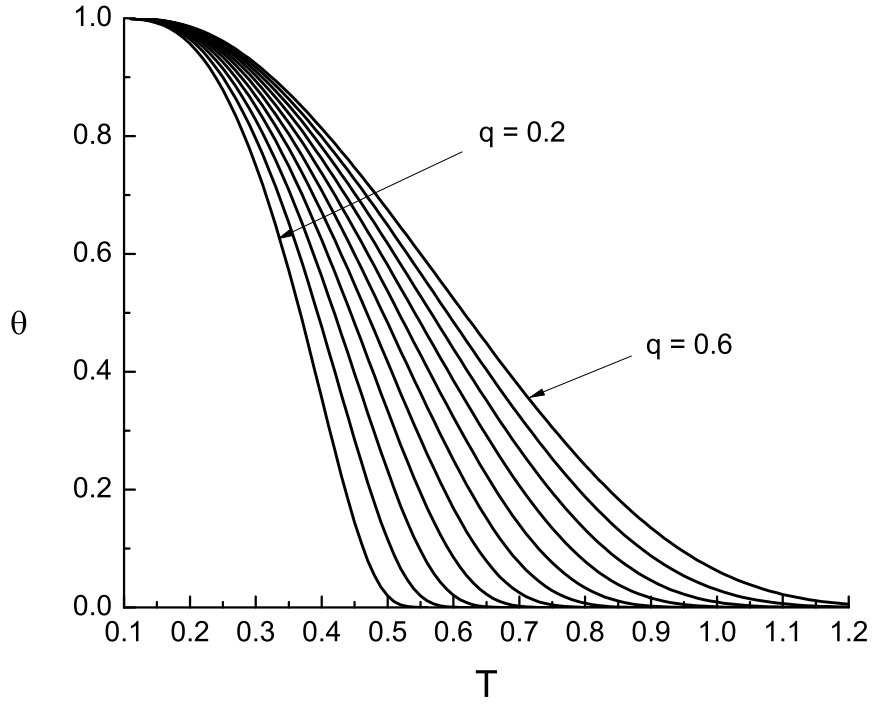

Fig 2.9. Helicity degree of Model B

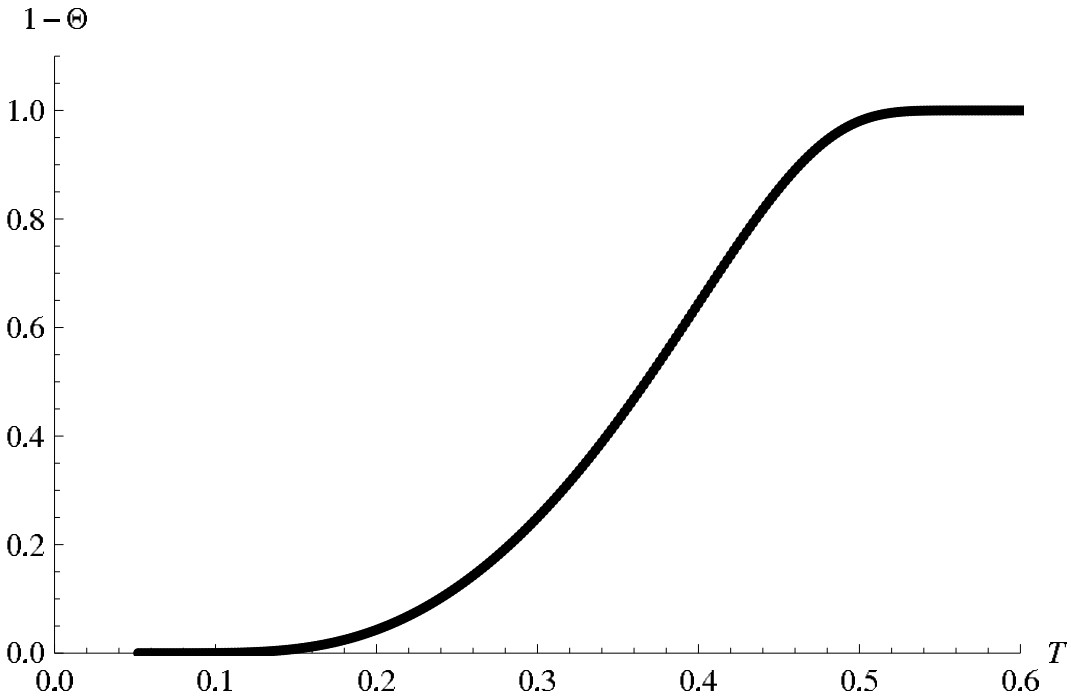

Fig 2.10. Dependence of the degree of denaturation, $\eta = 1 - \theta$ on the temperature.

The thermal dependence of the helicity degree is shown in Fig. 2.10. We conclude that the model describes the complete denaturation transition at the

finite temperature T_c.

To address the fluctuations of the base pair formation, we define the pairwise correlation function as

$$g(i,j) = P(i,j) - P(i)P(j), \tag{2.38}$$

where $P(i,j)$ is the probability of the i-th and j-th base pair formation:

$$P(i,j) = \frac{\Lambda_{1,i}\Lambda_{i,j}\Lambda_{j,L}}{\Lambda_{1,L}}, \tag{2.39}$$

and $P(i)$ is the probability of the i-th base pair formation:

$$P(i) = \frac{\Lambda_{1,i}\Lambda_{i,L}}{\Lambda_{1,L}}. \tag{2.40}$$

By substituting (3.31) and (2.40) in (2.38), the correlation function can be expressed as

$$g(i,j) = \frac{\Lambda_{1,i}\Lambda_{i,j}\Lambda_{j,L}}{\Lambda_{1,L}} - \frac{\Lambda_{1,i}\Lambda_{i,L}}{\Lambda_{1,L}}\frac{\Lambda_{1,j}\Lambda_{j,L}}{\Lambda_{1,L}}. \tag{2.41}$$

In the case of the long DNA, the points i and j are far away from the ends of the chain, but the distance $|i-j|$ is finite. Thus, we need an expression for the Λ_L for large, but finite L ($T < T_c$). Taking into account Eq. (2.30) and Eqs. (2.37, 2.41), we obtain the correlation function for $T < T_c$ in the form

$$g(r) \simeq \frac{\theta w}{(1-w)^2}\frac{\exp(-\frac{r}{\xi})}{r\log^2(r)}, \tag{2.42}$$

where $r = |i-j|$, and the correlation length is

$$\xi = -\frac{1}{\log z_+}. \tag{2.43}$$

The asymptotics of the correlation length ξ at temperatures $T \to T_c - 0$ follows from Eqs. (2.55) and (2.34)

$$\xi \simeq \exp\left(\frac{T_c^2}{|\Delta U|(T_c - T)}\right). \tag{2.44}$$

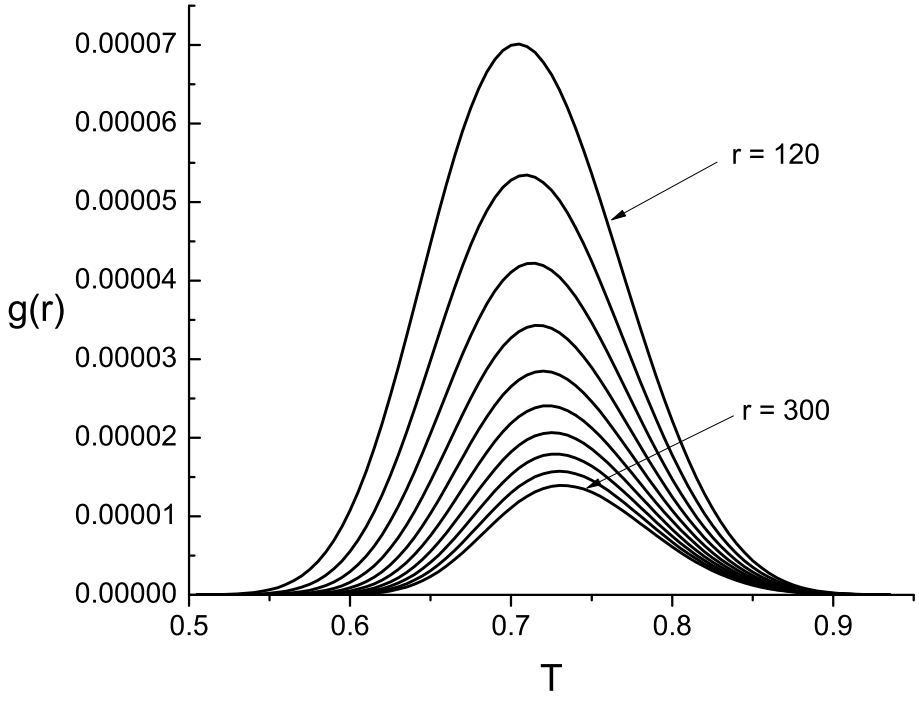

Fig 2.11. Correlation functionof Model B

2.5. Results and discussion of Model B

The main idea of the proposed approach is the mapping of two interacting three-dimensional polymer strands onto the single two-dimensional random walk interacting with the origin [55].

The completely denatured state ($\theta = 0$) with unbounded two DNA strands appears at the finite temperature $T = T_c$ (see Fig. 2.10). To understand the phase behavior of the model, the asymptotics of the density of the free energy $\mathfrak{F}$ near the transition point T_c has been found. The results are given by Eq. (2.35). This kind of temperature behavior can be identified with the infinite order phase transition. That is new for DNA, but considered formerly, for instance in [56–59].

At $T \to T_c - 0$ the tendency to 0 of the order parameter θ can be expressed in terms of the correlation length ξ (2.44):

$$\theta \simeq \frac{T_c^2}{\xi (T_c - T)^2}. \tag{2.45}$$

In the vicinity of T_c, the correlation length diverges at $T \to T_c - 0$. If in the

case of the second order phase transition the length of correlations diverges by the power-law $\sim |T_c - T|^{-x}$, then in our case it diverges qualitatively differently as $\sim \exp(\frac{const}{T_c - T})$ (see Eq. (2.44)). At the same time, the free energy (2.35) is continuous over the temperature, and the phase transition can be interpreted as an infinite order.

When $\xi \to \infty$ $(T \to T_c - 0)$ the correlation function $g(r)$ decays as the power-law

$$g(r) \sim \frac{1}{r \log^2(r)}. \tag{2.46}$$

The pre-factor of the correlation function $g(r)$ in Eq. (2.42) behaves as $\sim \frac{\theta w}{(1-w)^2}$. Since the helicity degree θ tends to zero as $\sim \exp(\frac{const}{T_c - T})$ (see Eq. (2.45)), the value of correlation as a function of temperature vanishes in the same way. Thus, we have unusual thermal behavior at $T \to T_c - 0$, where the correlation length diverges as an exponent but at the same time the value of correlation tends to zero.

The phase transition of the infinite order was obtained before, e.g., in [56, 57] for the one-dimensional classical spin model with long-range interactions and the singled out interaction center on the lattice. A similar result was presented in [59], where the Ising model on the growing network was addressed. There, the order parameter decays as $\sim \exp(-\frac{1}{\sqrt{T_c - T}})$, which is qualitatively similar to our results in (2.45). That and existence of long-range interactions in the system are the common point with our model for the denaturated loops arbitrary lengths.

The phase transition of the infinite order considered in this work takes place in the case of $c = 1$ at the end of melting, where $\theta = 0$. The given scenario is in agreement with the experimental data obtained in [60]. The sharp kink of the melting curve was interpreted in [60] as a phase transition of the second order. However, the number of experimental points is not enough to define the order of transition without fluctuational analysis. At the same time, the comparison of the melting curves presented in Fig. 2.10 and in Fig. 1 of Ref.[60] shows very close similarity. Thus, the phase transition obtained in [60] experimentally

could be of the infinite order but not the second order.

The order of phase transition in the double-stranded DNA is sensitive to the way of taking into account the loop entropy $\delta S(m) = -c \ln m$, where m is the length of the loop. In our approach, no assumptions about the value of c has been made. Considering denaturated loops explicitly in some approximation justified above we obtain $c = 1$. In contrast to the results derived in [40, 61, 62] we got the continuous phase transition of the infinite order. The various phase behavior is explained by the different consideration of the denaturated loops.

2.6. Random walks with stops at the origin (Model C)

In this section, we consider another kind of a random walk which admits a more detailed description of the interaction between polymer chains [63, 64].

One of the key-points of the double-stranded DNA denaturation is the so-called "loop factor" describing the entropy of the one-loop formation, $\Delta S(m) = -c \ln m$, where m is the length of the loop. The phase behavior, e.g., the order of the phase transition depends on the value of the c factor [34, 40, 61, 62, 65]. This factor was considered in many semi-empirical mean field theories [62, 66, 67] as a modification of Stockmayer's theory for sufficiently long chains. In e.g. [40], the excluded-volume interactions within denatured loops were taken into account and, on the basis of the Poland and Scheraga model [34] phase transitions of different order were shown to arise depending on the value of a loop exponent. However, in spite of extensive research the real-life phase behavior of the double-strand DNA still remains unclear. For instance, in [40] and references therein a phase transition of the first order was reported. At the same time, in [60] evidence is provided for second order phase transition at the end of the double-strand DNA melting. Such diversity of experimental results is mainly caused by different experimental conditions. Conformational statistics of the long loops and parameter c are highly sensitive to the counter-ion concentration, pH etc. The problem which still remains unsolved is the relationship between very

diverse experimental conditions of and the value of the c factor.

As well as in the previous sections, the main idea of the proposed approach is the mapping of two interacting three-dimensional polymer chains to a single two-dimensional random walk interacting with the origin.

Our approach has a number of advantages. First of all, we have taken into account the self-avoiding effects of each chain, since the helix initialization (first base-pair formation in the helix) takes place only at the first return of the two-dimensional random walk. Second, the approach used permits one to avoid the meander- and knot-like conformations. The traditional approach using $\Delta S(m)$ deals with any return of the random walk. At the same time, to address the loop entropy $\Delta S(m)$ correctly, it is necessary to use only the *first* returns of the random walk, as in our case. Thus, our approach deals with two-strand polynucleotides without any preliminary assumptions concerning $\Delta S(m)$.

The interaction between two strands and the self-interaction inside each strand in the coil regions include mainly Van-der-Waals and electrostatic interactions. The latter is the most long-range one among the non-helical interactions. Happily, the DNA-solvent system as a whole can be considered as electro-neutral one, since it contains different salts and other low-molecular compounds which screen the electrostatic interactions on the length scale r_D, where r_D is the Debye radius.

We discuss two ways of the hydrogen bond formation. The first one is an instant contact between the polymer chains which leads to the creation of a single hydrogen bond with subsequent free evolution of both polymers. This contact interaction is compensated by the short range intermolecular repulsion, and we ascribe the energy $U_1 > 0$ to it. The second way is the creation of a sequence of hydrogen bonds. This case corresponds to "glued" polymer chains in the helical phase where intermolecular repulsion is suppressed. We ascribe the energy $U_2 < 0$ to the monomer-monomer contacts in the helical region. This energy actually is the sum of two terms: the energy of hydrogen bonds and the

energy of stacking interaction between the given base pair and the preceding base pair in the helical region.

Each nucleotide is a group of atoms described by internal degrees of freedom, the dihedral angles. The base pair formation gains the energy but results in the entropy loss. To address the issue of the internal structure of nucleotides, we introduce a new factor $q = e^{\Delta S}$, where ΔS is the entropy loss caused by one base pair formation. We suppose that there is no another interaction in the middle part of the loop because the radius r_D is short enough at the physiological conditions.

The one-strand loop is presented as a walk of the particle. Effectively, we multiply the whole weight of the random walk trajectory by $k_1 = e^{\frac{-U_1}{T}}$ for visiting the origin and by $k_2 = e^{\frac{-U_2}{T}}$ for staying at the origin. The partition function for this model is

$$\Lambda = \sum_{j=0}^{\infty} (qk_1)^j F(z)^j \left(\sum_{m=0}^{\infty} (qk_2 z)^m \right)^j |_{z^N} = \frac{1}{2\pi i} \oint\limits_{C_0} \frac{1}{z^{N+1}} \frac{dz}{1 - \frac{qk_1 F(z)}{1-qk_2 z}}. \qquad (2.47)$$

The effective interaction $U_1 > 0$ depends on the result of competition between the repulsive interaction and the binding energy. $U_2 < 0$ corresponds to the attractive standing at the origin. The factor q with $q < 1$ mimics the fact that each base pair formation is unfavorable entropically. A microscopic study of these competing interactions using the analogy with the self-avoiding random walks was considered in [40].

2.6.1. The analysis of the partition function

Now let us discuss the partition function (2.47)

$$\Lambda = \frac{1}{2\pi i} \oint\limits_{C_0} \frac{1}{z^{N+1}} \frac{(1 - qk_2 z)dz}{1 - qk_2 z - qk_1 F(z)}, \qquad (2.48)$$

where we consider two cases depending on values k_1, k_2 and q.

If $k_1 + k_2 > \frac{1}{q}$ for $0 < q < 1$ we have a simple pole only at the positive

point z_+ which can be found by solving the equation

$$k_2 z_+ + k_1 F(z_+) = \frac{1}{q}. \tag{2.49}$$

For the integral Λ we derive (2.54)

In the second case, when $k_1 + k_2 < \frac{1}{q}$, there are no poles inside the contour C_1 except 0, so it is necessary to estimate an integral along indentation around the points ± 1:

$$\Lambda_{MP} = \frac{1}{2\pi i} \int\limits_M^P \frac{1}{z^{N+1}} \frac{(1-qk_2 z)dz}{1-qk_2 z - qk_1\left(1 - \frac{1}{-\frac{1}{\pi}\log(1-z)}\right)} = \int\limits_0^\delta \frac{1}{(z+1)^{N+1}}$$
$$\times \frac{\frac{1}{\pi}(1-qk_2(z+1))(qk_1 + qk_2(1+z)\frac{1}{\pi}\log(z))dz}{\left[qk_1 - (1-qk_1-qk_2(z+1))\frac{1}{\pi}\log(z)\right]^2 + \left[1-qk_1-qk_2(z+1)\right]^2}. \tag{2.50}$$

Using the fact that δ is small, we get

$$\begin{aligned} \Lambda_{MP} &= \frac{(1-qk_2)k_1\pi}{(1-qk_1-qk_2)^2} \int\limits_0^\delta \frac{1}{(z+1)^{N+1}} \frac{dz}{\log^2(z)} \\ &= \frac{(1-qk_2)qk_1\pi}{(1-qk_1-qk_2)^2} \frac{1}{N\log^2(N)}. \end{aligned} \tag{2.51}$$

In the same way, we can show that

$$\Lambda_{M'P'} = \frac{(1+qk_2)qk_1\pi}{(1-qk_1+qk_2)^2} \frac{1}{N\log^2(N)}. \tag{2.52}$$

Thus, for the whole integral Λ we obtain

$$\Lambda = \left(\frac{(1-qk_2)qk_1\pi}{(1-qk_1-qk_2)^2} + \frac{(1+qk_2)qk_1\pi}{(1-qk_1+qk_2)^2} \right) \frac{1}{N\log^2(N)} \tag{2.53}$$

Finally, to find an average energy per step and helicity, we substitute (2.53) in formulas (2.56) and obtain 0 for both of them in the limit of large N.

2.7. Results and discussion of Model C

The thermal behavior of the system is encoded in the singularities of the integrand of the partition function (2.47). Depending on the parameters of the

model the critical temperature T_c exists where the singular behavior is changed. The analysis of possible cases is presented in 2.6.1. Specifically, we derive for $T < T_c$

$$\Lambda = \frac{1}{z_+^{N+1}} \frac{1 - qk_2 z_+}{qk_2 + qk_1 F'(z_+)}. \tag{2.54}$$

and for $T > T_c$ Λ tends to 0 as a $\frac{1}{N \log^2(N)}$. The critical temperature T_c is determined from the equation

$$k_1 + k_2 = \frac{1}{q}. \tag{2.55}$$

To find the average energy and helicity, we use the formulas generalizing (2.22)

$$\begin{aligned} \overline{E} &= -\frac{1}{N} k_1 \log k_1 \frac{\partial \log \Lambda}{\partial k_1} - \frac{1}{N} k_2 \log k_2 \frac{\partial \log \Lambda}{\partial k_2} \\ \theta &= \frac{1}{N} k_1 \frac{\partial \log \Lambda}{\partial k_1} + \frac{1}{N} k_2 \frac{\partial \log \Lambda}{\partial k_2}, \end{aligned} \tag{2.56}$$

which give for $T < T_c$:

$$\begin{aligned} \overline{E} &= -\frac{k_1 \log k_1 F(z_+)}{z_+(k_2 + k_1 F'(z_+))} + \frac{qk_1 k_2 \log k_2 F(z_+)}{(k_2 + k_1 F'(z_+))(qz_+ k_2 - 1)} \\ \theta &= -\frac{k_1 F(z_+)}{z_+(k_2 + k_1 F'(z_+))(qz_+ k_2 - 1)}. \end{aligned} \tag{2.57}$$

The circles on Fig. 2.12 show that the helicity degree completely vanishes above the critical temperature T_c. This is in contrast with the simple random walk model with $k > 1$ shown by squares, where θ tends to zero asymptotically due to entropy effects. We conclude that the model with stops at the origin describes the helix-coil sharp transition. The same behavior takes place for the average energy.

If the contact interaction is attractive with $k > 1$ ($U < 0$), the system exhibits a gradual helix-coil transition. In the case of repulsive interaction with $k < 1$ ($U > 0$) we have zero helicity in the double-stranded region. More interesting behavior appears for a competing interaction of the random walk with the origin when $k_1 < 1$ ($U_1 > 0$) for the instantaneous contacts between

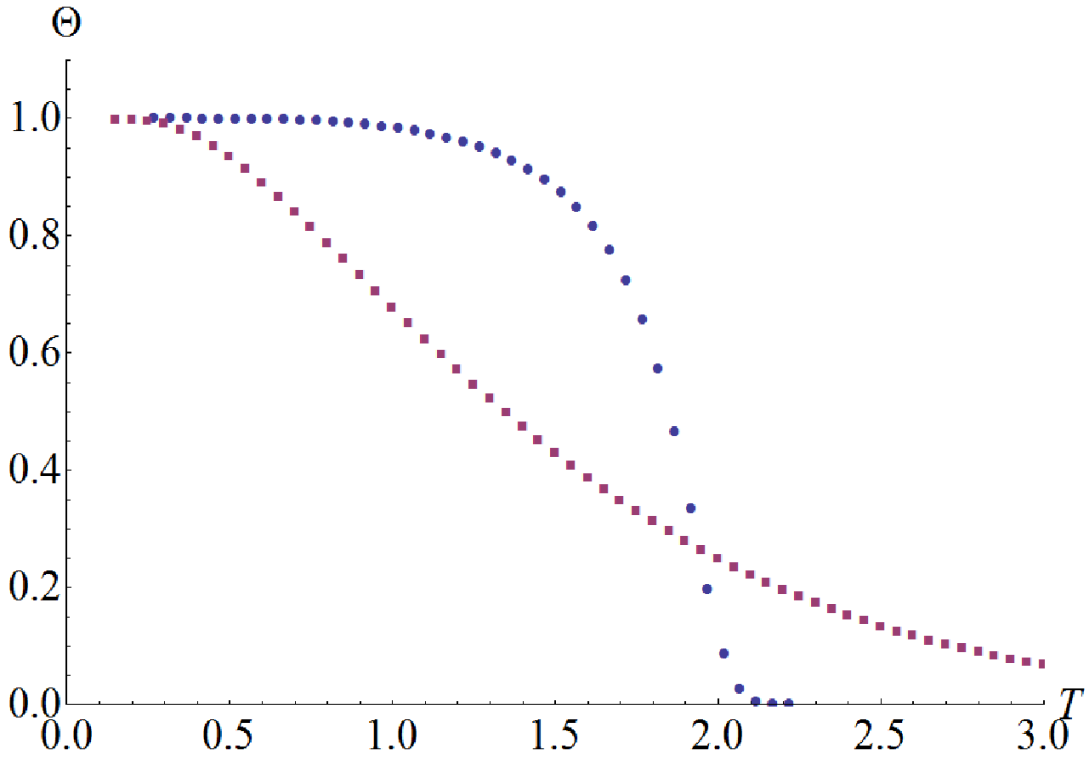

Fig 2.12. Dependence of helicity degree on the temperature. Squares show the case of the simple random walk model. Circles show the behavior of the random walk with stops at the origin ($U_1 = 1, U_2 = 1.5, q = 0.4, U = -1.5$).

polymer chains and $k_2 > 1$ ($U_2 < 0$) for their long contact. We also introduce a factor q which accounts for the entropy decrease in the base pair formation.

