

# Description of Helix-Coil Transition in Heteropolymers in Frame of GMPC

A. V. Asatryan<sup>a, b, \*</sup>, A. K. Kalashyan<sup>c</sup>, and V. F. Morozov<sup>b</sup>

<sup>a</sup> Institute of Chemical Physics after A.B. Nalbandyan, NAS of Armenia, Yerevan, Armenia

<sup>b</sup> Yerevan State University, Yerevan, Armenia

<sup>c</sup> National Polytechnic University of Armenia, Vanadzor Branch, Vanadzor, Armenia

\*e-mail: arevik.asatryan@ysu.am

Received September 9, 2021; revised September 22, 2021; accepted October 3, 2021

**Abstract**—The work is devoted to the study of the behavior of heterogeneous biopolymers during the melting of the secondary structure. A random bimodal biopolymer is generated. Each component corresponds to its transfer matrix of GMPC, by multiplying these transfer matrices the free energy of the system is calculated. It is shown that for a length of 5000 repeating units, the reduced free energy does not depend on the realization of the system (primary structure). The corresponding denaturation curves show that the transition point is well defined as a linear combination of the corresponding homopolymers, but the profile of the curves depends on the implementation. The behavior of the number of junctions between the helical and coiled states is calculated. An unexpected result is obtained that the racemic heteropolymer with a length of 5000 repeated units is divided into about 10 sections during the melting process. We associate this with the behavior of correlation in the system, which will be investigated further.

**Keywords:** heteropolymer, helix-coil transition, biopolymer, denaturation, transfer-matrix

**DOI:** 10.3103/S106833722104006X

## 1. INTRODUCTION

In biopolymers, various processes in biological systems often result in or are accompanied by a helix-coil transition. Despite the successes achieved, the problem of the helix-coil transition is still one of the most popular problems in structural biology, and at present interest in the problem is unabated [1–5]. In this publication, we base ourselves on the microscopic theory of the helix-coil transition [6–11]. This became possible thanks to the approaches and methods of modern theoretical physics, which make it possible to use in the statistical physics of polymers a large number of theoretical models that describe the subject of study more adequately than is accepted today and do not require the introduction of an additional phenomenological parameter to take into account the cooperativity of the system. The theory of the helix-coil transition in polypeptides was developed based on the Potts-like model [12–18]. It has also been shown that neglecting large-scale loops, the characteristic equation for the DNA model is the same as that for the Generalized Polypeptide Chain Model (GMPC) [15]. The difference from the polypeptide model is that the number of repeated units fixed to a single hydrogen bond enters the GMPC as a parameter. As a result the description of Helix-coil transition in polypeptides, as well as in polynucleotides, and also for the study of the influence of factors of different natures can be done by the same model GMPC with different parameters. Taking into account that further research should be carried out based on the Generalised Model of Polypeptide Chain, let us bring its fundamentals.

## 2. GMPC MODEL

Suppose one has a biopolymer containing  $N$  repeating units. It is assumed that each repeating unit can be in  $Q$  discrete states corresponding to different discrete values of conformations (the rotational isomer approximation), the energies of these states are assumed to be the same and do not depend on the states of neighboring units. Let us describe the state of the  $i$ -th unit using the spin variable  $\gamma_i$ , so  $\gamma_i = 1, \dots, Q$ . A hydrogen bond, resulting in the formation of a helical structure, is formed when the  $\Delta$  adjacent repeat-

ing units are in the same defined conformation. Let this be called the number one conformation. The Hamiltonian of such GMPC model has the form

$$-\beta H = J \sum_{i=1}^N \prod_{k=\Delta-1}^0 \delta(\gamma_{i-k}, 1) = J \sum_{i=1}^N \delta_i^{(\Delta)}, \quad (1)$$

where  $\delta_i^{(\Delta)} = \prod_{k=\Delta-1}^0 \delta(\gamma_{i-k}, 1)$  is the product of Kronecker symbols, which provides energy release  $J$  only in the case when  $\Delta$  consecutively adjacent repeating units are in a helical conformation, designated by us as number 1. The transfer matrix GMPC ( $\Delta \times \Delta$ ) can be constructed for specific values of  $\Delta$  and  $Q$  based on the Hamiltonian (1). It has the following properties [15–21]:

- the element (1, 1) is equal to  $W = e^J$ ;
- all elements of the last line are equal to 1;
- the elements of the first upper pseudo-diagonal are equal to 1;
- elements ( $\Delta - 1, \Delta$ ) and ( $\Delta, \Delta$ ) are equal to  $Q-1$ ;
- all other elements are zero.