Under these conditions the system exhibits a sharp denaturation transition. The completely denatured state ($\theta = 0$) with two completely unbound DNA strands appears at finite temperatures $T > T_c$.

The key point of our study is the entropic nature of the parameter $q = e^{\Delta S}$, where ΔS is the entropy loss caused by the base-pair formation. Also, we obligatorily need a repulsion between non-paired nucleotides to obtain a sharp denaturation transition. In the opposite case, where there is attraction between non-paired nucleotides or there is no interaction between non-paired nucleotides, we have smooth denaturation. The given result is in qualitative agreement with [40], where the sharpness of the DNA melting was also assigned to repulsive interactions.

The main characteristics of the melting curve θ on the temperature T are

the melting temperature T_m and the interval of transition ΔT. The melting temperature is the measure of stability of the helical structure defined by the condition $\theta = \frac{1}{2}$ [34]. The melting interval ΔT is usually considered as a measure of cooperativity of a helix-coil transition [34]. It is characterized by the slope of the melting curve at the point T_m, $\Delta T = |\frac{d\theta}{dT}|^{-1}_{T=T_m}$. Figure 2.13 clearly shows that the helix stability increases with the strength of attraction U_2. At the same time, the melting cooperativity substantially decreases with the strength of attraction. The growth of stability is quite natural because attraction U_2 stabilizes the double-helix.

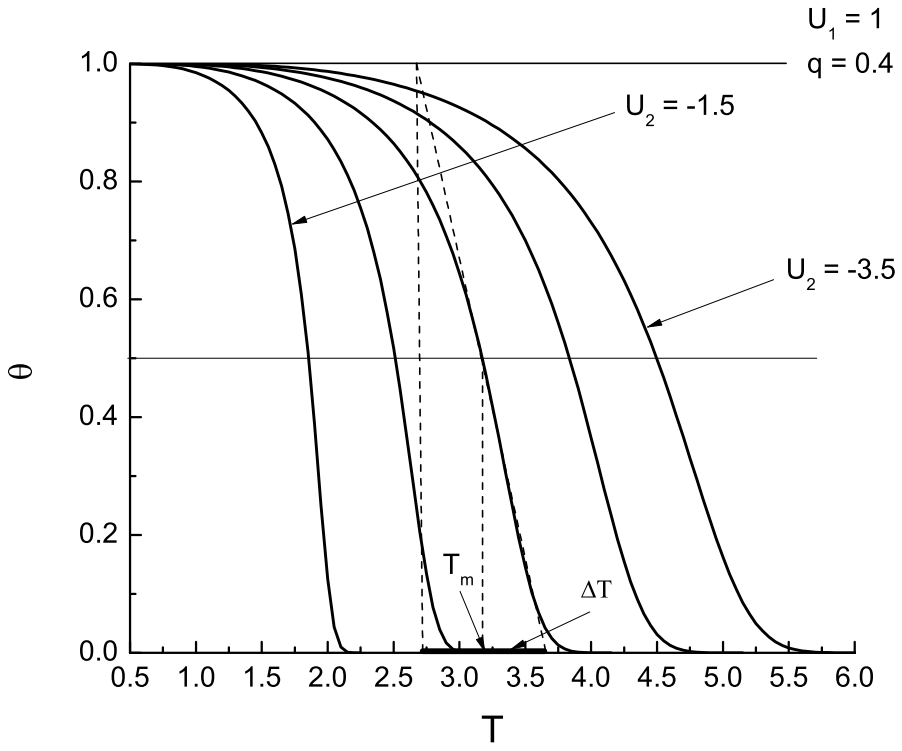

Fig 2.13. Helicity degree of Model C. Dependence of the helicity degree on the temperature in the case $q = 0.4, U_1 = 1$ and different values of U_2. The melting temperature T_m defined at the point where $\theta = 0.5$. The melting interval $\Delta T = |\frac{d\theta}{dT}|^{-1}_{T=T_m}$ defined as the slope of the melting curve at the melting temperature T_m.

Thus, using a very simple random walk model, one is able to describe the essentially complex behavior of the double-stranded polynucleotide.

The proposed model is in qualitative agreement with experimental results presented in [60], where sharp transition is exhibited at the end of the melting

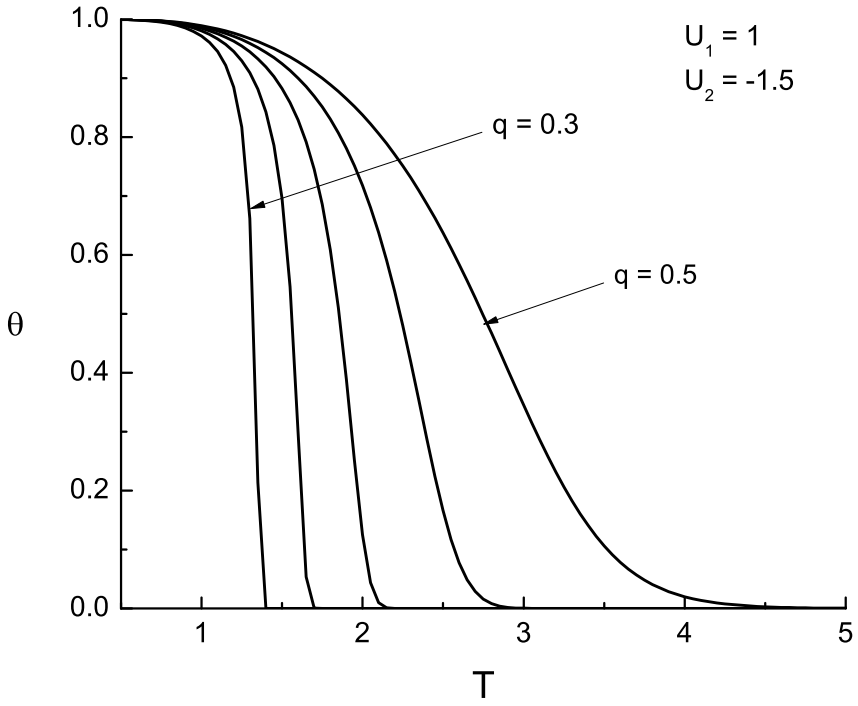

Fig 2.14. Helicity degree of Model C

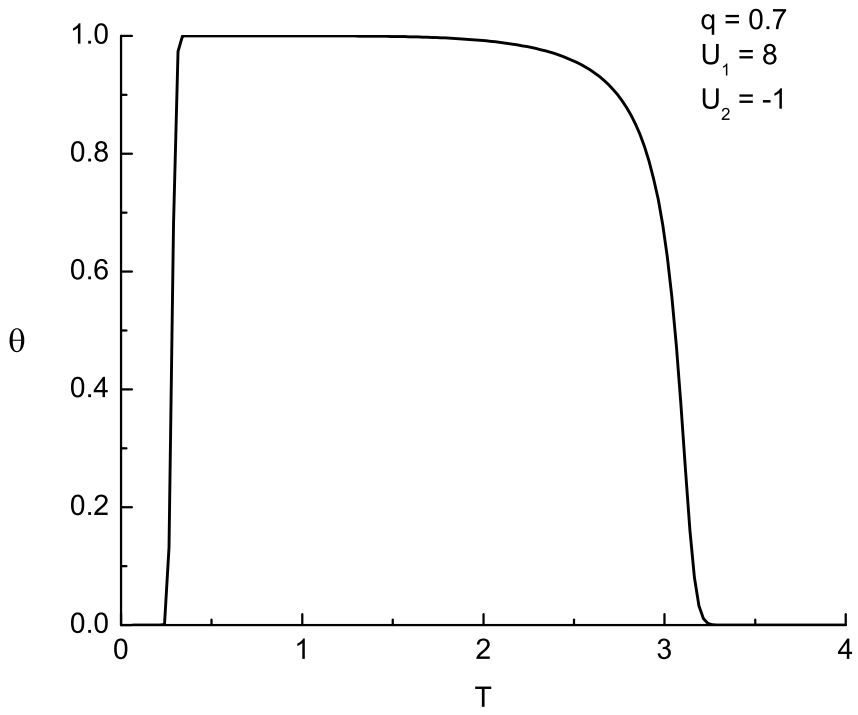

Fig 2.15. Helicity degree of Model C: cold denaturation

transition ($\theta = 0$). To our knowledge this is the first case when the theoretical phase behavior is confirmed by experiment. The order of the transition was interpreted by authors [60] as the second order. However, the number of measured

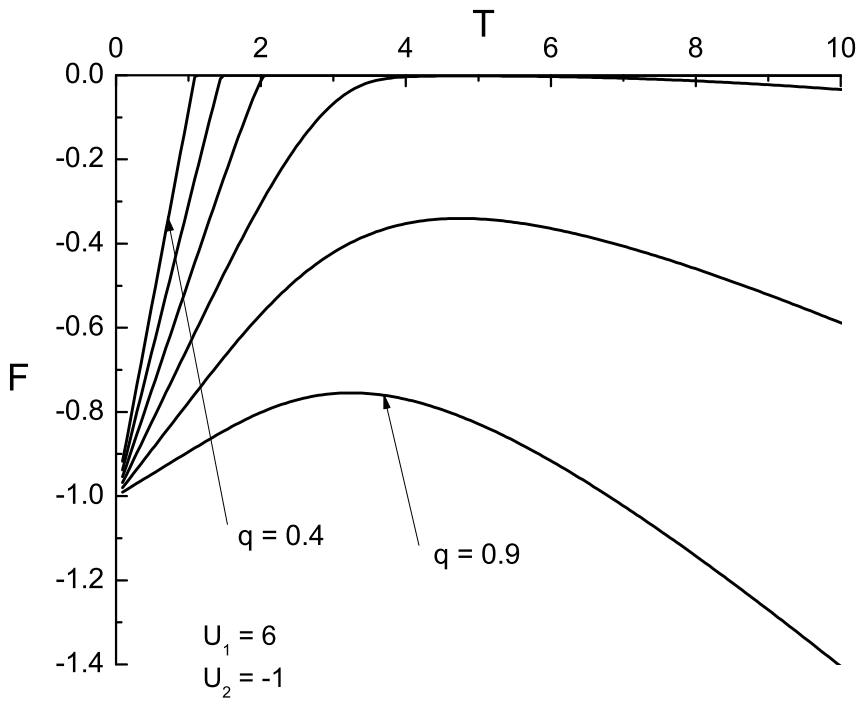

Fig 2.16. Free energy of Model C

points seems not to be enough to observe this experimentally. Our model clearly shows a sharp but continuous denaturation transition at the temperature $T = T_c$ and at the same time, gives a melting curve which qualitatively agrees with experiment [60].

Chapter 3

The secondary structural transitions in single-stranded RNA. The basic model.

3.1. Statement of the problem.

Single – stranded RNA (ssRNA) plays a central role in molecular biology. In addition to transmitting genetic information from DNA to proteins, RNA molecules participate actively in a variety of cellular processes [68]. Examples are translation (rRNA, tRNA, and tmRNA), editing of mRNA, intracellular protein targeting, nuclear splicing of pre-mRNA, and X-chromosome inactivation. Secondary structure of ssRNA is usually much more stable than tertiary structure. It can be explained by stronger interactions like hydrogen bonds an stacking-interactions, stabilizing secondary structure in comparison with tertiary [13]. Another explanation is the additional entropy loss, necessary for the stable tertiary formation, as it was shown in [69]. Thus, independent on the specific origin of the higher stability of secondary structure, the secondary structure prediction is possible without taking into account the tertiary structure formation.

Since the pioneering work of Higgs and Morgan [29, 70] and Bundschuh and Hwa [71, 72], several authors have studied the statistical physics of RNA secondary structures both for homopolymeric and heterogeneous RNAs and [71–75]. In dependence on model peculiarities ssRNA exhibits rich phase behavior including folding transitions, continuous freezing transition between molten and glass phase etc. Not much is known about the freezing transition, even from numerical work; indeed its localization is non-trivial [76]. Better studied numerically is the glass phase at strong disorder, or equivalently zero temperature [71, 73, 77, 78]. However, the nature of the freezing transition and of the low-temperature phase are still poorly understood, and contradictory results are

reported [78]. The main problem is to address the effect of the sequence disorder on the thermodynamics of ssRNA. The commonly used replica approach [79–84] still remain non-effective for ssRNA secondary structure investigation. The glass phase appears in the solution of [71, 72] for the partition function for $n = 2$ replicas (instead of $n = 0$ relevant for the disordered system) and in numerical simulations [71, 73, 78, 85, 86].

The main goal of the present report is to develop an approach to investigate thermodynamics of ssRNA with taking into account sequence heterogeneity.

3.2. The constrained annealing approach

We propose to study random ssRNA sequences composed of A, C, G, and U bases. Pairing is permitted only between A and U and between C and G bases. The topological rules that determine which structures are allowed are the essential feature that makes workable the numerical calculation of the free energy of secondary structure. The main rule is elimination of so called pseudoknots (Fig.3.1) from the set of available secondary structures as in most other work on ssRNA [29].

In this case the full partition function Z_N for the ssRNA chain of the length N can be calculated recursively at any given temperature T [29, 71]. To make the sequence effect tractable analytically we propose to use an approach developed by M. Serva and G. Paladin in [87]. Following by [87–89], the free energy of ssRNA with quenched random sequence of nucleotides can be estimaed on the basis of annealed averages of the partition function with appropriate constraints. Given approach is substantially variational and can be realized by the aid of Lagrange multipliers, which serve as a variational parameters. The relationship between the quenched and annealed disorder in ssRNA has been addressed numerically in [90].

Disordered systems like spin glasses or random heteropolymers are characterized by two types of degrees of freedom: annealed which arrange themselves to

minimize the free energy and quenched which can be considered as constant in time. In case of ssRNA annealed degrees of freedom are Watson-Crick base pairs. The nucleotides sequence can be addressed as a set of quenched degrees of freedom. According to [87] the free energy of the ssRNA with random quenched sequence of nucleotides can be estimated as

$$f \geq g(T, \mu) \geq f_A, \tag{3.1}$$

where f and f_A are the reduced quenched and annealed free energy per nucleotide, correspondingly and

$$g(T, \mu) = -\frac{1}{N} \ln \overline{Z(seq) e^{-N\mu\alpha(seq)}}. \tag{3.2}$$

$Z(seq)$ is the parition function of ssRNA with given sequence realization seq and $\alpha(seq)$ is the appropriate self-averaging quenched quantity. $\overline{\mathcal{O}}$ means the average over sequence distribution function. We will refer given approach below as a "constraint annealing approach".

3.3. The model

A primary RNA structure is fully determined by the base sequence which is a list of nucleotides, cytosine (C), guanine (G), adenine (A), or uracil (U) with N entries. In agreement with previous treatments, a valid secondary structure is a list of all base pairs with the constraint that a base can be part of at most one pair. In addition, pseudoknots are not allowed, i.e., for any two base pairs (i, j) and (k, l) with $i < j$, $k < l$, and $i < k$ we have either $i < k < l < j$ or $i < j < k < l$.

Hamiltonian of the model is written as

$$\mathcal{H}(\hat{m}, \{h\}) = \sum_{i<j} m_{ij}(\epsilon_0 + \epsilon h_i h_j), \tag{3.3}$$

where sum is taken over all non-repeated base pairs, $m_{ij} = 1$ if the bases i and j are paired and $m_{ij} = 0$ otherwise. The partition function for the ssRNA chain

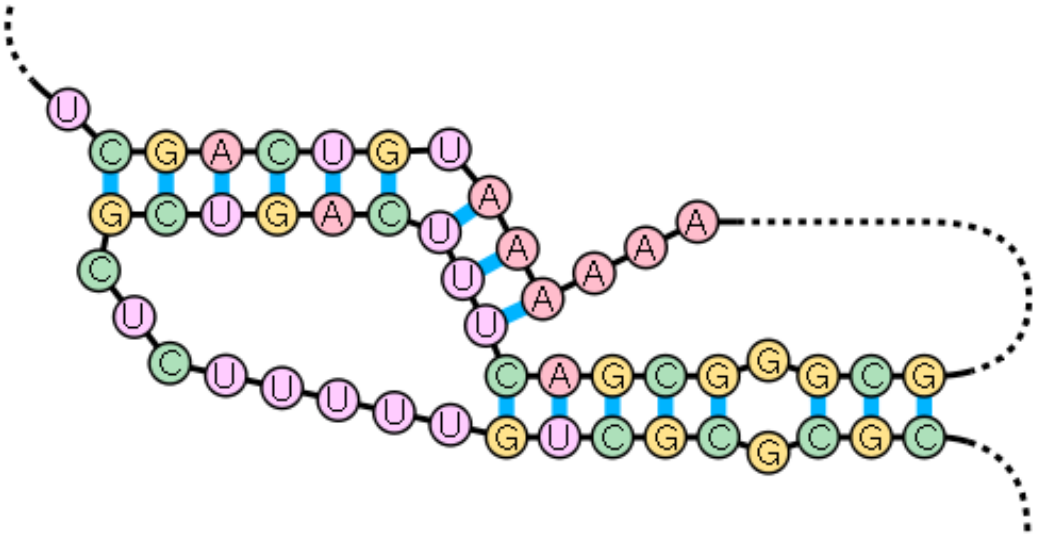

Fig 3.1. A pseudoknot is an RNA secondary structure containing at least two stem-loop structures in which half of one stem is intercalated between the two halves of another stem.

of N nucleotides is written as

$$Z_N(\{h\}) = \sum\nolimits'_{\hat{m}} \exp[-\beta \mathcal{H}(\hat{m}, \{h\})], \tag{3.4}$$

where $\beta = \frac{1}{k_B T}$ and the sum is taken over all realizations of the matrix $\hat{m}$, which are not include pseudoknots and containing not more than one unity on each row or column. The latter condition describes the saturation of base pairing.

3.3.1. Gaussian disorder

Let us consider first the case of Gaussian disorder. Then, the distribution function for the sequence $\{h\}$ is written

$$\mathcal{P}\{h\} = \prod_{i=1}^{N} \rho(h_i), \tag{3.5}$$

where $\rho(h_i) = (2\pi D)^{-1/2} e^{-\frac{h_i^2}{2D}}$. The reduced free energy per nucleotide is written as

$$f\{h\} = -\frac{1}{N} \ln Z_N(\{h\}) \tag{3.6}$$

In the thermodynamic limit $N \to \infty$ the free energy becomes a non-random quantity and $f\{h\} = f$, where f is the quenched free energy

$$f = -\frac{1}{N}\overline{\ln Z_N(\{h\})} \tag{3.7}$$

Following by [87] the quenched free energy can be estimated as $\max_\mu g(T,\mu)$ using the inequality (4.5). Let us introduce the following constraints for the quenched variables $\{h\}$

$$\alpha_1\{h\} = \frac{1}{N}\sum_{i=1}^{N} h_i \tag{3.8}$$

$$\alpha_2\{h\} = \frac{1}{N}\sum_{i=1}^{N}(h_i^2 - D)$$

The effective partition function is written

$$\mathcal{Z}_N = \overline{e^{-N\mu_1\alpha_1\{h\}-N\mu_2\alpha_2\{h\}}Z_N(\{h\})} \tag{3.9}$$

and can be presented as (see section 3.5 for details)

$$\mathcal{Z}_N = \omega^N \mathcal{Z}_N^0(\tilde{\epsilon}), \tag{3.10}$$

where

$$\mathcal{Z}_N^0(\tilde{\epsilon}) = \sum_{\hat{m}}{}' e^{-\beta(\epsilon_0+\tilde{\epsilon})\sum_{i<j} m_{ij}}$$

and

$$\omega = \frac{e^{\mu_2 D + \frac{\tilde{D}\mu_1^2}{2}}\sqrt{\tilde{D}}}{\sqrt{D}} \tag{3.11}$$

$$\tilde{D} = \frac{D}{1+2D\mu_2}$$

$$\beta\tilde{\epsilon} = \frac{\beta\epsilon\mu_1^2\tilde{D}^2}{1+\beta\epsilon\tilde{D}} + \frac{1}{2}\ln\left[1-(\beta\epsilon\tilde{D})^2\right]$$

$\mathcal{Z}_N^0(\tilde{\epsilon})$ is the partition function of homopolymeric ssRNA with effective energy $\epsilon = \epsilon_0 + \tilde{\epsilon}$. As it was shown in [71],

$$\mathcal{Z}_N^0(\tilde{\epsilon}) = A_0(Q)N^{-\theta_0}z(Q)^N, \tag{3.12}$$

where $\theta_0 = 3/2$, $Q = e^{\beta(\epsilon_0+\tilde{\epsilon})}$ and

$$z(Q) = 1 + 2\sqrt{Q} \tag{3.13}$$
$$A_0(Q) = \sqrt{\frac{1+2\sqrt{Q}}{4\pi Q^{3/2}}}$$

The variational reduced free enery $g(\beta, \mu_1, \mu_2) = -\frac{1}{N}\ln \mathcal{Z}_N$ is written

$$g(\beta,\mu_1,\mu_2) = -\frac{1}{2}\ln\frac{\tilde{D}}{D} - \mu_2 D - \frac{\tilde{D}\mu_1^2}{2} - \ln\left(1+2[1-(\beta\epsilon\tilde{D})^2]^{-1/4} e^{-\frac{\beta\epsilon_0}{2} - \frac{\mu_1^2}{2}\frac{\beta\epsilon\tilde{D}^2}{1+\beta\epsilon\tilde{D}}}\right) \tag{3.14}$$

$g(\beta, \mu_1, \mu_2)$ reach maximal value at $\mu_1 = 0$ (see section 3.6 for details). Thus, we need to maximize the variational free energy over the variable $\tilde{D}$

$$g(\beta,\tilde{D}) = -\frac{1}{2}\ln\frac{\tilde{D}}{D} + \frac{1}{2} - \frac{D}{2\tilde{D}} - \ln\left(1+2[1-(\beta\epsilon\tilde{D})^2]^{-1/4} e^{-\frac{\beta\epsilon_0}{2}}\right) \tag{3.15}$$

Maximization results to the equation

$$\frac{D}{\tilde{D}} = 1 + \frac{(\beta\epsilon\tilde{D})^2}{1-(\beta\epsilon\tilde{D})^2}\Theta(\ln 2 - \frac{\beta\epsilon_0}{2} - \frac{1}{4}\ln[1-(\beta\epsilon\tilde{D})^2]), \tag{3.16}$$

where $\Theta(x) = \frac{e^x}{1+e^x}$ is logical function. Equation (3.16) can be solved numerically, and its solution is unique, positive and continuously changing with temperature.

The free energy per monomer of the system is estimated as

$$f(\beta) = \frac{g(\beta)}{\beta}, \tag{3.17}$$

where $g(\beta) = g(\beta, \tilde{D}_0)$ and $\tilde{D}_0$ is solution of the equation (3.16). The entropy per monomer is written as

$$s(\beta) = -g(\beta) + \beta\frac{dg(\beta)}{d\beta} \tag{3.18}$$

and the specific heat

$$c_V(\beta) = -\beta^2\frac{d^2 g(\beta)}{d\beta^2} \tag{3.19}$$

Let us define helicity degree as the mean part of Watson-Crick base pairs

$$\theta = \frac{2}{N}\overline{\langle\sum_{i<j} m_{ij}\rangle} = \Theta(\ln 2 - \frac{\beta\epsilon_0}{2} - \frac{1}{4}\ln[1-(\beta\epsilon\tilde{D}_0)^2]), \tag{3.20}$$

where $\langle \mathcal{O} \rangle$ is thermodynamic average and $\tilde{D}_0$ is the solution of the equation (3.16). Temperature behavior of the thermodynamic parameters is calculated on the basis of equations (3.16,3.17,3.18,3.19,3.30).

The entropy of the model with disorder is substantially less than those for homopolymer. In the low-temperature limit the entropy of the model with Gaussian disorder exhibits logarithmic divergence with temperature as (for details see section 3.7).

$$s \simeq -ln(\epsilon \beta D) \tag{3.21}$$

Thus, at low enough, but finite temperatures entropy becomes negative $s < 0$. However, despite of ordinary and spin glasses, entropy crisis itself does not characterize the glass phase appearance, because of our model contains continuous degrees of freedom ($\{h\}$).

In dependence on parameters ϵ_0, ϵ and D model exhibits different temperature behavior of thermodynamic parameters. At the definite choice of the parameters the temperature behavior of specific heat exhibits two peaks. The high-temperature peak corresponds to the melting transition. It is necessary to mention that even at the infinitely large temperatures helicity degree still remain $\theta = 2/3$. In the area of the low-temperature peak the helicity degree $\theta \approx 1$. From the equations (3.21) and $c_V = T(ds/dT)$ it straightforwardly follows that $c_V(T = 0) > 0$. Thus, at low- and even zero-temperatures system has available degrees of freedom, although the entropy is drastically decreased in comparison with homopolymeric case. Similar temperature behavior of the specific heat has been observed recently e.g. in some models displaying glassy behavior at zero temperature due to entropic barriers [91]. From the one side we have a low-energy ground state (at $T = 0$) practically without unbounded base pairs ($\theta \approx 1$). From another side, specific heat behaves as a classical model with Maxwell-Boltzmann statistics at the high-temperature area. Thus, at low temperatures we have a highly ordered system ($\theta \approx 1$) with definite conformational freedom, which is signalling about the possibility of the low-temperature glassy state

appearance. However, Gaussian model does not provide enough evidence for the glass transition and we need to investigate more realistic model of ssRNA.

3.3.2. Bimodal disorder

In previous considerations of ssRNA folding the sequence disorder usually supposed to be Gaussian [71, 92–96] to make model tractable analytically. However, the real RNA sequence is composed from four-literal alphabet. For the sake of simplicity we consider the case of two-literal sequence to assign variable $h_i = \pm 1$ to each i-th nucleotide. It corresponds e.g. to random $poly(AU)$ sequence.

Then, the distribution function for the sequence $\{h\}$ is written

$$\mathcal{P}\{h\} = \prod_{i=1}^{N} P(h_i), \tag{3.22}$$

where $P(h_i) = q\delta(h_i - 1) + (1 - q)\delta(h_i + 1)$.

Let us introduce the following constraints for the quenched variables $\{h\}$

$$a_1\{h\} = \frac{1}{N}\sum_{i=1}^{N}[h_i - (2q - 1)] \tag{3.23}$$

$$a_2\{h\} = \frac{1}{N}\sum_{i=1}^{N}[(h_i - (2q - 1))^2 - 4q(1 - q)]$$

It is obvious that $\overline{a_1\{h\}} = 0$ and $\overline{a_2\{h\}} = 0$. The effective partition function is written

$$\mathcal{Z}_N = \overline{e^{-N\mu_1 a_1\{h\} - N\mu_2 a_2\{h\}} Z_N(\{h\})} \tag{3.24}$$

and can be presented as (see section 3.8 for details)

$$\mathcal{Z}_N = e^{N\mu(2q-1)}\Omega^N \mathcal{Z}_N^0(\bar{\epsilon}), \tag{3.25}$$

where

$$\mathcal{Z}_N^0(\bar{\epsilon}) = \sum_{\hat{m}}{}' e^{-\beta(\epsilon_0 + \bar{\epsilon})\sum_{i<j} m_{ij}}, \tag{3.26}$$

$\mu = \mu_1 - 2(2q-1)\mu_2$ and

$$\Omega(\mu) = qe^{-\mu} + (1-q)e^{\mu} \tag{3.27}$$

$$W(\mu, \beta, \epsilon) = e^{-\beta\epsilon}[q^2 e^{-2\mu} + (1-q)^2 e^{2\mu}] + 2q(1-q)e^{\beta\epsilon}$$

$$e^{-\beta\bar{\epsilon}} = \frac{W(\mu, \beta, \epsilon)}{\Omega(\mu)^2}$$

Thus, parameters μ_1 and μ_2 presented in $\mathcal{Z}_N$ only as $\mu = \mu_1 - 2(2q-1)\mu_2$ and variational problem becomes one-parameter. We need to maximize the variational potential

$$g(\beta, \mu) = -\mu(2q-1) - \ln\Omega(\mu) - \ln(1 + 2\sqrt{\bar{Q}}) \tag{3.28}$$

over μ, where $\bar{Q} = e^{-\beta(\epsilon_0 + \bar{\epsilon})}$. Maximization results to the equation

$$2q - 1 = \left[\frac{2\sqrt{\bar{Q}}}{1 + 2\sqrt{\bar{Q}}} - 1\right]\frac{d\ln\Omega(\mu)}{d\mu} - \frac{1}{2}\frac{2\sqrt{\bar{Q}}}{1 + 2\sqrt{\bar{Q}}}\frac{\partial \ln W(\mu, \beta, \epsilon)}{\partial\mu} \tag{3.29}$$

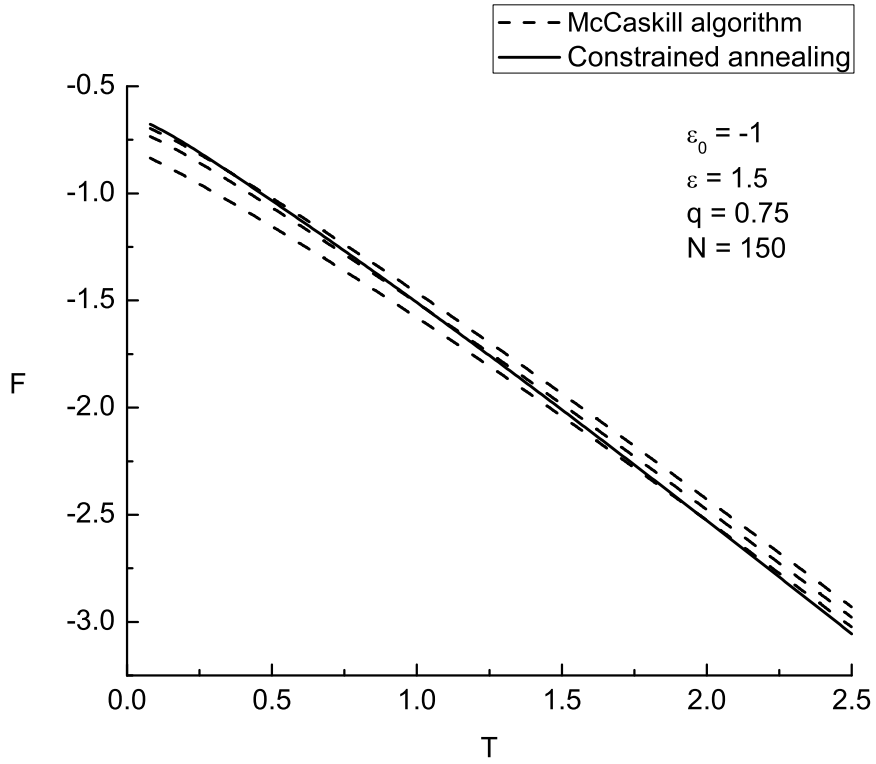

Fig 3.2. Dependence of reduced free energy on temperature.

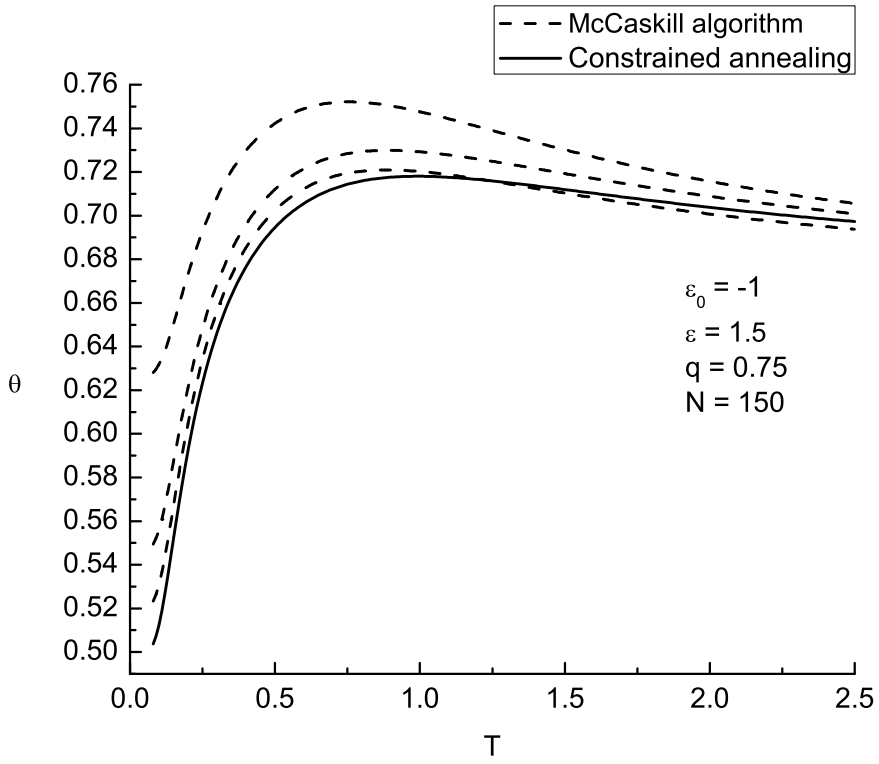

Fig 3.3. Dependence of helicity degree on temperature.

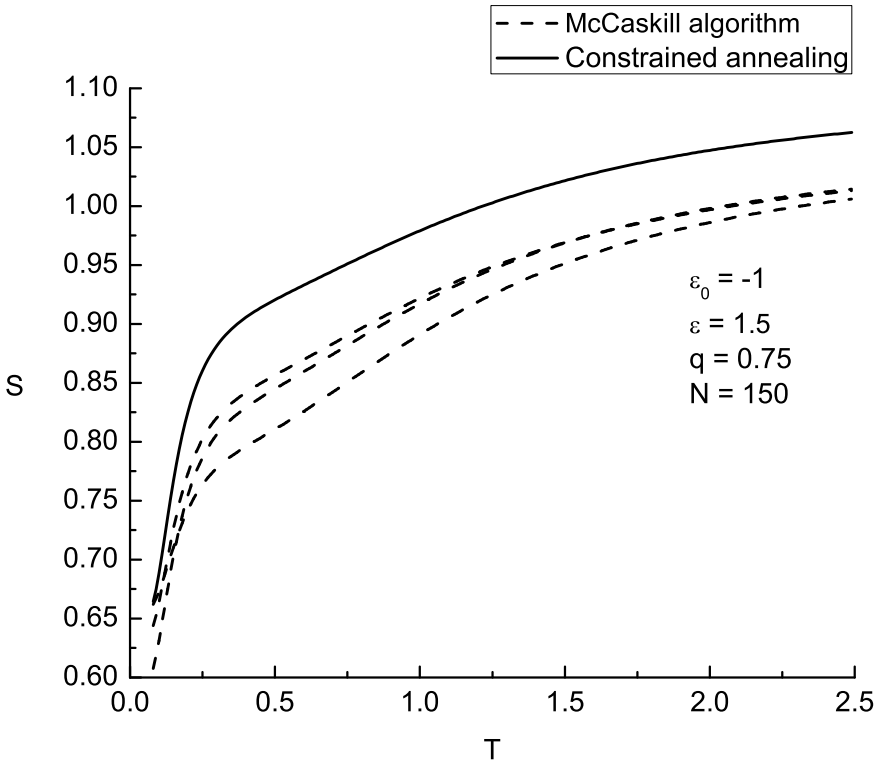

Fig 3.4. Dependence of entropy on temperature.

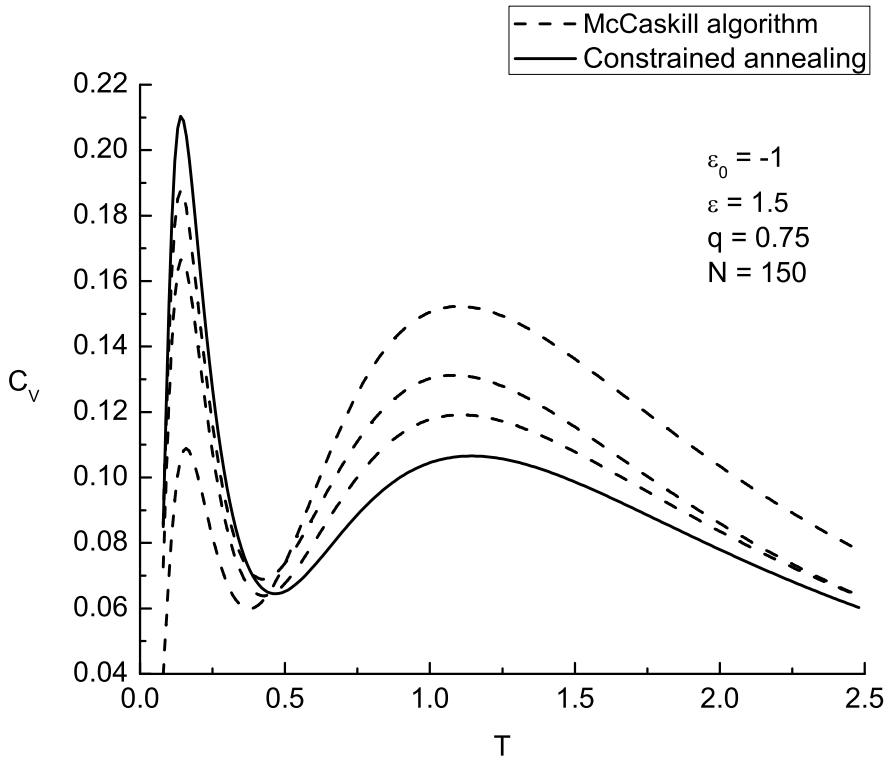

Fig 3.5. Dependence of specific heat on temperature.

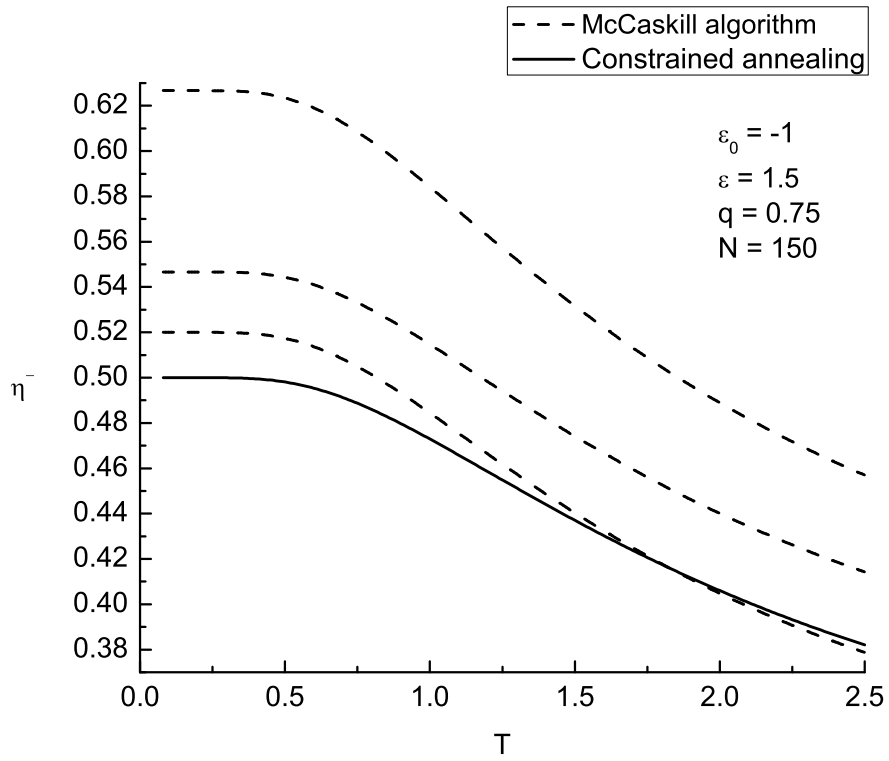

Fig 3.6. Dependence of the portion of the energetically favorable contacts on temperature.

3.4. Results and Discussion

In this section we presents results mainly relating to the case of bimodal disorder. We believe that this case is the most relevant for the RNA sequence.

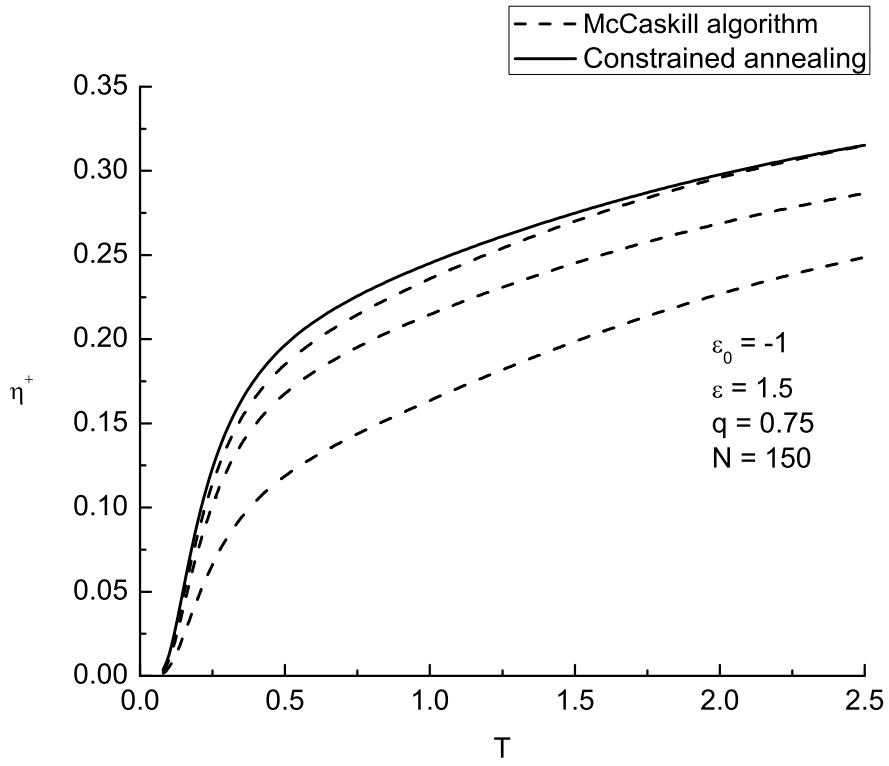

Fig 3.7. Dependence of the portion of the energetically unfavorable contacts on temperature.

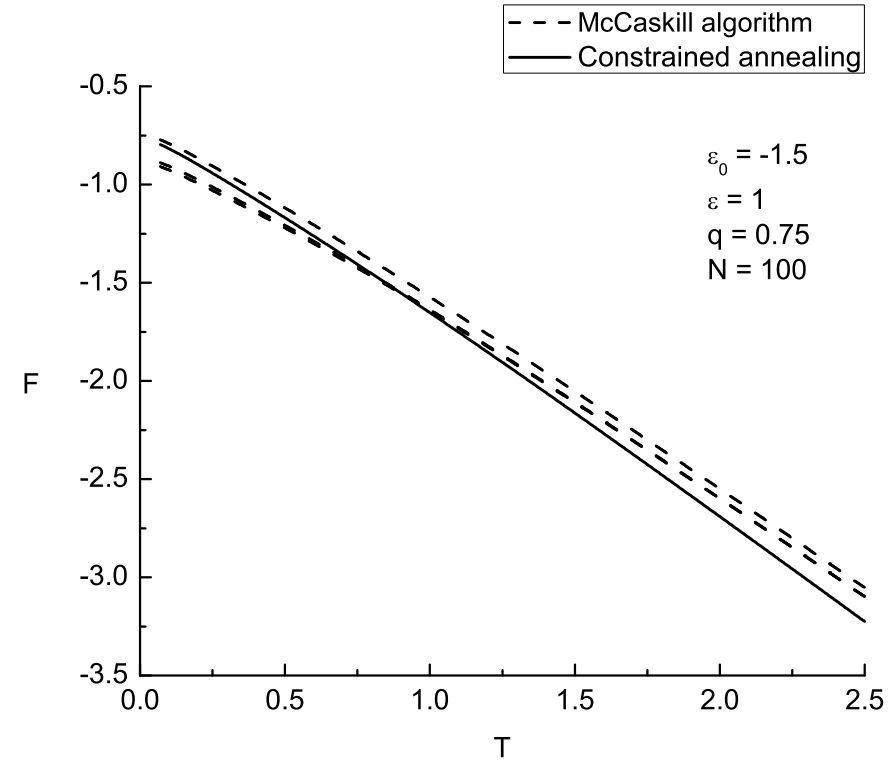

Fig 3.8. Dependence of reduced free energy on temperature.