Generalizing to the case of arbitrary finite  $\Delta$  and  $Q$ , we obtain the transfer matrix of the model

$$G = \begin{pmatrix} e^J & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & Q-1 \\ 1 & 1 & 1 & \dots & 1 & 1 & Q-1 \end{pmatrix}. \quad (2)$$

For the homopolymer of sufficiently large length, the partition function and the characteristic equation for the transfer matrix of the model have the form:

$$Z = \text{Tr} G^N, \quad (3)$$

$$\lambda^{(\Delta-1)}(\lambda - e^J)(\lambda - Q) = (e^J - 1)(Q - 1). \quad (4)$$

As mentioned above [5]:  $J$  is the energy of the hydrogen bond, reduced to temperature;  $Q$  is the number of conformations of repeating units, the conformation number 1 corresponds to the helical state;  $\Delta$  is the number of repeating units fixed by one hydrogen bond;  $W = e^J$  is the temperature parameter that determines the energy properties of repeating units. The latter parameters define various properties of the system. This Hamiltonian corresponds to the transfer matrix (2) and the secular equation  $W = e^J$  (4).

For the basic model defined above, the temperature parameter  $W$  has the form  $W = e^{-U/(KT)}$ , where  $U$  is the hydrogen bond energy, and the entropy parameter  $Q$  is a constant. However, when the interaction with the solvent and lianas is taken into account,  $W$  becomes a complex function of temperature, and  $Q$  also becomes temperature-dependent [11].

### 3. FREE ENERGY OF A BIMODAL HETEROPOLYMER

Consider the bimodal heteropolymer consisting of repeating units of two classes “A” and “B”. In this case, the GMPC Hamiltonian will have the form

$$-\beta H = \sum_{i=1}^N J_i \delta_i^{\Delta}, \quad (5)$$

where  $J_i = \begin{cases} J_A \\ J_B \end{cases}$ , and  $Q_i = \begin{cases} Q_A \\ Q_B \end{cases}$ .

Because we will investigate the dependence of the free energy on  $N$ , the partition function of such a system will have the form [18], although, for large  $N$ , one can use the expression (3)

$$Z = J^* \times \prod_{i=1}^N G_i \times J, \quad (6)$$

where  $G_i$  is the transfer matrix GMPC (2) corresponding to the given model,  $J^* = (0 \ 0 \ \dots \ 1)$ ,  $\Delta$  is the component row vector, and  $J = (0 \ 0 \ \dots \ 0 \ 1)^T$  is the column vector.

Because the product of matrices is not commutative, then, generally speaking, the partition function depends on the sequence of matrices, which corresponds to the primary sequence of repeating units. However, it was shown earlier that for a random sequence at sufficiently large  $N$  the value  $\ln Z/N$  tends to a finite limit [22], as well as the reduced free energy, then we need to determine the order of magnitude of  $N$  when the system can be considered as a thermodynamic with a given degree of accuracy.

#### 4. HELICITY DEGREE

From the definition of the degree of helicity  $\theta$  as the average fraction of repeating units in a helix state

$$\theta_N = N_h/N, \quad (7)$$

where  $N_h$  is the number of repeating units in a Helix state. For bimodal case, we suppose  $J_A = J_0 + \Delta J$ ,  $J_B = J_0 - \Delta J$ ,  $J_0 = (J_A + J_B)/2$ ,  $\Delta J = (J_A - J_B)/2$ . Then the magnitude of the statistical sum will be determined by the following expression:

$$Z = \sum_{\{\gamma_i\}} e^{\sum J_i \delta_i^\Delta}, \quad (8)$$

$J_i = J_0 + \Delta J \varepsilon$  where  $\varepsilon$  takes the values  $\pm 1$ . Hence, from (6),  $N_h = \frac{1}{Z} \sum_{\{\gamma_i\}} \sum_{i=1}^N \delta_i^\Delta e^{\sum J_i \delta_i^\Delta} = \frac{1}{Z} \frac{\partial Z}{\partial J_0}$ . The degree of helicity will have the form:

$$\theta_N = \frac{1}{NZ} \tilde{J} (G'_1 G_2 G_3 \dots G_N + G_1 G'_2 G_3 \dots G_N + \dots + G_1 G_2 G'_3 G_4 \dots G_N + \dots + G_1 G_2 G_3 G'_4 \dots G_N) J, \quad (9)$$

where  $G'_i$  is the matrix  $G_i$ , where all elements except the first are zero. Following [9, 10], the expression in parenthesis can be written as  $(E \ O) \prod_{i=1}^N M_i \begin{pmatrix} E \\ O \end{pmatrix}$ , where  $M_i$  is a supermatrix of the form  $\begin{pmatrix} G_i & G'_i \\ O & G_i \end{pmatrix}$ ,  $E$  is the identity matrix of size  $\Delta$ , and  $O$  is the null matrix of the same size. Finally, we obtain

$$\theta_N = \frac{1}{NZ_N} J^* (E \ O) \prod_{i=1}^N M_i \begin{pmatrix} E \\ O \end{pmatrix} J = \frac{1}{N} \frac{J_\theta^* \prod_{i=1}^N M_i J_\theta}{J^* \prod_{i=1}^N G_i J}, \quad (10)$$

Here  $J_\theta^* = (0 \ 0 \ \dots \ 0 \ 0 \ 1 \ \dots \ 1)$ ,  $J = (0 \ 0 \ \dots \ 1 \ 1)^T$  vector row and vector column with size  $2\Delta$ .

#### 5. THE AVERAGE PORTION OF JUNCTIONS BETWEEN HELIX AND COILED SECTIONS

The average portion of junctions  $\eta_N$  is defined as the fraction of pairs of the helical state and any other state

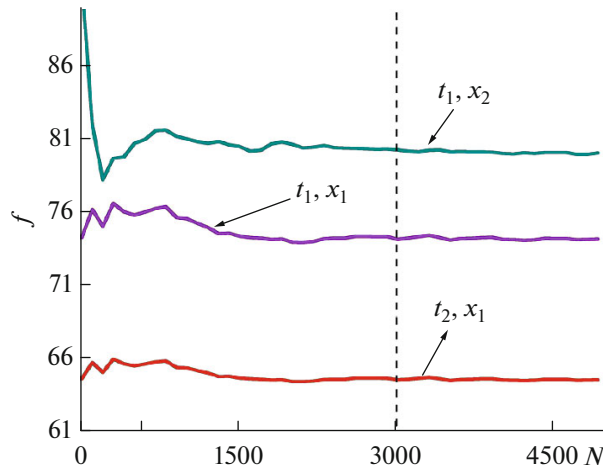
$$\langle \eta_N \rangle = p(\gamma_i = 1, \gamma_{i+1} = k), \quad k \neq 1. \quad (11)$$

Hence,  $\eta_N = \theta - p(hh)$ . Similarly to (8), we obtain

$$p(hh) = \frac{1}{NZ} J^* (G'_1 G_2 G_3 G_4 \dots G_N + G_1 G'_2 G_3 \dots G_N + \dots + G_1 G_2 \dots G'_{N-1} G'_N) J. \quad (12)$$

Similarly to how the expression for  $\theta$  was obtained, we introduce the supermatrix  $L_i$

$$L_i = \begin{pmatrix} G_i & G'_i & O \\ O & O & G'_i \\ O & O & G_i \end{pmatrix}. \quad (13)$$



**Fig. 1.** Dependence of the reduced free energy on the number of repeating units for different  $x$  and temperatures.  $x_1 = 0.4$ ,  $x_2 = 0.5$ ,  $\Delta = 4$ ,  $t_1 = 0.213$ ,  $t_2 = 0.219$ ,  $U_A = 1$ ,  $U_B = 0.8$ ,  $Q_A = 71$ ,  $Q_B = 51$ .

The expression in parentheses in (12) is obtained in terms of  $L_i$  as

$$(E \ O \ O) \prod_{i=1}^{N-1} L_i (O \ O \ E)^T. \tag{14}$$

from which  $\eta_N = \frac{1}{ZN} J_L^* \prod_{i=1}^{N-1} L_i J_L$ , where  $J_L^*$  is the row vector of size  $3\Delta$   $(0 \ 0 \ \dots \ 0 \ 0 \ 1 \ \dots \ 1)$ , and  $J_L$  is the column vector of the same size  $(0 \ 0 \ \dots \ 1 \ 1)^T$ . Thus, the average part of joints

$$\eta_N = \frac{1}{ZN} J_\theta^* \prod_{i=1}^N M_i J_\theta - \frac{1}{ZN} J_\eta^* \prod_{i=1}^{N-1} L_i J_\eta, \tag{15}$$

where  $Z$  is defined from (6). From here, one can go to the expression for the average length of the helix section. Because the average length of the helix region is  $v_N = \theta_N / \eta_N$ , we obtain

$$v_N = \left( 1 - \frac{J_\eta^* \prod_{i=1}^{N-1} L_i J_\eta}{J_\theta^* \prod_{i=1}^N M_i J_\theta} \right)^{-1}. \tag{16}$$