In Fig.3.14a we compare free energies obtained by constrained annealing *versus* directly calculated using the recursion algorithm, based on Eq. (1.5). The mean value of the free energy calculated numerically is in a good agreement

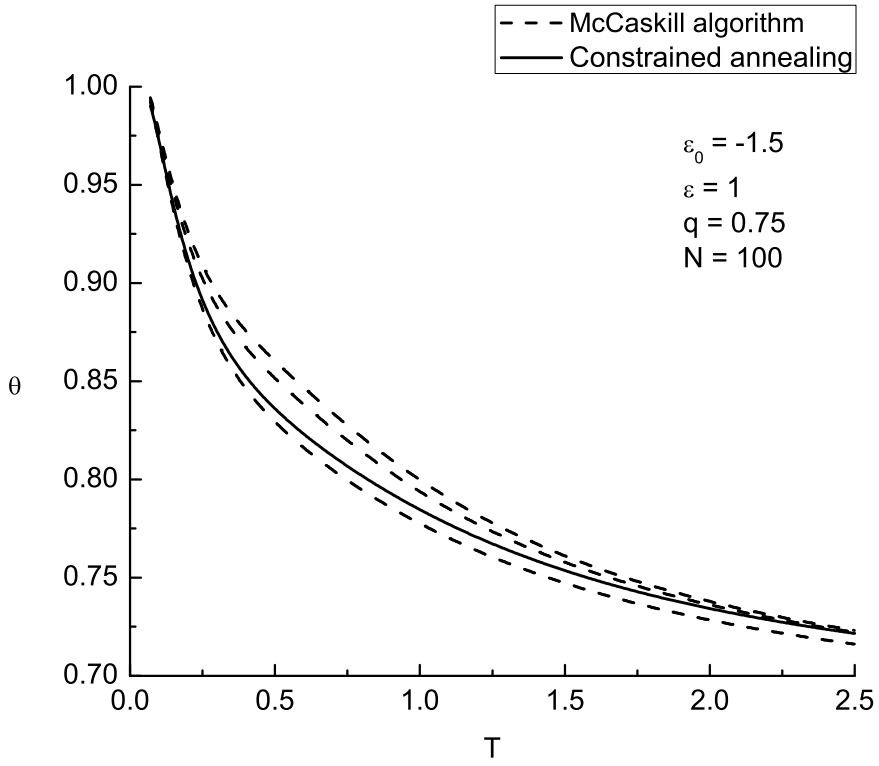

Fig 3.9. Dependence of helicity degree on temperature.

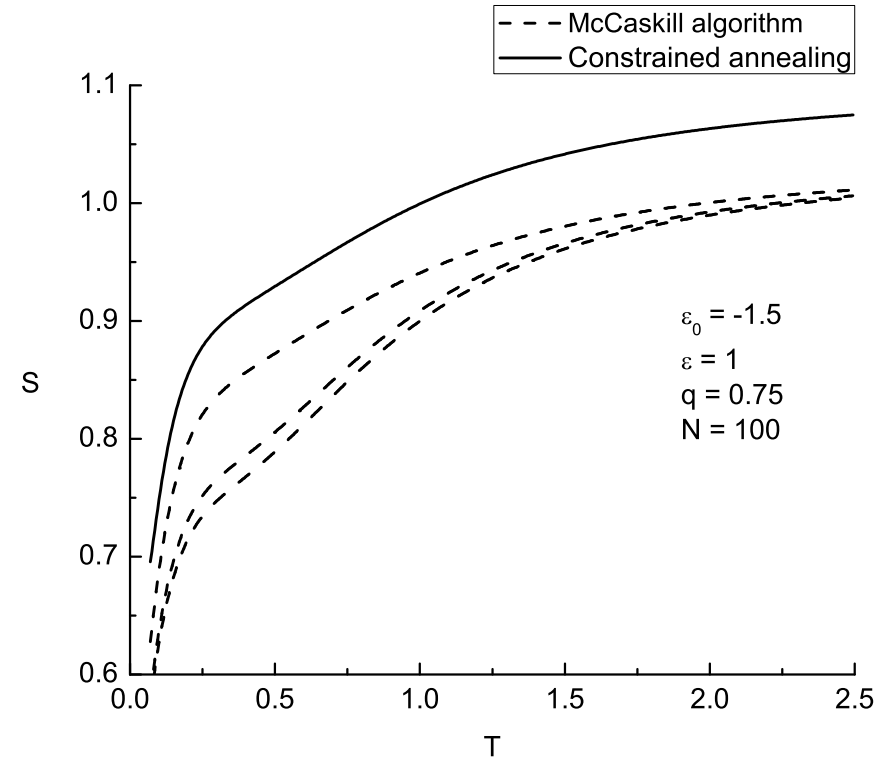

Fig 3.10. Dependence of entropy on temperature.

with the free energy estimated with the help of constrained annealing method [97].

In Fig.3.14b, we compare specific heats, obtained by constrained annealing

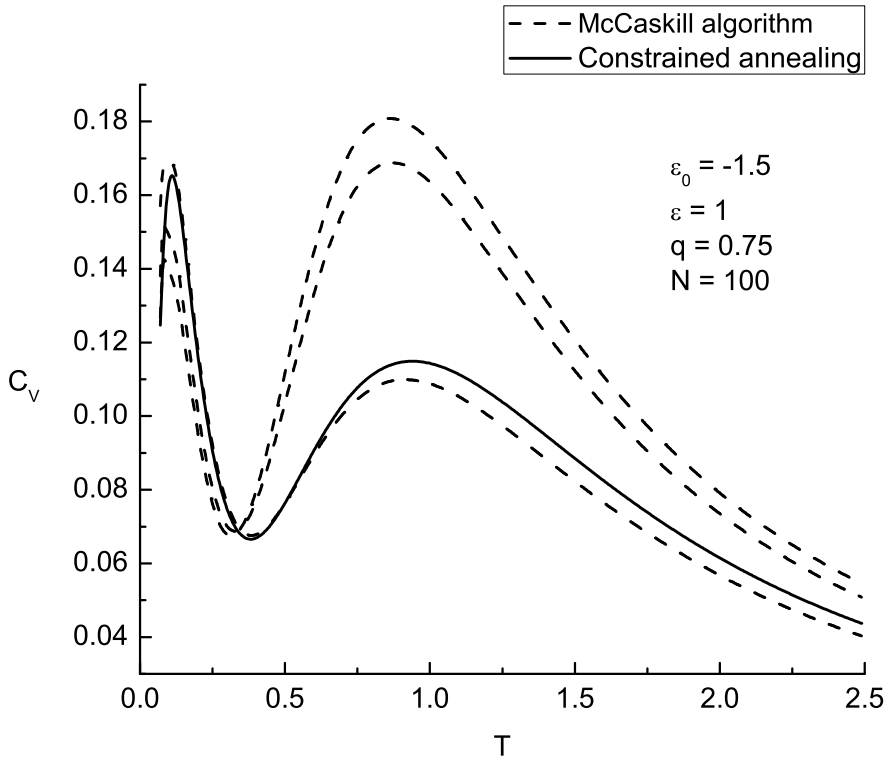

Fig 3.11. Dependence of specific heat on temperature.

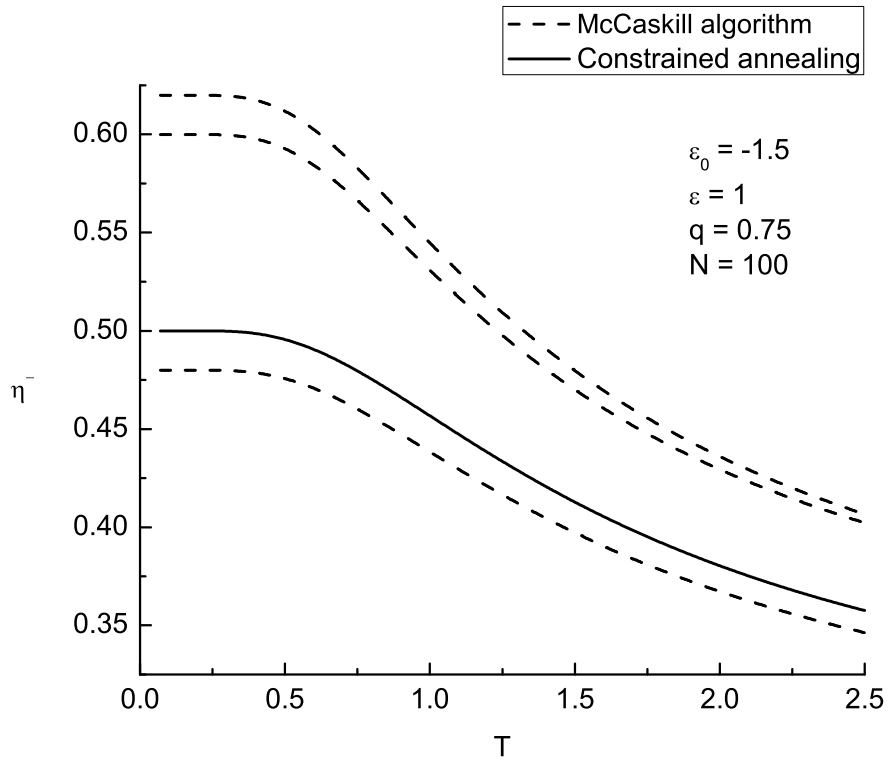

Fig 3.12. Dependence of the portion of the energetically favorable contacts on temperature.

versus the directly calculated using McCaskill's algorithm [32], based on Eq. (1.5) with subsequent numerical differentiation of the free energy by temperature. The temperature behavior of the specific heat is in a reasonable agreement

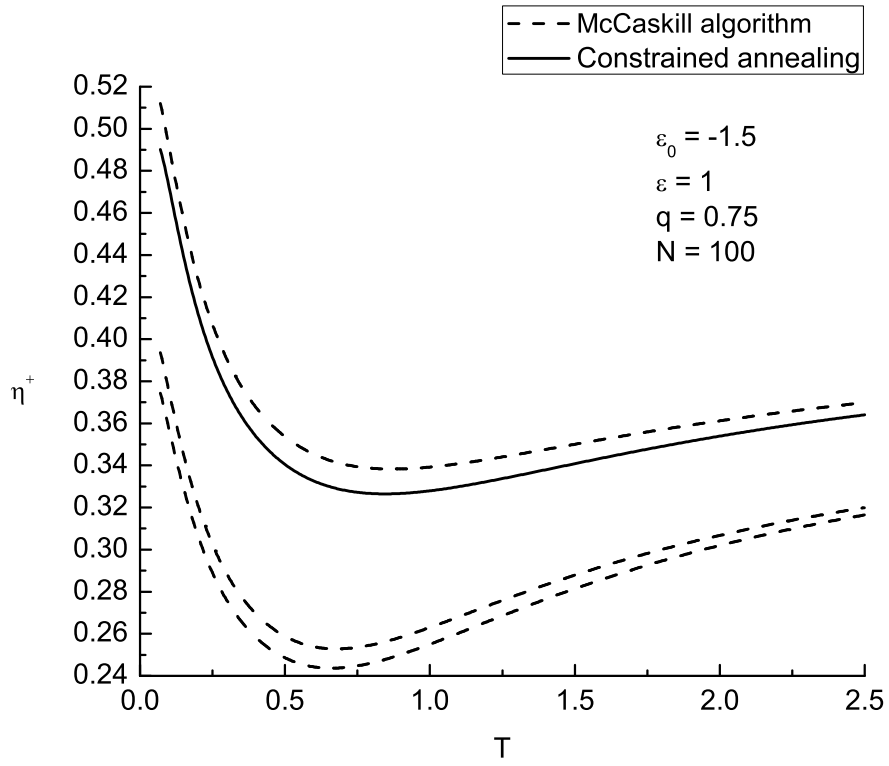

Fig 3.13. Dependence of the portion of the energetically unfavorable contacts on temperature.

with the McCaskill's results and exhibits two-peaks, a sign of two structural transitions.

To assign the specific heat behavior to the structural transformations of ssRNA let us define helicity degree as a mean portion of Watson-Crick base pairs

$$\theta = \frac{2}{N}\overline{\langle \sum_{i<j} m_{ij} \rangle} = \frac{2\sqrt{\bar{Q}}}{1+2\sqrt{\bar{Q}}}, \tag{3.30}$$

where $\langle \mathcal{O} \rangle$ is thermodynamic average. The r.h.s of the Eq. (3.30) is given by the expression of helicity degree of the homopolymeric RNA, straightforwardly obtained from the partition function of homopolymeric ssRNA [71]. Thus, in the constrained annealing approximation, helicity degree is written as for homopolymeric RNA with the effective statisticalcite weight $\bar{Q}$.

Out of Eq. (3.30) the helicity degree can be estimated numerically by making use of the probability of base pair formation between nucleotides i and j [71]

$$p_{ij} = \langle m_{ij} \rangle = \frac{Q_{ij} Z_{i+1,j-1} Z_{j+1,N+i-1}}{Z_{1,N}}. \tag{3.31}$$

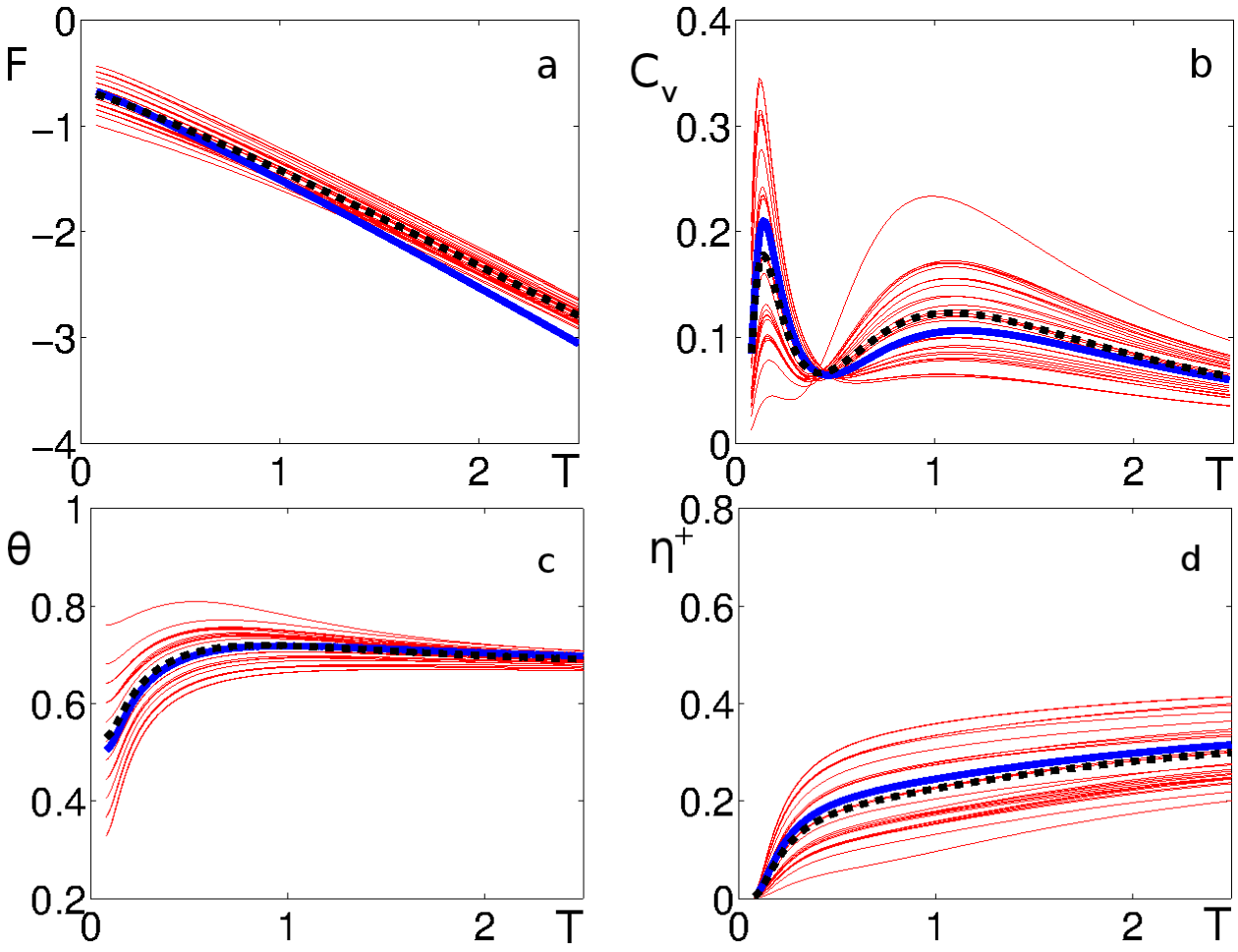

Fig 3.14. Free energy (a), specific heat (b) per nucleotide, helicity degree (c), and the fraction of unfavorable contacts (d) *vs* temperature $T = 1/\beta$. Thin red lines are calculated using McCaskill's algorithm for the 30 random realizations of the $N = 50$ nucleotides with parameters $\epsilon_0 = -1$, $\epsilon = 1.5$ and $q = 0.75$. The thick blue line is calculated in variational approximation $f \approx \max_\mu g(\beta, \mu)$ in the thermodynamic limit $N \to \infty$. The thick dashed black line is the mean value of the quantity, averaged over all random realisations.

Partition functions on the r.h.s. of the Eq. (3.31) have been calculated recursively (1.5) and the helicity degree for the specific realization of sequence of nucleotides is estimated as

$$\theta_{seq.} = \frac{2}{N} \sum_{i<j} p_{ij}. \tag{3.32}$$

In Fig. 3.14c we compare helicity degrees, obtained by the constrained annealing with those directly calculated using Eqs. (1.5,3.31,3.32) for the pool of randomly

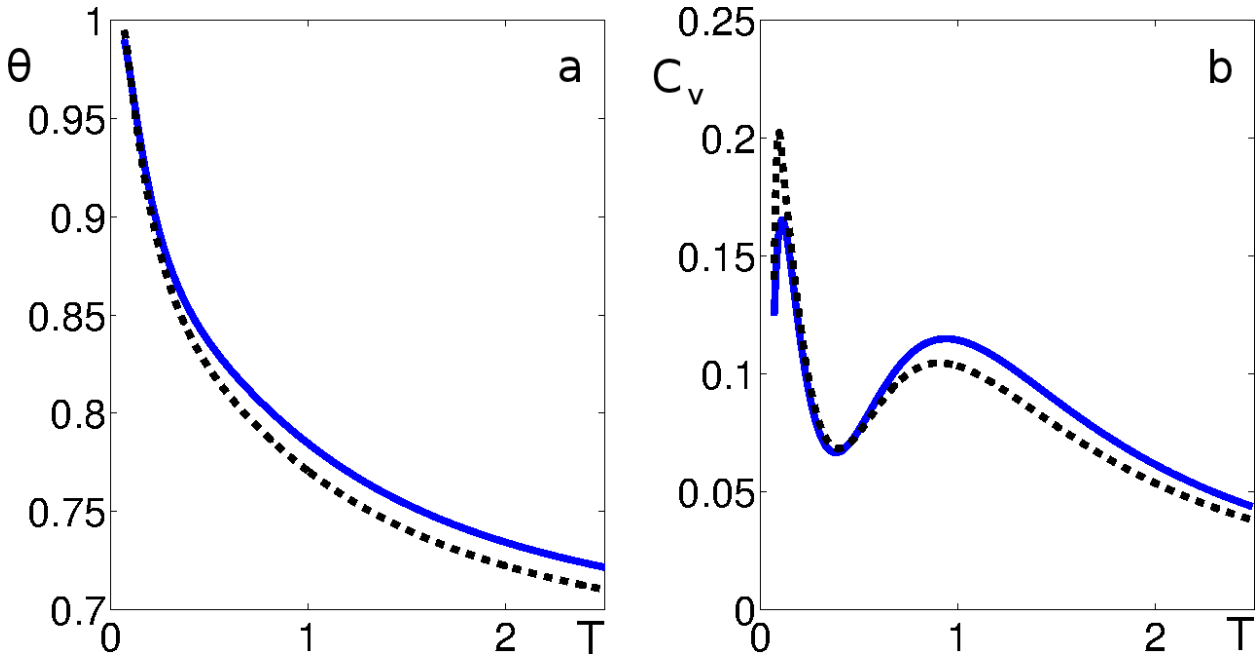

Fig 3.15. Temperature behavior of the helicity degree (a) and the specific heat per nucleotide (b). Black dashed lines are calculated using McCaskill algorithm and averaged over random realizations of the $N = 50$ nucleotides with parameters $\epsilon_0 = -1.5$, $\epsilon = 1.0$ and $q = 0.75$. The blue lines are the quantities, calculated by the constrained annealing method.

generated sequences. The mean value of the helicity degree, calculated numerically is in a good agreement with those, calculated with the help of constrained annealing method. As seen from Fig. 3.14c, helicity degree abruptly increases with temperature and then, after some temperature around $T = 0.5$ point, begins decreasing. Such reentrance of helicity degree indicates the presence of both high- and low-temperature melting and, perhaps denaturation.

The high temperature limit corresponds to the homopolymeric case, where the impact of inter-nucleotide interactions is not so essential. For the sake of simplicity the temperature dependence of the (free) energy of base pair formation is neglected and $\lim_{T\to\infty}\theta = 2/3$. For more realistic choice *e.g.* $\epsilon_0 = \Delta H - T\Delta S$ the high-temperature limit of the helicity degree will be defined mainly by the loss of entropy ΔS of one base pair formation. Here ΔH is the enthalpy per one base pair. When compared against Figs. 3.14b the low-temperature peak of heat capacity could be assigned to low-temperature (cold)

denaturation, while the high-temperature one to the usual hot denaturation.

Helicity degree can be also represented through the fractions of energetically unfavorable (between similar nucleotides) and favorable (between different nucleotides) contacts as $\theta = \eta^+ + \eta^-$, where

$$\eta^{\pm} = \frac{2}{N}\overline{\langle \sum_{i<j} \delta(h_i h_j \mp 1) m_{ij} \rangle}, \tag{3.33}$$

normalized by the maximal number of base pairs, $\frac{N}{2}$. The consideration of temperature dependencies of these quantities reveals the origin of low-temperature melting. The fraction of unfavorable contacts can be written as $\eta^+ = \frac{1}{2}(\theta + \eta)$, where the auxiliary quantity η is written

$$\eta = \frac{2}{N}\overline{\langle \sum_{i<j} h_i h_j m_{ij} \rangle} = 2\frac{\partial g(\beta, \mu_0)}{\partial(\beta\epsilon)}. \tag{3.34}$$

θ and η quantities are calculated analytically (see SI for details). We also estimate η numerically, using the same approach as for helicity degree (see Eqs. (3.31,3.32)). In Fig.3.14d, we compare the fractions of unfavorable contacts η^+, obtained using the Eqs. (3.30,3.34) and those calculated numerically.

In Fig.3.14d, we show the decrease of the fraction of unfavorable contacts with lowered temperature. It is quite natural, since for unfavorable contacts the Boltzmann weight $Q_{ij} < 1$ and tends to zero at low temperature. At the same time, the fraction of favorable contacts η^- increases with the temperature. For the bilateral A and U alphabet, the probability to find the unfavorable pair of nucleotides is higher than to find the favorable one (see SI). That is why the decrease of η^+ results in low temperature melting.

To the best of our knowledge the double-peaked behavior of specific heat has never resulted before. Pagniani *et al* considered equal probabilities ($q = \frac{1}{2}$) for two letters (A and U) to appear and reported a single-peaked specific heat [86]. From our Eqs. (3.28,3.29) it straightforwardly follows that if done so, $\mu_0(\beta) = 0$ and the completely annealed case with a single peak of heat capacity and no low temperature melting results. In compliment to findings of Pagniani

et al [86], our results indicate that single peak of heat capacity results for $q = 0.5$ case only, and for other values of q there are always two peaks.

To address the effect of interaction parameters we distinguish two cases. The first, when the similar nucleotides (AA or UU) are repulsive and the second one, when they are still attractive with less strength than AU. The difference between the two cases is characterised by $\Delta = \epsilon_0 + \epsilon$ parameter. The low-temperature melting, described above (see Fig. 3.14c) takes place if $\Delta > 0$ and the similar nucleotides are repulsive. On the other hand, if the similar nucleotides are attractive ($\Delta < 0$), temperature behavior of helicity degree changes drastically and low-temperature melting disappears (see Fig.3.15a). Specific heat behavior remains the same as for $\Delta > 0$ (see in Fig.3.15b). Given scenario confirms our suggestion, that the reason for low-temperature melting is the decrease of the fraction of energetically unfavorable contacts η^+.

Fig. 3.16 summarizes the obtained results in a *phase* diagram. However, q parameter values belong to the interval $0 \leq q \leq 1$, we consider only $0.5 \leq q \leq 1$, because of system behavior is symmetric with respect to $q = 0.5$. In the upper half of the diagram the temperature behavior of helicity degree is presented, in dependence on the energy of similar nucleotides interactions, $\epsilon_0 + \epsilon$ for the typical value of $q = 0.75$. While the similar nucleotides interaction changes from attraction to repulsion, the system goes from the thermal melting scenario to the both cold and thermal one. In the bottom half of the diagram the temperature behavior of the helicity degree is presented in dependence on the probability q values. In the left-bottom corner the similar nucleotides attraction is addressed and in the right-bottom, the repulsion one. If in case of attraction, the growth of the probablity q just decreases the helicity degree, the similar nucleotides repulsion is characterised by more complicated behavior. While the probability q is growing in the interval $0 \leq q \leq 1$, the helicity degree behavior changes from the purely hot to the purely cold melting. At the intermediate values of q the system exhibits both cold and hot melting.

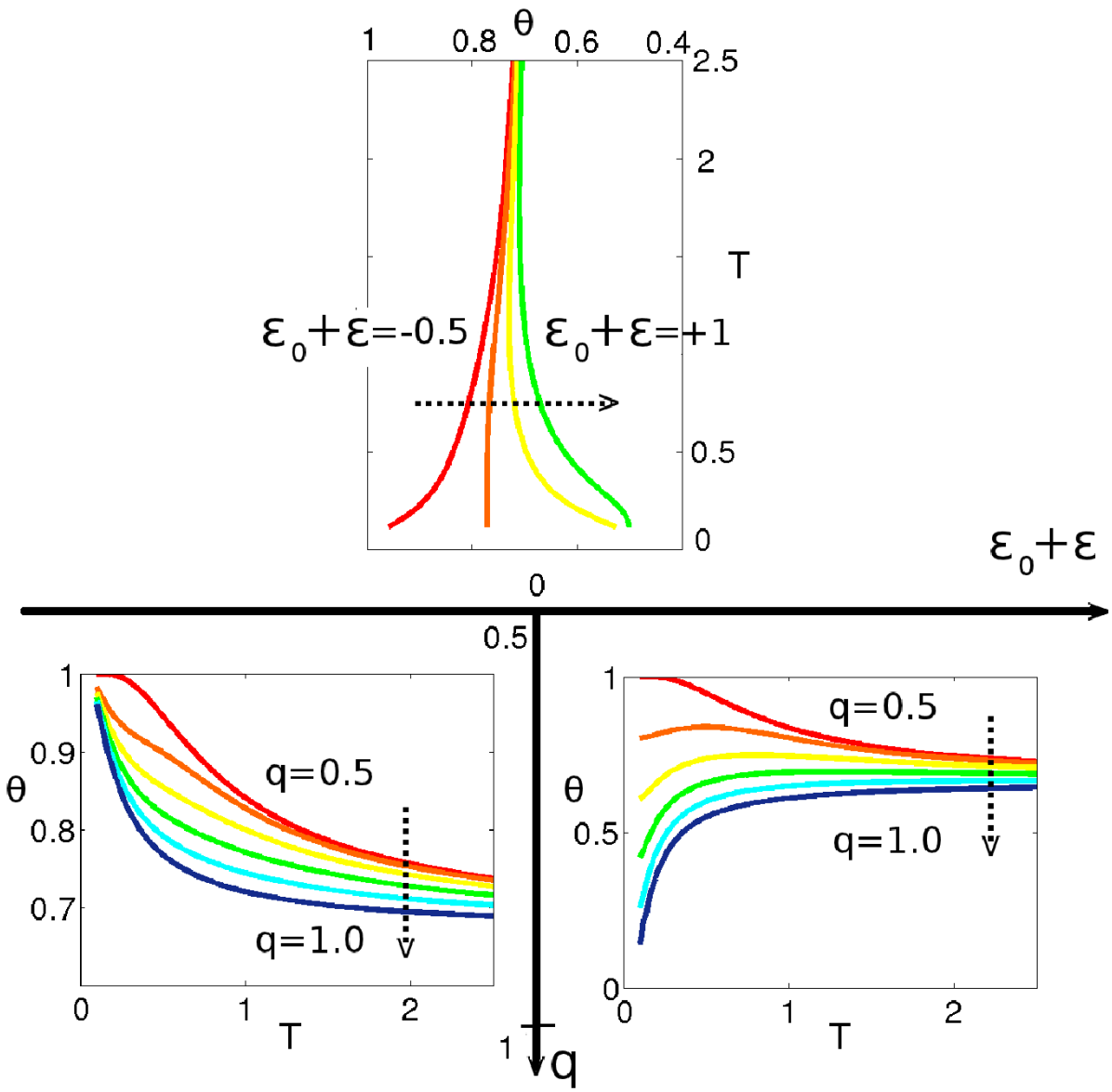

Fig 3.16. Phase diagram $\epsilon_0 + \epsilon$, q.

Temperature behavior of the specific heat is depicted in Fig. 3.17. While in case $q = 0.5$ only high-temperature peak is survived, in (homopolymeric) case $q = 1$ specific heat exhibits only low-temperature peak. In the crossover regime $0.5 < q < 1$ specific heat exhibits two-peak behavior that is corresponding to the hot and cold melting in the right-bottom corner in the Fig. 3.16.

The obtained theoretically cold melting, gives insight into the sequence effect on the cold denaturation [98]. Cold denaturation usually assigned to the positive specific heat difference between the denaturated and native states [98–100] or to the competing between the inter- and intramolecular hydrogen bonds [101]. According to our consideration, the two transitions takes place only if $q \neq$

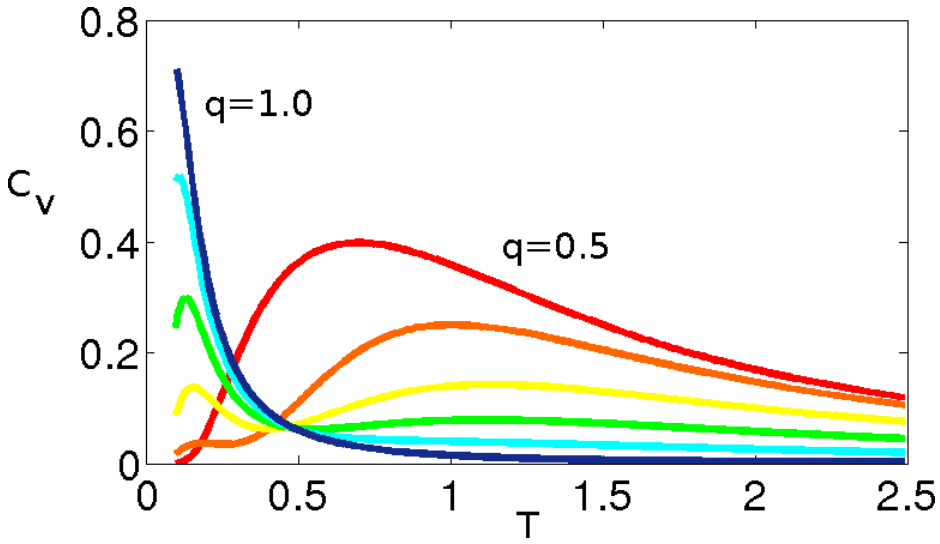

Fig 3.17. Temperature behavior of the specific heat for the different values of q parameter and $\epsilon_0 + \epsilon > 0$ ($\epsilon_0 = -1$, $\epsilon = 1.5$). Value of the probability q is changing from $q = 0.5$ up to $q = 1.0$, while the line color is changing from the red to the blue one.

1/2. The reason is that the probability to find the unfavorable pair of nucleotides is higher than to find the favorable one. These probabilities are equal only if $q = 1/2$. Thus, the potential number of unfavorable contacts seems to be one of the main prerequisites of the cold melting and, perhaps the cold denaturation. At the same time, cold melting requires $\Delta > 0$, where the similar nucleotides are repulsive. That is the free energy change, caused by non-Watson-Crick pairs formation should be positive. Thus, the experimental conditions, suitable for the cold denaturation are based on the interplay between the potential number of unfavorable contacts (sequence) and the non-Watson-Crick pairs stability.

3.5. Effective partition function: Gaussian case.

Let us obtain effective partition function with constraints defined by equations

$$\alpha_1\{h\} = \frac{1}{N}\sum_{i=1}^{N}(h_i - \bar{h}) \tag{3.35}$$

$$\alpha_2\{h\} = \frac{1}{N}\sum_{i=1}^{N}((h_i - \bar{h})^2 - D),$$

where $\bar{h}$ is the mean value of h_i and corresponding distribution function is given by

$$\rho(h_i) = (2\pi D)^{-1/2} e^{-\frac{(h_i-\bar{h})^2}{2D}} \tag{3.36}$$

The effective partition function $\mathcal{Z}_N = \overline{e^{-N\mu_1\alpha_1\{h\}-N\mu_2\alpha_2\{h\}} Z_N(\{h\})}$ is transformed as

$$\mathcal{Z}_N = e^{N(\mu_1\bar{h}+\mu_2 D)} \sum\nolimits'_{\hat{m}} e^{-\beta\epsilon_0 \sum_{i<j} m_{ij}} \int \mathcal{D}h\mathcal{P}\{h\} e^{-\frac{\beta\epsilon}{2}(\mathbf{h},\hat{m}\mathbf{h})-\mu_1(\mathbf{e},\mathbf{h})-\mu_2\sum_j (h_j-\bar{h})^2}, \tag{3.37}$$

where $(\mathbf{a},\mathbf{b})$ is the scalar product of the vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{h} = (h_1, h_2, ..., h_N)$ and $\mathbf{e} = (1, 1, ..., 1)$. Let us average over the distribution function $\mathcal{P}\{h\}$

$$\begin{aligned}
&\int \mathcal{D}h\mathcal{P}\{h\} e^{-\frac{\beta\epsilon}{2}(\mathbf{h},\hat{m}\mathbf{h})-\mu_1(\mathbf{e},\mathbf{h})-\mu_2\sum_j (h_j-\bar{h})^2} = \\
&= \int \mathcal{D}h\mathcal{P}\{h\} \exp\{-\frac{\beta\epsilon}{2}(\mathbf{h},\hat{m}\mathbf{h}) - \mu_1(\mathbf{e},\mathbf{h}) - \\
&-\mu_2(\mathbf{h},\hat{e}\mathbf{h}) + 2\mu_2\bar{h}(\mathbf{e},\mathbf{h}) - \mu_2 N\bar{h}^2\} = \\
&= e^{-N\bar{h}^2(\mu_2+\frac{1}{2D})}(2\pi D)^{-N/2} \int \mathcal{D}h \exp\left\{ -\frac{1}{2}\left(\mathbf{h}, \left[\beta\epsilon\hat{m} + \hat{e}(2\mu_2 + 1/D)\right]\mathbf{h}\right) - \right. \\
&\left. -(\mu_1 - 2\mu_2\bar{h} - \bar{h}/D)(\mathbf{e},\mathbf{h})\right\},
\end{aligned} \tag{3.38}$$

where $\hat{e}$ is the unit matrix. Thus, the effective partition function (3.37) is written

$$\begin{aligned}
&\mathcal{Z}_N = \frac{e^{N(\mu_1\bar{h}+\mu_2 D)-\frac{N\bar{h}^2}{2}(2\mu_2+\frac{1}{D})}}{(2\pi D)^{N/2}} \sum\nolimits'_{\hat{m}} e^{-\beta\epsilon_0\sum_{i<j} m_{ij}} \times \\
&\times \int \mathcal{D}h \exp\left\{ -\frac{1}{2}\left(\mathbf{h}, \left[\beta\epsilon\hat{m} + \hat{e}(2\mu_2 + 1/D)\right]\mathbf{h}\right) - \right. \\
&\left. -(\mu_1 - \bar{h}(2\mu_2 + 1/D))(\mathbf{e},\mathbf{h})\right\} = \\
&= \frac{e^{N(\mu_1\bar{h}+\mu_2 D)-\frac{N\bar{h}^2}{2}(2\mu_2+\frac{1}{D})}}{(2\pi D)^{N/2}} \sum\nolimits'_{\hat{m}} \exp\left\{ -\frac{1}{2}\ln\det\left(\frac{\beta\epsilon\hat{m} + \hat{e}(2\mu_2 + 1/D)}{2\pi}\right) + \right. \\
&\left. +\frac{1}{2}[\mu_1 - \bar{h}(2\mu_2 + 1/D)]^2 \sum_{ij} [\beta\epsilon\hat{m} + \hat{e}(2\mu_2 + 1/D)]^{-1}_{ij}\right\}
\end{aligned} \tag{3.39}$$

Let us calculate first the ln det term in the last equation

$$\begin{aligned} &\ln\det\left(\frac{\beta\epsilon\hat{m}+\hat{e}(2\mu_2+1/D)}{2\pi}\right)= \\ &= Tr\ln\left(\frac{\beta\epsilon\hat{m}+\hat{e}(2\mu_2+1/D)}{2\pi}\right)= \\ &= -N\ln(2\pi\tilde{D})+\sum_{k=1}^{\infty}\frac{(-1)^k}{k}(\beta\epsilon\tilde{D})^k Tr(\hat{m}^k), \end{aligned} \tag{3.40}$$

where $\tilde{D}=\frac{D}{1+2D\mu_2}$. Let us calculate a few first terms $Tr(\hat{m}^k)$.

- $(k=1)\ Tr(\hat{m})=0$
- $(k=2)\ Tr(\hat{m}^2)=\sum_i\sum_j m_{ij}m_{ji}=\sum_{ij}m_{ij}$
- $(k=3)$

  $$Tr(\hat{m}^3)=\sum_i\sum_{ij}m_{ij}m_{jk}m_{ki}=\sum_i\sum_{jk}m_{ij}m_{jk}m_{ki}\delta_{ij}\delta_{ik},$$

  because of e.g. $m_{ij}m_{jk}\neq 0$ only if $i=k$. Thus,

  $$Tr(\hat{m}^3)=\sum_i\sum_{jk}m_{ij}m_{jk}m_{ki}\delta_{ij}\delta_{ik}=\sum_{ij}m_{ij}m_{ji}m_{ii}\delta_{ij}=0$$

- $(k=4)$

  $$Tr(\hat{m}^4)=\sum_i\sum_{jkl}m_{ij}m_{jk}m_{kl}m_{li}=\sum_i\sum_{jkl}m_{ij}m_{jk}m_{kl}m_{li}\delta_{ik}\delta_{jl}\delta_{ki}=\sum_{ij}m_{ij}$$

- $(k=5)$ in the same manner we can show that $Tr(\hat{m}^5)=0$ etc.

Thus,

$$Tr(\hat{m}^k)=\begin{cases}\sum_{ij}m_{ij} & \text{, if k is even}\\ 0 & \text{, if k is odd}\end{cases} \tag{3.41}$$

and eq.(3.40) is written as

$$\begin{aligned} &\ln\det\left(\frac{\beta\epsilon\hat{m}+\hat{e}(2\mu_2+1/D)}{2\pi}\right)= \\ &= -N\ln(2\pi\tilde{D})+\sum_{ij}m_{ij}\sum_{m=1}^{\infty}\frac{(-1)^{2m+1}}{2m}(\beta\epsilon\tilde{D})^{2m}= \\ &= -N\ln(2\pi\tilde{D})+\frac{1}{2}\ln[1-(\beta\epsilon\tilde{D})^2] \end{aligned} \tag{3.42}$$

The sum of the elements of inverse matrix in the exponent of eq.(3.39) is written

$$\sum_{ij}[\beta\epsilon\hat{m}+\hat{e}(2\mu_2+1/D)]^{-1}_{ij}=\tilde{D}\sum_{ij}[\beta\epsilon\tilde{D}\hat{m}+\hat{e}]^{-1}_{ij} \tag{3.43}$$

The inverse matrix is expanded as $[\beta\epsilon\tilde{D}\hat{m}+\hat{e}]^{-1}=\sum_{l=0}^{\infty}(-1)^l(\beta\epsilon\tilde{D})^l\hat{m}^l$. Thus, eq.(3.43) is rewritten as

$$\sum_{ij}[\beta\epsilon\hat{m}+\hat{e}(2\mu_2+1/D)]^{-1}_{ij}=\tilde{D}\sum_{l=0}^{\infty}(-1)^l(\beta\epsilon\tilde{D})^l\sum_{ij}(\hat{m}^l)_{ij} \tag{3.44}$$

In analogy with $Tr(\hat{m}^l)$ we can show that

$$\sum_{ij}(\hat{m}^l)ij=\begin{cases}\sum_{ij}m_{ij} & \text{, if l=1,2,3,...}\\ N & \text{, if l=0}\end{cases} \tag{3.45}$$

Thus,

$$\sum_{ij}[\beta\epsilon\hat{m}+\hat{e}(2\mu_2+1/D)]^{-1}_{ij}=N\tilde{D}-\frac{\beta\epsilon\tilde{D}^2}{1+\beta\epsilon\tilde{D}} \tag{3.46}$$

Finally, the effective partition function is written

$$\mathcal{Z}_N=\left(\frac{\tilde{D}}{D}\right)^{N/2}\exp\left[N\mu_1\bar{h}+N\mu_2D+\frac{N\tilde{D}}{2}\mu_1(\mu_1-2\bar{h}/\tilde{D})\right]\times \tag{3.47}$$
$$\times\sum\nolimits'_{\hat{m}}e^{-\beta(\epsilon_0+\tilde{\epsilon})\sum_{i<j}m_{ij}},$$

where

$$\beta\tilde{\epsilon}=\frac{1}{2}\ln[1-(\beta\epsilon\tilde{D})^2]+(\mu_1-\bar{h}/\tilde{D})^2\frac{\beta\epsilon\tilde{D}^2}{1+\beta\epsilon\tilde{D}} \tag{3.48}$$

3.6. Variational equation.

Let us address the case, where $\bar{h}=0$. With taking into account notations $\tilde{D}=D/(1+2D\mu_2)$ and $\mu_1=\mu$ variational reduced free energy is written

$$g(\beta,\mu,\tilde{D})=-\frac{1}{2}\ln\frac{\tilde{D}}{D}-\frac{D}{2}+\frac{1}{2}-\frac{\tilde{D}\mu^2}{2}- \tag{3.49}$$
$$-\ln\left(1+2[1-(\beta\epsilon\tilde{D})^2]^{-1/4}e^{-\frac{\beta\epsilon_0}{2}-\frac{\mu^2}{2}\frac{\beta\epsilon\tilde{D}^2}{1+\beta\epsilon D}}\right)$$

Let us find the point of extrema over the μ variable

$$0 = \frac{\partial g}{\partial \mu} = -\tilde{D}\mu + \mu \frac{\beta\epsilon\tilde{D}^2}{1+\beta\epsilon\tilde{D}} S(X), \tag{3.50}$$

where $X = \ln 2 - \frac{\beta\epsilon_0}{2} - \frac{\mu^2}{2}\frac{\beta\epsilon\tilde{D}^2}{1+\beta\epsilon\tilde{D}} - \frac{1}{4}\ln[1-(\beta\epsilon\tilde{D})^2]$ and $S(x) = \frac{e^x}{1+e^x}$ is so called logical function. Eq.(3.50) has two solutions, $\mu = 0$ and

$$1 = \frac{\beta\epsilon\tilde{D}}{1+\beta\epsilon\tilde{D}} S(X) \tag{3.51}$$

Because of $0 < S(x) < 1$ at any finite value of x, the right side of the last equation is less than 1. That is why eq.(3.51) has no solution and the unique solution of the eq.(3.50) is $\mu = 0$.

3.7. Entropy: low-temperature limit.

To estimate entropy in the low-temperature area let us estimate value of $\tilde{D}$, maximizing variational free energy $g_0(\beta, \tilde{D}) = g(\beta, 0, \tilde{D})$, where $g(\beta, 0, \tilde{D})$ is defined by eq.(3.49). From the equation $0 = \frac{\partial g_0}{\partial \tilde{D}}$ it follows that

$$\frac{D}{\tilde{D}} = 1 + \frac{(\beta\epsilon\tilde{D})^2}{1-(\beta\epsilon\tilde{D})^2} S(X), \tag{3.52}$$

where $X = \ln 2 - \frac{\beta\epsilon_0}{2} - \frac{\mu^2}{2}\frac{\beta\epsilon\tilde{D}^2}{1+\beta\epsilon\tilde{D}} - \frac{1}{4}\ln[1-(\beta\epsilon\tilde{D})^2]$. Let us suppose, $\beta\epsilon\tilde{D} \to 1$ ($\beta \to \infty$) as $(\beta\epsilon\tilde{D})^2 = 1 - r$, where $r \ll 1$. Thus, eq.(3.52) is written

$$\beta\epsilon(1 + r/2) \approx 1 + \left(\frac{1}{r} - 1\right)\left[1 - \frac{r^{1/4}}{2} e^{\beta\epsilon_0/2} + \mathcal{O}(e^{\beta\epsilon_0}\sqrt{r})\right] \tag{3.53}$$

We are focused on the case $\epsilon_0 < 0$. The last equation can be expanded up to the linear term over r and written as

$$\beta\epsilon D r \simeq 1 - \frac{r^{1/4} e^{\beta\epsilon_0/2}}{2} + \mathcal{O}(e^{\beta\epsilon_0}\sqrt{r}) \tag{3.54}$$

Thus, in the limit $\beta \to \infty$

$$r \simeq \frac{1}{\beta\epsilon D} \tag{3.55}$$

In the low-temperature area

$$g_0(\beta,\tilde{D}) \simeq \frac{1}{4}\ln(\beta\epsilon D) + \frac{1}{4\beta\epsilon D} + \frac{1}{4} - \frac{\beta\epsilon D}{2} - \ln 2 + \frac{\beta\epsilon_0}{2} \tag{3.56}$$

and entropy $s = -g_0 + \beta\frac{dg_0}{d\beta}$ in the limit $\beta \to \infty$ is estimated as

$$s_0 \simeq -\frac{1}{4}\ln(\beta\epsilon D) \tag{3.57}$$

3.8. Effective partition function: bimodal disorder.

Let us obtain effective partition function with constraints defined by equations

$$a_1\{h\} = \frac{1}{N}\sum_{i=1}^{N}[h_i - (2q-1)] \tag{3.58}$$

$$a_2\{h\} = \frac{1}{N}\sum_{i=1}^{N}[(h_i - (2q-1))^2 - 4q(1-q)],$$

where the corresponding distribution function is given by

$$P(h_i) = q\delta(h_i - 1) + (1-q)\delta(h_i + 1) \tag{3.59}$$

The effective partition function $\mathcal{Z}_N = \overline{e^{-N\mu_1 a_1\{h\} - N\mu_2 a_2\{h\}} Z_N(\{h\})}$ is transformed as

$$\mathcal{Z}_N = e^{N(2q-1)\mu}\sum_{\{h\}}\prod_{j=1}^{N} P(h_j)e^{-\mu h_j} Z_N(\{h\}), \tag{3.60}$$

where $\mu = \mu_1 - 2(2q-1)\mu_2$. Let us calculate separately

$$\sum_{\{h\}}\prod_{j=1}^{N} P(h_j)e^{-\mu h_j} Z_N(\{h\}) = \tag{3.61}$$

$$= \sum_{\hat{m}}{}' e^{-\beta\epsilon_0\sum_{i<j} m_{ij}} \sum_{\{h\}}\prod_{j=1}^{N} P(h_j)e^{-\mu h_j}\prod_{k<l}(1 + m_{kl}V_{kl}),$$

where $V_{kl} = e^{-\beta\epsilon h_k h_l} - 1$. Thus,

$$\sum_{\{h\}}\prod_{j=1}^{N} P(h_j)e^{-\mu h_j} Z_N(\{h\}) = \sum_{\hat{m}}{}' e^{-\beta\epsilon_0\sum_{i<j} m_{ij}} \sum_{\{h\}}\prod_{j=1}^{N} P(h_j)e^{-\mu h_j} \times \tag{3.62}$$

$$\times\left\{1 + \sum_{k<l} m_{kl}V_{kl} + \sum_{k<l}\sum_{p<q} m_{kl}m_{pq}V_{kl}V_{pq} + \sum_{k<l}\sum_{p<q}\sum_{t<n} m_{kl}m_{pq}m_{tn}V_{kl}V_{pq}V_{tn} + ...\right\},$$

Summation in the last equation is taken over p non-repeating pairs of nucleotides $i_\alpha < j_\alpha$, where $p = 1, 2, 3, ...$ At the same time each of pairs differs from other pairs at least by one nucleotide. Thus, each sum of p pairs of nucleotides can be divided by two parts. The first one, containing no common nucleotides and the second one, containing at least one common nucleotide. Let us address the sum over p non-repeating base pairs containing at least one common nucleotide, e.g. J

$$\sum_{i_1<j_1} ... \sum_{i_k} ... \sum_{j_l} ... \sum_{i_p<j_p} m_{i_1j_1}...m_{i_kJ}...m_{Jj_l}...m_{i_pj_p}V_{i_1j_1}...V_{i_kJ}...V_{Jj_l}...V_{i_pj_p} = \tag{3.63}$$
$$= \sum_{i_1<j_1} ... \sum_{i_k} ... \sum_{j_l} ... \sum_{i_p<j_p} m_{i_1j_1}...m_{i_kJ}...\delta_{i_kj_l}...m_{Jj_l}...m_{i_pj_p}V_{i_1j_1}...V_{i_kJ}...V_{Jj_l}...V_{i_pj_p}$$

because of $m_{i_kJ}m_{Jj_l} \neq 0$ only if $i_k = j_l$. At the same time $\delta_{i_kj_l}$ in eq.(3.63) is always equal to zero because of each pair in the sum (3.63) differs from other pairs at least by one nucleotide and consequently $i_k \neq j_l$. Thus,

$$\sum_{i_1<j_1} ... \sum_{i_p<j_p} m_{i_1j_1}...m_{i_pj_p}V_{i_1j_1}...V_{i_pj_p} = \sum_{(i_1<j_1)} ... \sum_{(i_p<j_p)} m_{i_1j_1}...m_{i_pj_p}V_{i_1j_1}...V_{i_pj_p} \tag{3.64}$$

is taken over p non-repeated pairs of nucleotides without any common nucleotide. Thus, we can average factors $V_{i_1j_1}$,..., $V_{i_pj_p}$ in the eq.(3.62) independently and

$$\sum_{\{h\}} \prod_{j=1}^{N} P(h_j)e^{-\mu h_j}V_{i_1j_1}...V_{i_pj_p} = \tag{3.65}$$
$$= \left(\sum_{h=\pm 1} P(h)e^{-\mu h} \right)^{N-2p} \left(\sum_{h,h'=\pm 1} P(h)P(h')e^{-\mu(h+h')}[e^{-\beta\epsilon hh'} - 1] \right)^p =$$
$$= \Omega^{N-2p} \left(-\Omega^2 + e^{-\beta\epsilon}[q^2e^{-2\mu} + (1-q)^2e^{2\mu}] + 2q(1-q)e^{\beta\epsilon} \right)^p = \Omega^N \bar{V}^p,$$

where

$$\Omega(\mu) = qe^{-\mu} + (1-q)e^{\mu} \tag{3.66}$$
$$W(\mu, \beta, \epsilon) = e^{-\beta\epsilon}[q^2e^{-2\mu} + (1-q)^2e^{2\mu}] + 2q(1-q)e^{\beta\epsilon}$$
$$\bar{V} = \frac{W(\mu, \beta, \epsilon)}{\Omega(\mu)^2} - 1$$

Finally obtaining from the eqs.(3.60,3.62 and 3.65)

$$\mathcal{Z}_N = e^{N(2q-1)\mu}\Omega^N \sum_{\hat{m}}{}' e^{-\beta(\epsilon_0+\bar{\epsilon})\sum_{i<j} m_{ij}}, \tag{3.67}$$

where $e^{-\beta\bar{\epsilon}} = \bar{V} + 1$.

3.9. Probabilities.

Energetically unfavorable contacts are $++$ and $--$. In assumption of statistical independence of the h_i variables the probability to find the unfavorable pair of nucleotides is

$$\mathcal{P}_{unfav.} = q^2 + (1-q)^2 = 1 - 2q(1-q) \tag{3.68}$$

and the probability to find the favorable one is written

$$\mathcal{P}_{fav.} = 2q(1-q) \tag{3.69}$$

Thus, always

$$\mathcal{P}_{unfav.} > \mathcal{P}_{fav.}, \tag{3.70}$$

if $q \neq \frac{1}{2}$ and $\mathcal{P}_{unfav.} = \mathcal{P}_{fav.}$ if $q = \frac{1}{2}$ (see Figure 3.18).

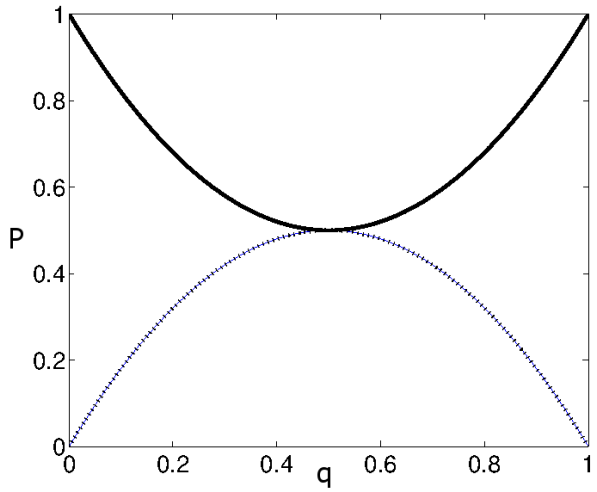

Fig 3.18. Probabilities of favorable and unfavorable pairs of nucleotides vs. q are given by dashed and solid lines correspondingly.

Chapter 4

The secondary structural transitions in single-stranded RNA. Account of the loop formation.

4.1. Statement of the problem

In this chapter we develop the model of random ssRNA by taking into consideration loop formation. As has became common in theory,conformational weight ascribed to a loop with length m is m^{-c}. Thus, entropic impact of such loop will be $-c \ln m$. The role of loop factor c is critical in secondary structural phase transitions, particularly, in thermal-induced phase transitions. Loop structures such as hairpin loops,internal loops, multi-loops with three or more emerging loops, which are common for ssRNA are characterized with the value of $c \approx 2.1$. However, according to [74], mentioned phase transitions occur in the specific range of loop exponent $2 < c < 2.479$. Our aim is to modify the model introduced in the previous chapter, and study the dependence of thermodynamic parameters, such as specific heat and helicity degree, on the value of loop exponent c. We concentrate our attention on the case of bimodal disorder, since it is relatively simple, but, at the same time, it displays good description of the phenomena. However, as a rule, a sequence of a ssRNA is not completely random, usually it is optimized for distinct native structure. Nevertheless, to understand the role of such optimization we have to study thermodynamics of ssRNA with bimodal disorder.

We exploit the partition function of homopolymeric ssRNA in the presence of loops [74, 102] by combining it with the modified statistical weight $Q = e^{-\beta(\epsilon_0+\bar{\epsilon})}$.

4.2. The model

For the sake of simplicity we propose to study random ssRNA sequences composed only of A and U bases. The topological rules that determine allowed structures are essential for efficient numerical calculation of the free energy of secondary structures. The main rule is the elimination of so-called pseudoknots from the set of available secondary structures as in most other works on ssRNA. Thus, for any two base pairs (i, j) and (k, l) with $i < j$, $k < l$, and $i < k$ we have either $i < k < l < j$ or $i < j < k < l$ [103]. Besides, a valid secondary structure is a list of all base pairs with the constraint that a base can be part of at most one pair.

Hamiltonian of the model reads

$$\mathcal{H}(\hat{m}, \{h_i\}) = \sum_{i<j} m_{ij}\epsilon_{ij}, \tag{4.1}$$

where the interaction constants $\epsilon_{ij} = \epsilon_0 + \epsilon h_i h_j$, sum is taken over all non-repeated base pairs, $m_{ij} = 1$ if the bases i and j are paired and $m_{ij} = 0$ otherwise. Variables $\{h_i\}$ describe the type of nucleotide and $h_i = \pm 1$, where $h_i = +1$ corresponds to A, and $h_i = -1$ to U. The partition function for the ssRNA chain of N nucleotides is written as

$$Z_N(\{h_i\}) = \sum\nolimits'_{\hat{m}} \exp[-\beta\mathcal{H}(\hat{m}, \{h_i\})], \tag{4.2}$$

where $\beta = \frac{1}{k_B T}$ and the sum is taken over all realizations without pseudoknots of the matrix $\hat{m}$, which contains no more than one unity on each row or column. The latter condition describes the saturation of base pairing. The $\{h_i\}$ sequence is supposed to be randomly generated with the distribution function

$$\mathcal{P}\{h\} = \prod_{i=1}^{N} \rho(h_i), \tag{4.3}$$

where $\rho(h_i) = q\delta(h_i - 1) + (1 - q)\delta(h_i + 1)$.

The reduced free energy for the given $\{h_i\}$ sequence of nucleotides is written as $f\{h_i\} = -\frac{1}{N} \ln Z_N(\{h_i\})$. Due to self-averaging, the free energy

in the thermodynamic limit $N \to \infty$ becomes a non-random quantity and

$$f\{h_i\} = f = -\frac{1}{N}\overline{\ln Z_N(\{h\})}, \tag{4.4}$$

where f is the reduced quenched free energy and $\overline{\mathcal{O}}$ means the average over sequence distribution function (4.3). According to [87], the free energy of the ssRNA with random quenched sequence of nucleotides satisfies the conditions

$$f \geq g(\beta, \mu) \geq f_a, \tag{4.5}$$

where f_a is the reduced annealed free energy and

$$g(\beta, \mu) = -\frac{1}{N}\ln \mathcal{Z}_N = -\frac{1}{N}\ln \overline{Z_N(\{h_i\})e^{-N\mu\alpha(\{h_i\})}}. \tag{4.6}$$

Here $\alpha(\{h_i\})$ is the appropriate self-averaging sequence-dependent quantity. Thus, $g(\beta, \mu)$ gives the lower bound of the quenched free energy f. According to [87], the best lower bound of the quenched free energy is given by $\max_\mu g(\beta, \mu)$ and we can estimate the free energy of the ssRNA molecule with randomly generated sequence as

$$f \approx \max_\mu g(\beta, \mu). \tag{4.7}$$

The simplest constraint imposed on the quenched variables $\{h_i\}$ is given by $\alpha(\{h_i\}) = \frac{1}{N}\sum_{i=1}^{N}[h_i - (2q-1)]$, which does not fix the types of individual monomers h_i, but just the mean value of the sum $\sum_i h_i$. After some algebra (see for details SI) the effective partition function $\mathcal{Z}_N$, defined in Eq. (4.6) reads

$$\mathcal{Z}_N = e^{N\mu(2q-1)}\Omega^N \mathcal{Z}_N^0(\epsilon_0 + \bar{\epsilon}), \tag{4.8}$$

where $\mathcal{Z}_N^0(\epsilon_0 + \bar{\epsilon})$ is the partition function (4.2) of the homopolymeric ssRNA with the effective interaction constant $\epsilon_{ij} = \epsilon_0 + \bar{\epsilon}$. Here

$$\begin{aligned} &\bar{\epsilon} = -\frac{1}{\beta}\ln\frac{W(\mu, \beta, \epsilon)}{\Omega(\mu)^2}, \\ &\Omega(\mu) = qe^{-\mu} + (1-q)e^{\mu}, \\ &W(\mu, \beta, \epsilon) = e^{-\beta\epsilon}[q^2e^{-2\mu} + (1-q)^2e^{2\mu}] + 2q(1-q)e^{\beta\epsilon}. \end{aligned} \tag{4.9}$$

4.3. Calculation of thermodynamic characteristics of the model with loops

As it was shown above, the variational reduced free energy $g(\mu,\beta) = -\frac{1}{N}\ln \mathcal{Z}_N$ could be written as

$$g(\beta,\mu) = -\mu(2q-1) - \ln\Omega(\mu) - \frac{1}{N}\ln \mathcal{Z}_N^0(Q(\mu,\beta)) \tag{4.10}$$

According to [74, 102]

$$\mathcal{Z}_N^0 \simeq z_d^{-N}, \tag{4.11}$$

where z_d is the dominant singularity of grand canonical partition function of homopolymeric ssRNA, which is defined as the singularity which is nearest to the origin in the complex z-plane. In particular, our grand partition function has two singularities: a branching point and a pole. Depending on the value of statistical weight Q we have

$$z_d = \begin{cases} z_p,\ Q < Q_c \\ z_b,\ Q > Q_c \end{cases}. \tag{4.12}$$

Here Q_c is critical value of Q:

$$Q_c = \frac{Li_{c-1}(1) - Li_c(1)}{\left(Li_{c-1}(1) - 2Li_c(1)\right)^2} \tag{4.13}$$

where the polylogarithm $Li_c(x) = \sum_{n=1}^{\infty} x^n/n^c$ is used. To obtain branching point z_b, we have to solve following system of equations

$$\begin{cases} \kappa\left(\kappa - 1\right) = QLi_c\left(z_b\kappa\right) \\ \kappa^2 = Q\left(Li_{c-1}\left(z_b\kappa\right) - Li_c\left(z_b\kappa\right)\right) \end{cases} \tag{4.14}$$

Whereas,

$$z_p = \frac{2}{1+\sqrt{1+4QLi_c(1)}} \tag{4.15}$$

From (4.14) by dividing equations we obtain

$$1 - \frac{1}{\kappa} = \frac{Li_c\left(z_b\kappa\right)}{Li_{c-1}\left(z_b\kappa\right) - Li_c\left(z_b\kappa\right)} \tag{4.16}$$

So,

$$\kappa = \frac{Li_{c-1}\left(z_b\kappa\right) - Li_c\left(z_b\kappa\right)}{Li_{c-1}\left(z_b\kappa\right) - 2Li_c\left(z_b\kappa\right)} \tag{4.17}$$

Here we introduce variable $t = z_b\kappa$. By substituting $\kappa = \frac{t}{z_b}$ in (4.17) we arrive to

$$\begin{cases} \frac{t}{z_b} = \frac{Li_{c-1}(t) - Li_c(t)}{Li_{c-1}(t) - 2Li_c(t)} \\ \frac{t^2}{z_b^2} = Q\left(Li_{c-1}\left(t\right) - Li_c\left(t\right)\right) \end{cases} \tag{4.18}$$

By modifying (4.18) we obtain

$$Li_{c-1}\left(t\right) - Li_c\left(t\right) = Q\left(Li_{c-1}\left(t\right) - 2Li_c\left(t\right)\right)^2 \tag{4.19}$$

and

$$z_b = t\frac{Li_{c-1}\left(t\right) - 2Li_c\left(t\right)}{Li_{c-1}\left(t\right) - Li_c\left(t\right)} \tag{4.20}$$

This transformation allows us to solve (4.14) numerically. First, we derive t from (4.19), and thereafter, by substituting it in (4.20), we obtain z_b. Now, let's return to calculation of free energy.According to constrained annealing approach we have to maximize free energy by solving

$$\frac{\partial g(\mu, \beta)}{\partial \mu} = 0 \tag{4.21}$$

for μ. Hereupon, we can write free energy as $g(\beta) = g(\mu_0, \beta)$, where μ_0 is solution of (4.21). The free energy per monomer of the system is estimated as

$$f(\beta) = \frac{g(\beta)}{\beta}, \tag{4.22}$$

The entropy per monomer is written as

$$s(\beta) = -g(\beta) + \beta\frac{dg(\beta)}{d\beta} \tag{4.23}$$

and the specific heat

$$c_V(\beta) = -\beta^2\frac{d^2g(\beta)}{d\beta^2}. \tag{4.24}$$

To calculate (4.23) and (4.36), we have to consider two cases (4.12). Let's first implement derivation of (4.23) for the cases $z_d = z_p$ when $Q < Q_c$.

$$\frac{\partial g(\beta)}{\partial \beta} = \frac{\partial}{\partial \beta} \ln z_p(Q(\mu_0, \beta)) = \frac{1}{z_p(Q)} \left(\frac{\partial z_p}{\partial \beta} \right) = \frac{1}{z_p(Q)} \left(\frac{\partial z_p}{\partial Q} \right) \left(\frac{\partial Q}{\partial \beta} \right) \quad (4.25)$$

where in accordance with (4.15) and (3.27)

$$\frac{\partial z_p}{\partial Q} = -\frac{4Li_c(1)}{\left(1 + \sqrt{1 + 4QLi_c(1)}\right)^2 \sqrt{1 + 4QLi_c(1)}}, \quad (4.26)$$

$$\frac{\partial Q}{\partial \beta} = \frac{\partial e^{-\beta(\epsilon_0 + \bar{\epsilon})}}{\partial \beta} = -\epsilon_0 Q + \frac{e^{-\beta\epsilon_0}}{\Omega^2} \left(\frac{\partial W}{\partial \beta} \right), \quad (4.27)$$

and

$$\frac{\partial Q}{\partial \beta} = -\epsilon e^{-\beta\epsilon} \left(q^2 e^{-2\mu} + (1-q)^2 e^{2\mu} \right) + 2\epsilon q(1-q) e^{\beta\epsilon}. \quad (4.28)$$

One can substitute (4.26) and (4.27) in (4.25) and obtain the final expression for derivative of the reduced free energy when $Q < Q_c$:

$$\frac{\partial g(\beta)}{\partial \beta} = -\frac{2Li_c(1)}{\left(1 + 4QLi_c(1) + \sqrt{1 + 4QLi_c(1)}\right)} \left(-\epsilon_0 Q + \frac{e^{-\beta\epsilon_0}}{\Omega^2} \left(\frac{\partial W}{\partial \beta} \right) \right). \quad (4.29)$$

where

$$\frac{\partial W}{\partial \beta} = -\epsilon e^{-\beta\epsilon} \left(q^2 e^{-2\mu} + (1-q)^2 e^{2\mu} \right) + 2\epsilon q(1-q) e^{\beta\epsilon}. \quad (4.30)$$

Therefore, the expression for entropy (when $Q < Q_c$) can be written as

$$s(\beta) = -\mu(2q-1) - \ln \Omega(\mu) - \ln \frac{2}{1 + \sqrt{1 + 4QLi_c(1)}} - \quad (4.31)$$
$$-\beta \frac{2Li_c(1)}{\left(1 + 4QLi_c(1) + \sqrt{1 + 4QLi_c(1)}\right)} \left(-\epsilon_0 Q + \frac{e^{-\beta\epsilon_0}}{\Omega^2} \left(\frac{\partial W}{\partial \beta} \right) \right).$$

To represent entropy for the case when $Q > Q_c$, we have to calculate $\frac{\partial z_b}{\partial Q}$. Since being unable to do it explicitly, we perform following transformation:

$$\frac{\partial z_b}{\partial Q} = \frac{\left(\frac{\partial z_b}{\partial t} \right)}{\left(\frac{\partial Q}{\partial t} \right)}. \quad (4.32)$$

Taking into account (4.19) and (4.20) we obtain

$$\frac{\partial z_b}{\partial Q} = \frac{tLi_c(t)\left(Li_{c-1}(t) - 2Li_c(t)\right)^3}{Li_{c-1}(t)\left(Li_{c-1}(t) - Li_c(t)\right)^2}. \tag{4.33}$$

One can combine (4.20), (4.27) and (4.33) in $\frac{\partial g(\beta)}{\partial \beta} = \frac{1}{z_b(Q)}\left(\frac{\partial z_b}{\partial Q}\right)\left(\frac{\partial Q}{\partial \beta}\right)$ and obtain

$$\frac{\partial g(\beta)}{\partial \beta} = \frac{Li_c(t)\left(Li_{c-1}(t) - 2Li_c(t)\right)^2}{Li_{c-1}(t)\left(Li_{c-1}(t) - Li_c(t)\right)}\left(-\epsilon_0 Q + \frac{e^{-\beta\epsilon_0}}{\Omega^2}\left(\frac{\partial W}{\partial \beta}\right)\right). \tag{4.34}$$

Hence,

$$s(\beta) = \mu(2q-1) - \ln\Omega(\mu) - \ln\frac{2}{1+\sqrt{1+4QLi_c(1)}} + \tag{4.35}$$
$$+\beta\frac{Li_c(t)\left(Li_{c-1}(t) - 2Li_c(t)\right)^2}{Li_{c-1}(t)\left(Li_{c-1}(t) - Li_c(t)\right)}\left(-\epsilon_0 Q + \frac{e^{-\beta\epsilon_0}}{\Omega^2}\left(\frac{\partial W}{\partial \beta}\right)\right)$$

Explicit analytic expressions for specific heat are more complicated, however, they could be obtained from general analytic expression

$$c_V(\beta) = -\beta\frac{ds}{d\beta} \tag{4.36}$$

Now let's refer to another important thermodynamic characteristic,namely, to helicity degree:

$$\theta = \frac{2}{N}\frac{d\ln\mathcal{Z}}{d\ln Q} \tag{4.37}$$

where $N \to \infty$. According to (4.11)

$$\theta = -2Q\frac{dz_d}{dQ} \tag{4.38}$$

4.4. Results and Discussion

Free energy in dependence on temperature is presented in Figs. 4.1,4.2. Just as in the Chapter 3, we distinguish two cases, according to the energy of interaction between similar nucleotides $\epsilon_0 + \epsilon$. There is no qualitative difference between the temperature behavior of the free energy in the cases of repulsion and attraction, characterized by $\epsilon_0 + \epsilon > 0$ and $\epsilon_0 + \epsilon < 0$ correspondingly.

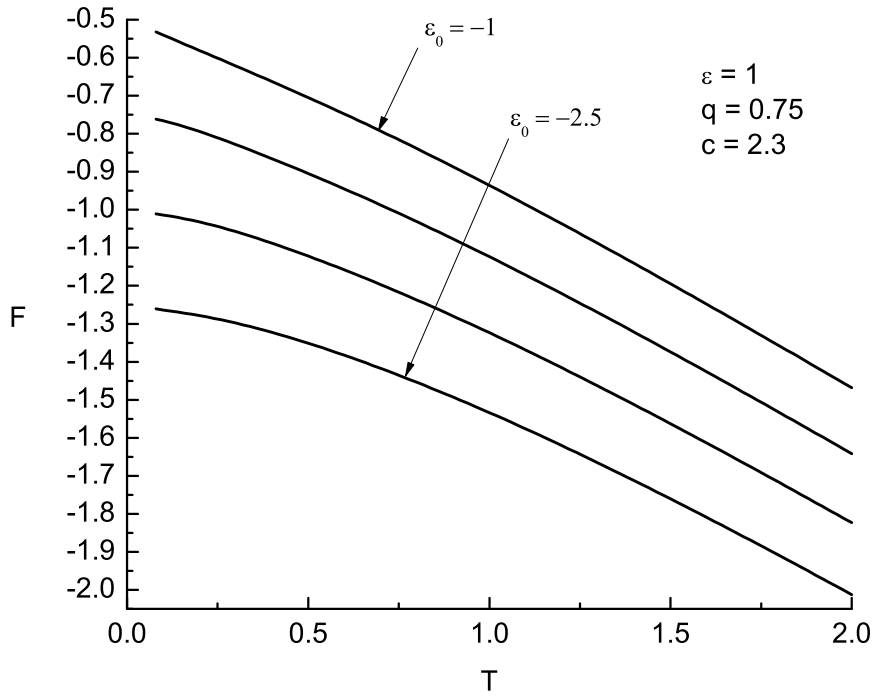

Fig 4.1. Dependence of reduced free energy on temperature , where $\epsilon_0 + \epsilon \leq 0$.

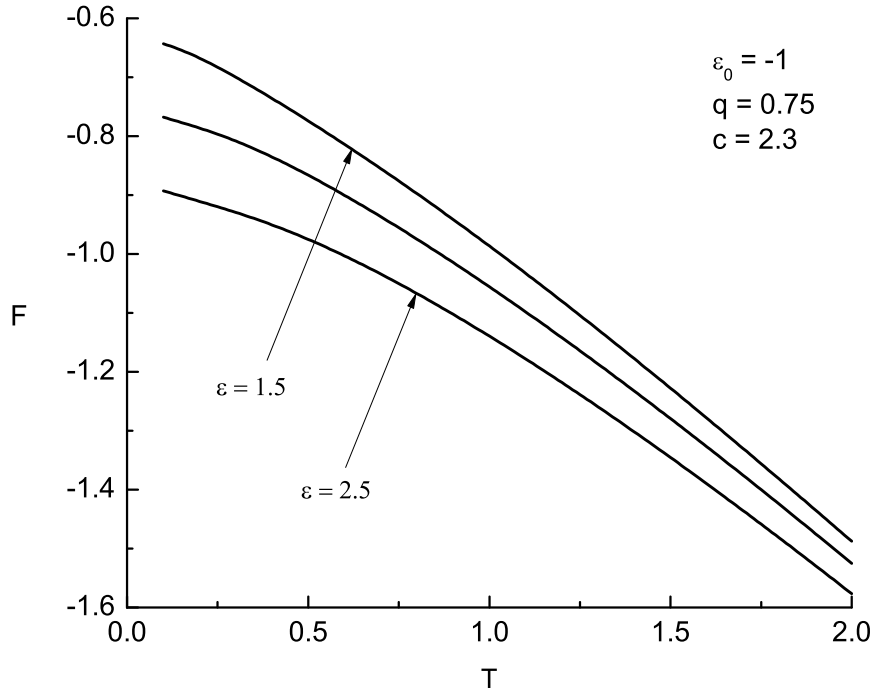

Fig 4.2. Dependence of reduced free energy on temperature, where $\epsilon_0 + \epsilon > 0$.

As we can see from Figs. 4.3,4.4 thermal behavior of the entropy does not change qualitatively until ϵ_0 becomes equal to $-\epsilon$. The rapid decrease of the entropy at low temperatures may lead us to the idea that there is a room for

structural transitions at that stage, and another phase transition is possible. In terms of effective attraction and repulsion we may interpret the case when $\epsilon + \epsilon_0 = 0$ as the situation when similar nucleotides (AA or UU) neither attract nor repulse each other, meanwhile, different nucleotides are attracted by each other.

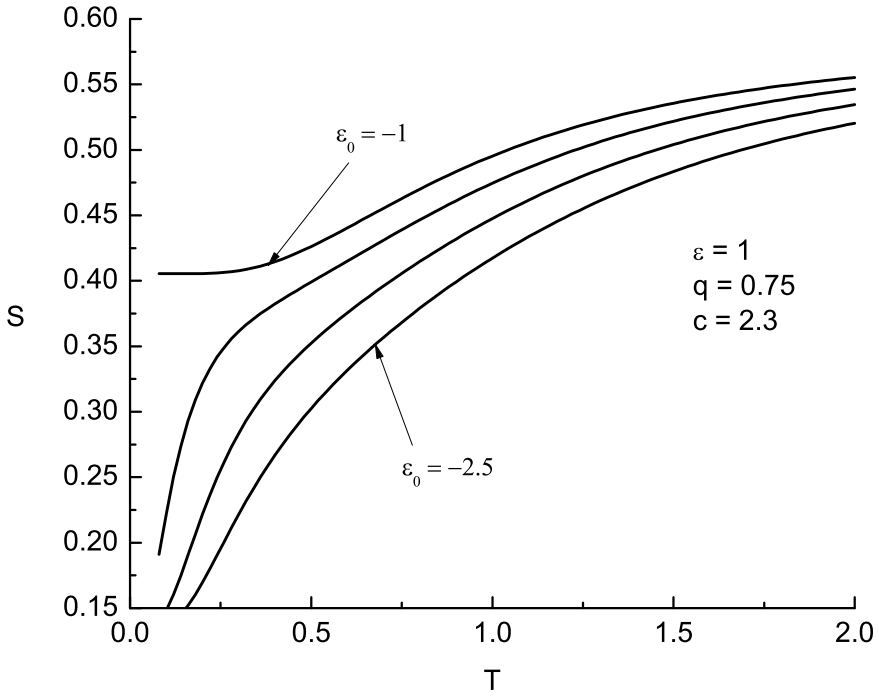

Fig 4.3. Dependence of entropy on temperature.

To provide thermal dependence of specific heat (Fig. 4.5) we implemented numeric calculations on the basis of the formula 4.36. The calculations were performed for different values of ϵ_0 with the increment equal to 0.5. In contrast to the thermal dependencies of the free energy and the entropy (Figs 4.1 and 4.3), the thermal dependence of specific heat is more sensitive to the change of ϵ_0, the difference between peaks flattens with the decrease of ϵ_0. Notice that, two peaks completely vanish when $\epsilon = -\epsilon_0$, and, therefore, there is only one phase transition. For the remaining cases $\epsilon_0 + \epsilon < 0$, and, therefore, according to (3.3), similar nucleotides weakly attract each other. Thus, we obtain non-trivial result that even if $\epsilon_0 + \epsilon < 0$ we observe existence of two phase transitions.

One may notice that the fact of attraction between similar nucleotides is

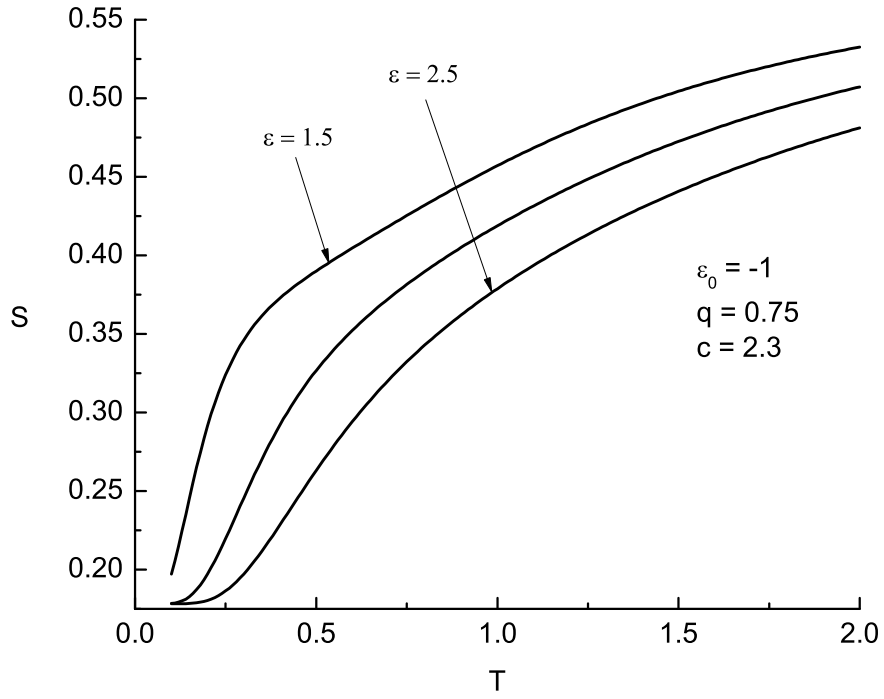

Fig 4.4. Dependence of entropy on temperature.

more evidently presented on Fig. 4.6. At low temperature limit helicity degree is equal to one, i.e., all nucleotides are involved in formation of complementary pairs. However, in the case when $\epsilon_0 + \epsilon = 0$, or, as it was mentioned above, attraction and repulsion between similar nucleotides is absent, helicity degree does not reach its maximal value equal to one.

In case of repulsion between similar nucleotides, $\epsilon_0+\epsilon > 0$ specific heat also exhibits two peaks presented in the Fig. 4.7. Thus, we observe two structural transitions. While the high-temperature peak corresponds to the usual thermal denaturation presented in the Fig.4.8, the low-temperature one indicates the existence of cold denaturation. In this case, the helicity degree in the Fig. 4.8 drastically drops down to the values lower than the high-temperature level of helicity degree.

Comparison of the temperature behavior of thermodynamic characteristics of random ssRNA elucidate a few common points of the secondary structure formation. Specific heat exhibits two-peaks behavior both without loops formation and with loops. Given effect is independent on the type of interaction between

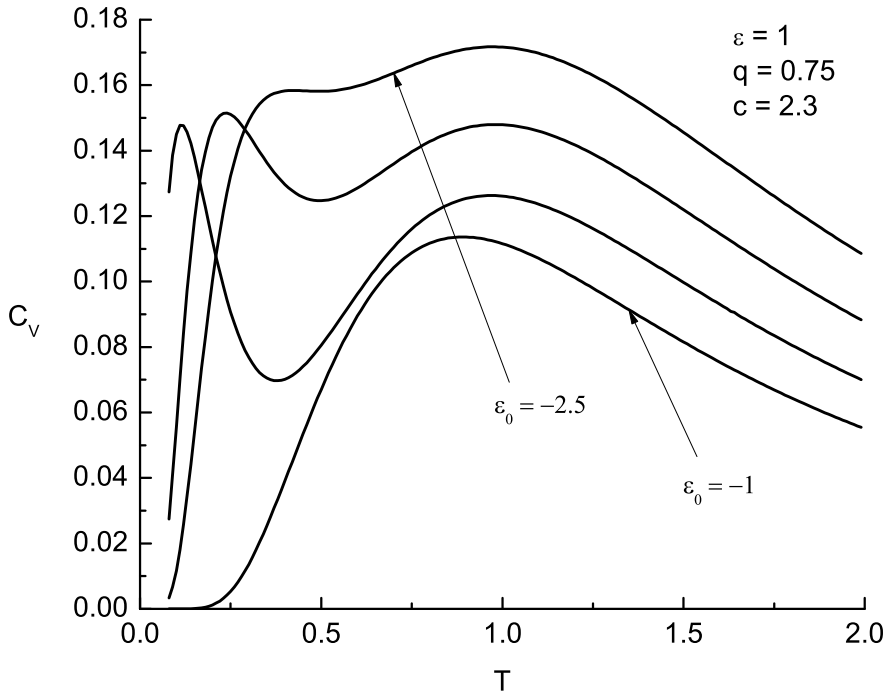

Fig 4.5. Dependence of specific heat on temperature.

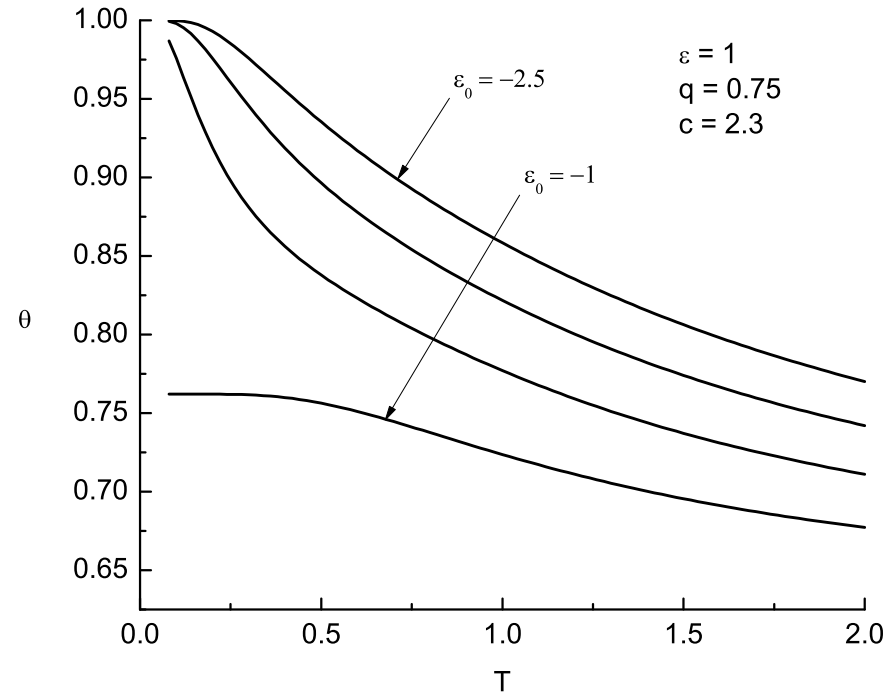

Fig 4.6. Dependence of helicity degree on temperature.

similar nucleotides. The energy of interaction between similar nucleotides is equal to $\epsilon_0 + \epsilon$. In the Figs. 4.11,4.12 is presented the temperature dependence of the specific heat in case of $\epsilon_0 + \epsilon < 0$ and $\epsilon_0 + \epsilon > 0$, correspondingly. Two-

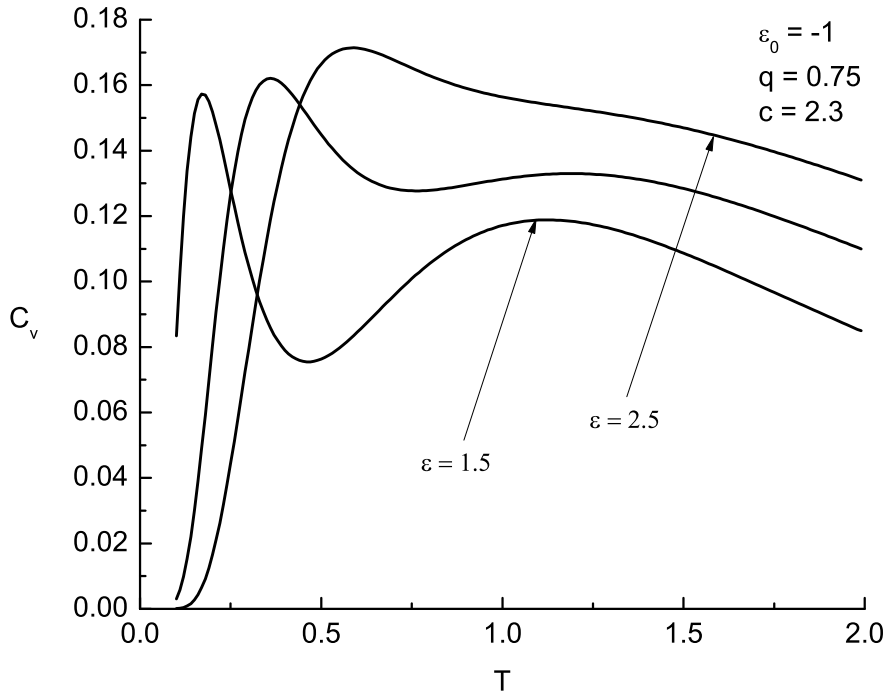

Fig 4.7. Dependence of specific heat on temperature.

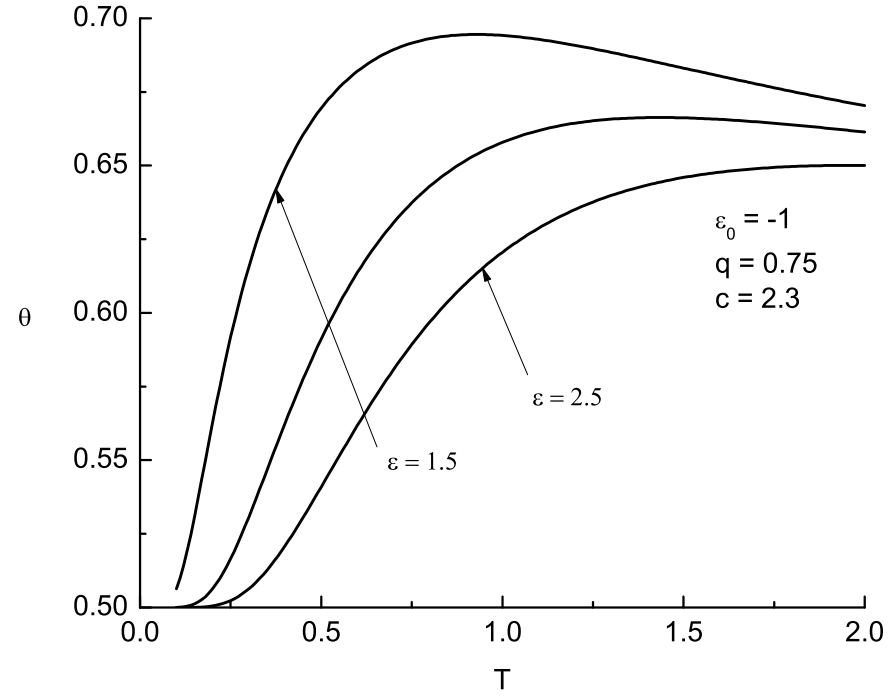

Fig 4.8. Dependence of helicity degree on temperature.

peaks behavior observed in both cases. Besides, the temperature of these two structural transitions remain almost the same with and without loops.

The main difference between specific heat behavior with and without

loops concerns the heat effect of transitions. The heat of the low-temperature transition in higher without loops, while the heat of high-temperature transition is higher with loops. The melting curves (see Figs. 4.9,4.10) are qualitatively the same with and without loops and fit well in the low-temperature region.

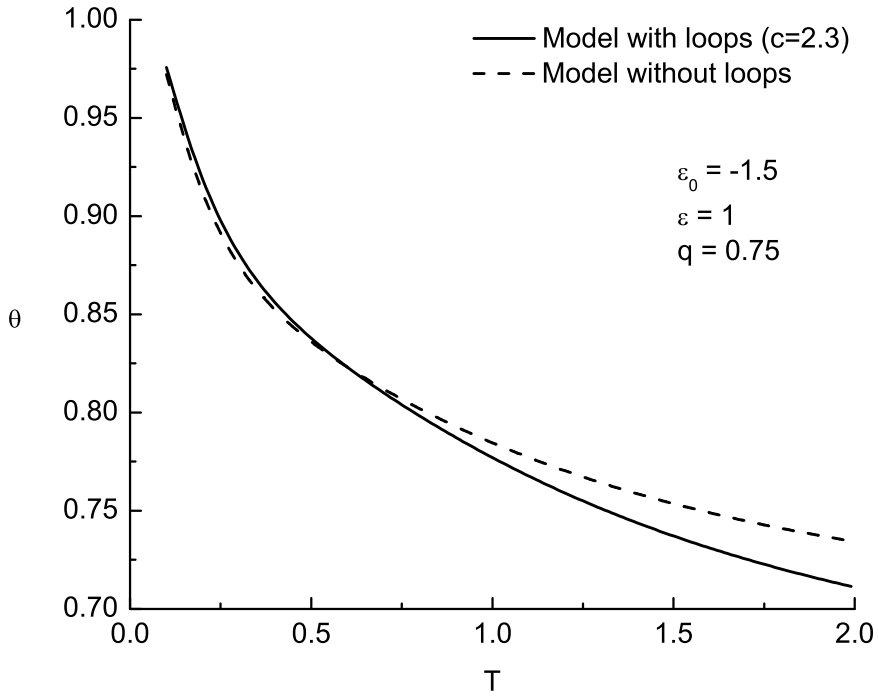

Fig 4.9. Dependence of helicity degree on temperature.

To interpret this result, let's take into account the fact that with decline of temperature amount of loops also decreases. Therefore, difference between the melting curves of models with and without loops vanishes. In case of repulsion between similar nucleotides, characterized by $\epsilon_0 + \epsilon > 0$, the high-temperature denaturation results to lower values of helicity degree with taking into account loops formation. At the same time, low-temperature maximum of specific heat evidences the existence of cold denaturation, which leads to small values of helicity degree at low temperatures. This results in increase of impact of loops in specific heat behavior in the case of repulsion between similar nucleotides. In case of attraction between similar nucleotides, $\epsilon_0 + \epsilon < 0$, the difference of high-temperature helicity degree with and without loops is not pronounced so well. Furthermore,the gap between specific heat with and without loops in

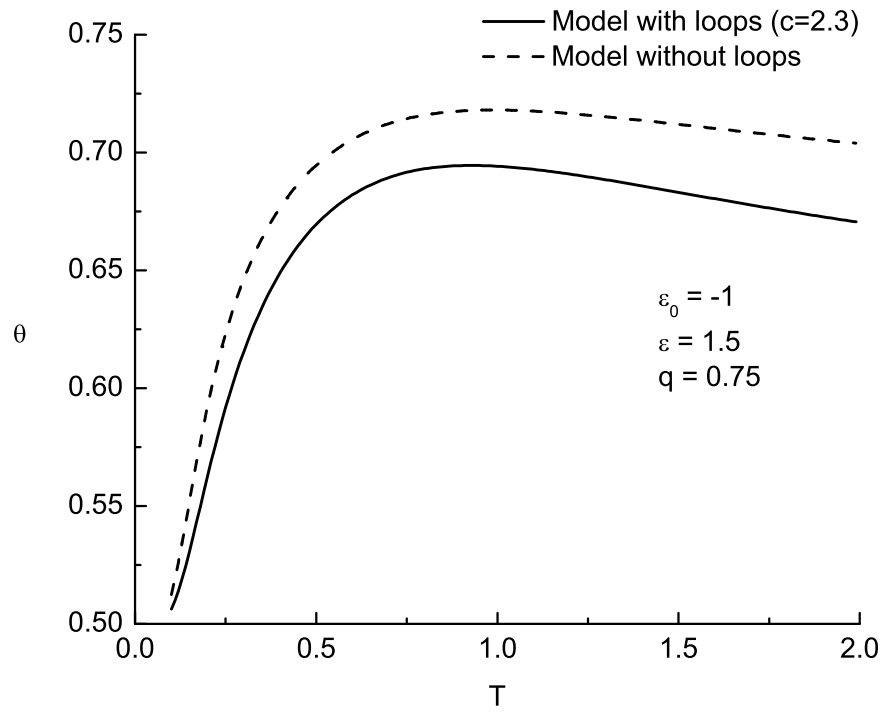

Fig 4.10. Dependence of helicity degree on temperature.

the high-temperature region is almost the same in both cases of attraction and repulsion.

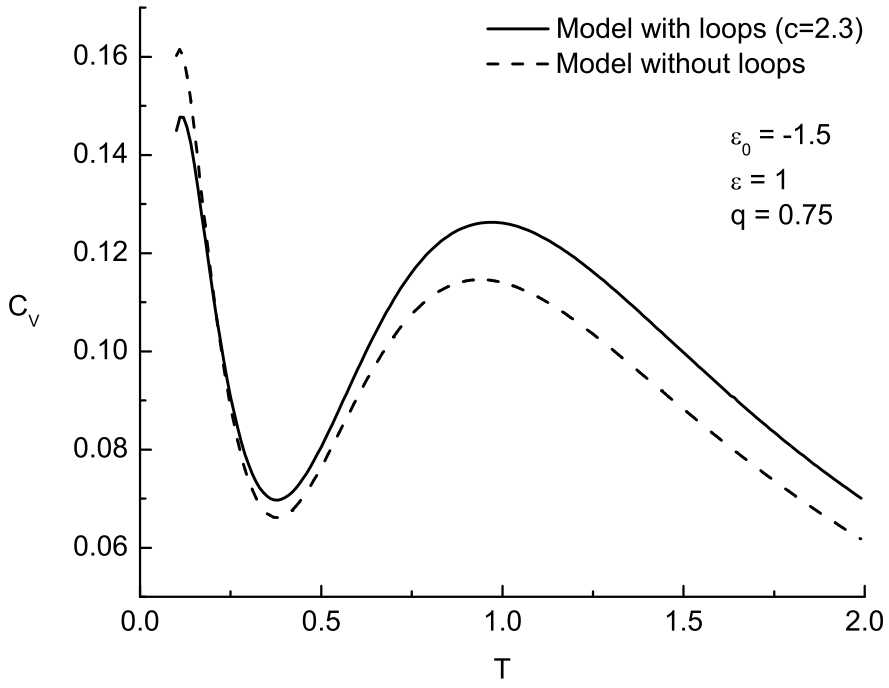

Fig 4.11. Dependence of specific heat degree on temperature.

Thus, the effect of long loops entropy on the phase behavior of single -

stranded RNA is not so important. Given scenario is substantially differs from those for double - stranded DNA, where the loop entropy factor c changes drastically phase behavior of the system. While in the homogeneous ssRNA phase transition exists only for the values $2 < c < c^* \approx 2.479$, homogeneous dsDNA exhibits much more rich phase behavior. Homogeneous dsDNA exhibits phase transition of the second order if $1 < c < 2$, and of the first order one, if $2 < c$. ssRNA exhibits much more smooth transition of the fourth order.

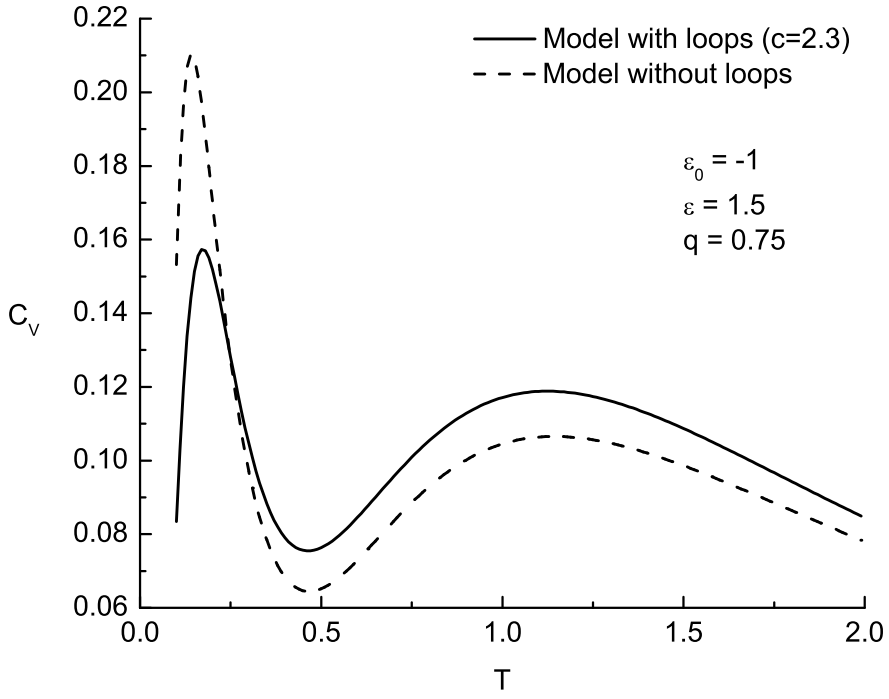

Fig 4.12. Dependence of specific heat degree on temperature.

In the frameworks of the proposed approach, the variational free energy is written in terms of the free energy of the homogeneous RNA with the effective parameters of interaction (see Eqn.(4.10)). Thus, the sequence disorder has no effect on the (fourth) order of the phase transition in ssRNA.

Conclusions

1. The impossibility of knot formation in melted regions of DNA and excluded volume interactions significantly affect the entropy of loop formation and give a value of loop factor $c = 1$.

2. A phase transition of infinite order takes place for the loop factor value $c = 1$ during denaturation of DNA double helix. The phase transition occurs almost in the end of the melting where helicity degree has small values. Above the phase transition temperature helicity degree is zero.

3. Near the critical temperature, the correlation length diverges, as it occurs during the phase transition of second order, whereas the amplitude of fluctuations tends to zero. Thus small but extended fluctuations should take place for the chains with finite length.

4. Single-stranded RNA with random bimodal nucleotide sequence shows two peaks in the temperature dependence of the specific heat of the system both for the attraction and repulsion between the same nucleotides and also for various percentages of the two types of nucleotides. Such behavior indicates the presence of two structural transitions.

5. For the case of repulsion between the nucleotides of the same type, low-temperature peak of specific heat corresponds to the cold melting of RNA when the helicity degree decreases significantly with decreasing temperature. This effect is caused by a large number of thermodynamically unfavorable contacts for a sequence consisting of two types of nucleotides.

6. The account of the entropy of long loop formation does not qualitatively affect the behavior of the specific heat and the helicity degree of single-stranded RNA with a bimodal sequence. The presence of two peaks and cold melting is observed at the same values of the interaction parameters

as without the account of loop entropy.

7. The sequence heterogeneity does not affect the existence of a phase transition of the forth order, which takes place when $2 \leq c \leq c^*$ in a homogeneous single-stranded RNA. When $c < 2$ in a single-stranded RNA phase transition does not occur both in homo- and heterogeneous sequences.

Bibliography

1. D. Soll. *The RNA World*, chapter Transfer RNA: an RNA for all seasons, pages 157–183. Cold Spring Harbor Laboratory Press, 1993.

2. M. Sprinzl, C. Steegborn, F. Hubel, and S. Steinberg. The transfer RNA database. *Nucl. Acids Res.*, 24:68–72, 1996.

3. P. B. Moore. *The RNA World*, chapter Ribosomes and the RNA world, pages 119–135. Cold Spring Harbor Laboratory Press, 1993.

4. H. F. Noller. *The RNA World*, chapter On the origin of the ribosome: coevolution of subdomains of RNA and RNA, pages 137–156. Cold Spring Harbor Laboratory Press, 1993.

5. R. A. Zimmermann and A. E. Dahlberg. *Protein Biosynthesis*, chapter Ribosomal RNA: Structure, Evolution, Processing, and Function. CRC Press, 1996.

6. Y. Van de Peer, A. Caers, P. De Rijk, and R. De Wachter. Database on the structure of small ribosomal subunit RNA. *Nucl. Acids Res.*, 26:179–182, 1998.

7. P. De Rijk, A. Caers Y. Van de Peer, and R. De Wachter. Database on the structure of small ribosomal subunit RNA. *Nucl. Acids Res.*, 26:183–186, 1998.

8. B. L. Maidak, J. R. Cole, C. T. Jr Parker, G. M. Garrity, N. Larsen, B. Li, T. G. Lilburn, M. J. McCaughey, G. J. Olsen, R. Overbeek, S. Pramanik, T. M. Schmidt, J.M. Tiedje, and C. R. Woese. A new version of the rdp (ribosomal database project). *Nucl. Acids Res.*, 27:171–173, 1999.

9. S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers,

T. Nielson, and D. H. Turner. Improved free energy parameters for prediction of RNA duplex stability. *PNAS*, 83:9373–9377, 1986.

10. J. Jr SantaLucia and D. H. Turner. Measuring the thermodynamics of RNA secondary structure formation. *Biopolymers*, 44:309–319, 1997.

11. A.M. Pyle and J. B. Green. Rna folding. *Curr. Opin. Struct. Biol.*, 5:303–310, 1995.

12. P. Brion and E. Westhof. Hierarchy and dynamics of RNA folding. *A. Rev. Biophys. Biomol. Struct.*, 26:113–137, 1997.

13. I. Jr. Tinoco and C. Bustamante. How RNA folds. *J. Mol. Biol.*, 293(2):271–281, 1999.

14. F.H.D. van Batenburg, A.P. Gultyaev, C.W.A. Pleij, J.Ng, and J. Oliehoek. Pseudobase: a database with RNA pseudoknots. *Nucl. Acids Res.*, 28:201–204, 2000.

15. E. Westhof and L. Jaeger. Rna pseudoknots. *Cur. Opin. Struct. Biol.*, 2:327–333, 1992.

16. C. W. Hilbers, P. J. A. Michiels, and H.A. Heus. New developments in structure determination of pseudoknots. *Biopolymers*, 48:137–153, 1998.

17. T. Hermann and D. J. Patel. Stitching together RNA tertiary architectures. *J. Mol. Biol.*, 294:829–849, 1999.

18. E. Rivas and S. R. Eddy. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.*, 285:2053–2067, 1999.

19. Vedenov A.A., Dykhne A.M., and Frank-Kamenetskii M.D. The helix-coil transition in DNA. *Sov.Phys. Usp.*, 14:715–736, 1972.

20. Yu. S. Lazurkin. *Physical Methods of Studying Proteins and Nucleic Acids.* Nauka, 1967.

21. M. J. Chamberlin. Comparative properties of DNA, RNA, and hybrid homopolymer pairs. *Fed. Proc.*, 24:1446–1457, 1965.

22. R. B. Inman and R. L. Baldwin. Helix-random coil transitions in DNA homopolymer pairs. *J. Mol. Biol.*, 8:452–469, 1964.

23. Haroutiunian S.G., Dalyan Y.B., Aslanian V.M., Lando D.Yu., and Akhrem A.A. A new method for detemining the relative effect of ligand on at- and gc- base pairs in DNA: application to metal ions, protons and two amino acids. *Nucl. Acids Res.*, 18(21):6413–6417, 1990.

24. Lando D.Yu., Haroutiunian S.G., Kulba A.M., Dalyan Y.B., Orioli P., Mangani R., and Akhrem A.A. Theoretical and experimental study of DNA helix-coil transition in acidic and alkaline medium. *J. Biomol. Struct. Dyn.*, 12:355–366, 1994.

25. J . Marmur and P. Doty. Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol. Biol.*, 5:109–118, 1962.

26. R. J . Owen, L. R. Hill, and S. P. Lapage. Determination of DNA base compositions from melting profiles in dilute buffers. *Biopolymers*, 7:503–516, 1969.

27. Yu. N. Kosaganov, Yu. S. Lazurkin, and N. V. Sidorenko. *Molekulyarnaya Biologiya*, 1:352, 1967.

28. J . Vinograd, J. Lebowitz, and R. Watson. Early and late helix-coil transitions in closed circular DNA. the number of superhelical turns in polyoma DNA. *J . Mol. Biol.*, 33(1):173–197, 1968.

29. P. G. Higgs. Rna secondary structure: physical and computational aspects. *Q. Rev. Biophys*, 33:199–253, 2000.

30. R. Nussinov and A. B. Jacobson. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl. Acad. Sci. USA*, 77(11):6309–6313, 1980.

31. M. Zuker and P. Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, 9(1):133–148, 1981.

32. J. S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.

33. B.H. Zimm. Theory of "melting"of the helical form in double chains of the DNA. *J. Chem. Phys.*, 33(5):1349–1356, 1960.

34. D. Poland and H. A. Scheraga. Occurrence of a phase transition in nucleic acid models. *J. Chem. Phys.*, 45:1464, 1966.

35. S. Lifson. Partition functions of linear chain molecules. *J. Chem. Phys.*, 40:3705, 1964.

36. A. Litan and S. Lifson. Statistical mechanical treatment of multistranded polynucleotide molecules. *J. Chem. Phys.*, 42:2528, 1965.

37. P.J. Flory. Theory of elastic mechanisms in fibrous proteins. *J. Am. Chem. Soc.*, 78:5222–5235, 1956.

38. M.E. Fisher. Effect of excluded volume on phase transition in biopolymers. *J. Chem. Phys.*, 45:1464, 1966.

39. Y. Kafri, D. Mukamel, and L. Peliti. Melting and unzipping of DNA. *Eur. Phys. J. B*, 27(1):135–146, 2002.

40. Y. Kafri, D. Mukamel, and L. Peliti. Why is the DNA denaturation transition first order? *Phys. Rev. Lett.*, 85:4988–4991, 2000.

41. M. Peyrard and A.R. Bishop. Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.*, 62(23):2755–2758, 1989.

42. T. Dauxois, M. Peyrard, and A. R. Bishop. Dynamics and thermodynamics of a nonlinear model for DNA denaturation. *Phys. Rev. E*, 47(1):684–695, 1993.

43. T. Dauxois, M. Peyrard, and A. R. Bishop. Entropy-driven DNA denaturation. *Rap. Comm. Phys. Rev. E*, 47(1):R44–R47, 1993.

44. N. Theodorakopulos, T. Dauxois, and M. Peyrard. Order of the phase transition in models of DNA thermal denaturation. *Phys. Rev. Lett.*, 85(1):6–9, 2000.

45. M. Peyrard. Nonlinear dynamics and statistical physics of DNA. *Nonlinearity*, 17:R1–R40, 2004.

46. W.B. Melchior Jr. and P.H. Von Hippel. Alteration of the relative stability of da-dt and dg-dc base pairs in DNA. *Proc. Nat. Acad. Sci. USA*, 70(2):298–302, 1973.

47. V.F. Morozov, E.Sh. Mamasakhlisov, Sh.A. Hayryan, and Chin-Kun Hu. Microscopical approach to the helix-coil transition in DNA. *Physica A*, 281:51–59, 2000.

48. G.N. Hayrapetyan, Y.Sh. Mamasakhlisov, V.F. Morozov, Vl.V. Papoyan, and V.B. Priezzhev. Two-dimensional random walk and loop factor in helix-coil transition theory of DNA. *J. Contemp. Phys.*, 46(5):242–246, 2011.

49. G.N. Hayrapetyan. Two-dimensional random walk and characteristics of helix-coil transition in DNA. *J. Contemp. Phys.*, 47(2):93–97, 2012.

50. E. W. Montroll. *Applied combinatorial mathematics*, chapter Lattice statistics. John Wiley and sons, Inc., New York, London, Sydney, 1964.

51. S.N Majumdar. An ideal polymer chain in arbitrary dimensions near attractive site. *Physica A*, 169:207–236, 1990.

52. R.J. Rubin. Random-walk model of chain-polymer adsorption at a surface. *J. Chem. Phys.*, 43:2392, 1965.

53. F. Spitzer. *Principles of Random Walk.* Van Nostrand, Toronto, New York, London, 1964.

54. P. Flajolet and R. Sedgewick. *Analytic combinatorics.* Cambridge University Press, 2009.

55. G.N. Hayrapetyan, Y.Sh. Mamasakhlisov, V.F. Morozov, V.V. Papoyan, and V.B. Priezzhev. Two-dimensional random walk and critical behavior of double-strand DNA. *J. Phys. A: Math. Theor.*, 46:035001, 2013.

56. O. Costin, R.D. Costin, and C.P. Grunfeld. Infinite-order phase transition in a classical spin system. *J. Stat. Phys.*, 59:1531–1546, 1990.

57. O. Costin and R.D. Costin. Limit probability distributions for an infinite-order phase transition model. *J. Stat. Phys.*, 64:193–205, 1991.

58. M. Bundaru and C.P. Grunfeld. On a phase transition in a one-dimensional non-homogeneous model. *J. Phys. A: Math. Gen.*, 32:875, 1999.

59. M. Bauer, S. Coulomb, and S.N. Dorogovtsev. Phase transition with the berezinskii-kosterlitz-thouless singularity in the ising model on a growing network. *Phys. Rev. Lett.*, 94:200602, 2005.

60. Yu. P. Blagoi, V.A. Sorokin, V.A. Valkeev, G.O. Gladchenko, S.A. Khomenko, and V.L. Galkin. On the possibility of true phase transition in

heterogeneous DNA during thermal denaturation under conditions close to equal stability of a+t and g+c pairs. *Biopolymers*, 18(9):2279, 1979.

61. A.Yu. Grosberg and A.R. Khokhlov. *StatisticalPhysics of Macromolecules*. AIP Press, 1994.

62. M.D. Frank-Kamenetskii. Biophysics of the DNA molecule. *Physics Reports*, 288:13–60, 1998.

63. G.N. Hayrapetyan, V.F. Morozov, Vl.V. Papoyan, S.S. Pogosyan, and V.B. Priezzhev. The helix-coil transition in double-stranded polynucleotide and two-dimensional random walk. *Mod. Phys. Lett. B*, 26:1250083, 2012.

64. G.N. Hayrapetyan, Y.Sh. Mamasakhlisov, S.S. Pogosyan, and Vl.V. Papoyan. The melting phenomenon in random-walk model of DNA. *Physics of Atomic Nuclei*, 75:1268–1271, 2012.

65. C. Richard and A. J. Guttmann. Ploand-sheraga models and the DNA denaturation transitions. *J. Stat. Phys.*, 115:925, 2004.

66. R.M. Wartell and A.S. Benight. Thermal denaturation of DNA molecules: A comparison of theory with experiment. *Physics Reports*, 126:67–107, 1985.

67. M.D. Frank-Kamenetskii. DNA topology. *Journal of Molecular Structure: THEOCHEM*, 336:235–243, 1995.

68. Ed. Raymond, F. Gesteland, and John F. Atkins, editors. *The RNA World*. Number 2. Cold Spring Harbor Laboratory Press, 1993.

69. Y. Sh. Mamasakhlisov, Sh. Hayryan, V.F. Morozov, and C.-K. Hu. Rna folding in the presence of counterions. *Phys. Rev. E*, 75:061907, 2007.

70. S. R. Morgan and P. G. Higgs. Barrier heights between groundstates in a model of RNA secondary structure. *J. Phys. A: Math. Gen.*, 31:3153–3170, 1998.

71. R. Bundschuh and T. Hwa. Statistical mechanics of secondary structures formed by random RNA sequences. *Phys. Rev. E*, 65:031903, 2002.

72. R. Bundschuh and T. Hwa. Phases of the secondary structures of RNA sequences. *Europhys. Lett.*, 59:903–909, 2002.

73. F. Krzakala, M. Mezard, and M. Mueller. Nature of the glassy phase of RNA secondary structure. *Europhys. Lett.*, 57(5):752–758, 2002.

74. T. R. Einert, P. Nager, H. Orland, and R. R. Netz. Impact of loop statistics on the thermodynamics of RNA folding. *Phys. Rev. Lett.*, 101:048103, 2008.

75. T. R. Einert and R. R. Netz. Theory for RNA folding, stretching, and melting including loops and salt. *Biophys. J.*, 10:2745–2753, 2011.

76. S. Hui and L.-H. Tang. Ground state and glass transition of the RNA secondary structure. *Eur. Phys. J. B*, 53:77–84, 2006.

77. M. Mueller, F. Krzakala, and M. Mezard. The secondary structure of RNA under tension. *Eur. Phys. J. E*, 9:67–77, 2002.

78. C. Monthus and T. Garel. Freezing transition of the random bond RNA model: statistical properties of the pairing weights. *Phys. Rev. E*, 75:031103, 2007.

79. M. Mezard, G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro. Replica symmetry breaking and the nature of the spin glass phase. *J. Physique*, 45:843–854, 1984.

80. M. Mezard, G. Parisi, and M. Virasoro. *Spin Glass Theory and Beyond.* World Scientific, 1987.

81. V. Dotsenko. *Introduction to the Replica Theory of Disordered Statistical Systems.* Cambridge University Press, 2001.

82. K. K. Binder and A. P. Young. Spin glasses: Experimental facts, theoretical concepts, and open questions. *Rev. Mod. Phys.*, 58:801–976, 1986.

83. C. D. Sfatos, A. M. Gutin, and E. I. Shakhnovich. Phase-diagram of random copolymers. *Phys. Rev. E*, 48:465–475, 1993.

84. V. S. Pande, A. Yu. Grosberg, and T. Tanaka. Heteropolymer freezing and design: Towards physical models of protein folding. *Rev. Mod. Phys.*, 72:259–314, 2000.

85. M. Mueller. Statistical physics of RNA folding. *Phys. Rev. E*, 67:021914, 2003.

86. A. Pagnani, G. Parisi, and F. Ricci-Tersenghi. Glassy transition in a disordered model for the RNA secondary structure. *Phys. Rev. Lett.*, 84:2026–2029, 2000.

87. M. Serva and G. Paladin. Gibbs thermodynamic potentials for disordered systems. *Phys. Rev. Lett.*, 70(2):105–108, 1993.

88. M. Pasquini and M. Serva. Two-dimensional frustrated ising model with four phases. *Phys. Rev. E*, 56:2751, 1997.

89. M. Pasquini and M. Serva. Sequence of constrained annealed averages for one-dimensional disordered systems. *Phys. Rev. E*, 51:2006, 1995.

90. T. Liu and R. Bundschuh. Quantification of the differences between

quenched and annealed averaging for RNA secondary structures. *Phys. Rev. E*, 72:061905, 2005.

91. L. Leuzzi and F. Ritort. The disordered backgammon model. *Phys. Rev. E*, 65:056125, 2002.

92. M. Lassig and K. J. Wiese. Freezing of random RNA. *Phys. Rev. Lett.*, 96(22):228101, 2006.

93. F. David and K. J. Wiese. Field theory of the RNA freezing transition. *J. Stat. Mech.*, (10):P10019, 2009.

94. F. David, C. Hagendorf, and K. J. Wiese. Random RNA under tension. *Europhys. Lett.*, 78:68003, 2007.

95. F. David, C. Hagendorf, and K. J. Wiese. A growth model for RNA secondary structures. *J. Stat. Mech.*, 4:P04008, 2008.

96. F. David and K. J. Wiese. Systematic field theory of the RNA glass transition. *Phys. Rev. Lett.*, 98:128102, 2007.

97. G.N. Hayrapetyan, H.L. Tsaturyan, Sh.A. Tonoyan, and Y.Sh. Mamasakhlisov. The melting phenomenon in random-walk model of DNA. *J. Contemp. Phys.*, 48(2):98–102, 2013.

98. P. J. Mikulecky and A. L. Feig. Cold denaturation of the hammerhead ribozyme. *J. Am. Chem. Soc.*, 124:890–891, 2002.

99. P.L. Privalov. Cold denaturation of proteins. *Crit. Rev. Biochem. Mol. Biol.*, 25(4):281–305, 1990.

100. A.V. Finkelstein. *Protein Physics*. Acad. Press, 2002.

101. A. V. Badasyan, Sh. A. Tonoyan, Y. Sh. Mamasakhlisov, A. Giacometti, A. S. Benight, and V. F. Morozov. Competition for hydrogen-bond

formation in the helix-coil transition and protein folding. *Phys. Rev. E*, 83:051903, 2011.

102. T. R. Einert, H. Orland, and R. R. Netz. Secondary structure formation of homopolymeric single-stranded nucleic acids including force and loop entropy: Implications for DNA hybridization. *Eur. Phys. J. E*, 34(6):1–15, 2011.

103. P. G. Higgs. Overlaps between RNA secondary structures. *Phys. Rev. Lett.*, 76:704–707, 1996.