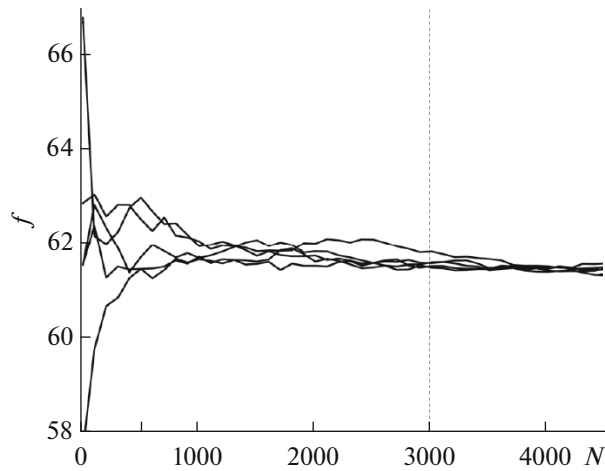
### 6. RESULTS AND DISCUSSION

Let us enter the value  $x$  as the probability of repeating units “A” in a random sequence, and  $1-x$  – “B”, respectively. So, one can generate a random sequence of repeating units “A” and “B” of a given length  $N$ . Accordingly,  $C_A = N_A/N$ ,  $C_B = N_B/N$ , where  $N_A$  and  $N_B$  are the numbers of repeating units of a given type in a given sequence. Figure 1 shows the dependence of the quantity  $\ln Z/N$  on  $N$ . It can be seen that for a sufficiently long chain  $N > 3000$ , the reduced free energy tends to a finite limit. For short chains, the reduced free energy strongly depends on the implementation of the sequence (primary structure). Figure 1 shows the behavior of the reduced free energy for different  $x$  and two temperatures for the same realization.

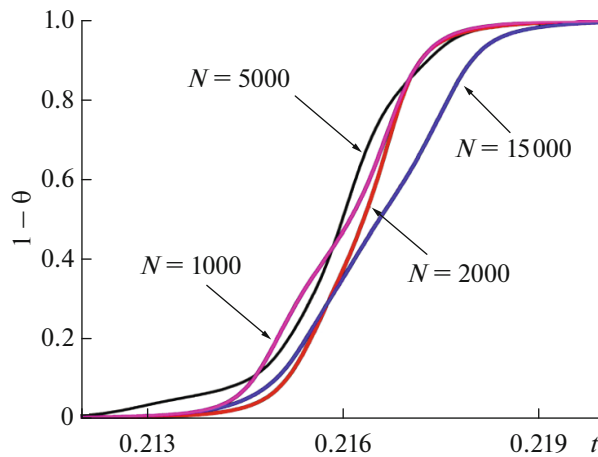
It should be noted that for all graphs the temperature is indicated in the given values  $t = J_A^{-1}$ , this method of introducing the temperature makes it possible to qualitatively compare the graphs of the dependence of the averaged parameters.

With a good degree of accuracy, the number  $N^* \approx 3000$  is confirmed for a wide set of realizations, as evidenced by Fig. 2. For sufficiently large lengths, the free energy does not depend on the realization. Thus, in terms of the reduced free energy, it can be argued at  $N > 3000$  that the heteropolymer behaves like a homopolymer.

Figure 2 shows the behavior of the reduced free energy at the same value of  $x$  and reduced temperature. However, it can be seen that the situation changes upon passing to the degree of helicity.



**Fig. 2.** Dependence of the reduced free energy on the number of repeating units for different sequences  $x_1 = 0.4$ ,  $t = 0.219$ ,  $U_A = 1$ ,  $U_A = 0.8$ ,  $Q_A = 71$ ,  $Q_B = 51$ .



**Fig. 3.** Temperature dependence of  $1 - \theta$  for sequences with different  $N$ .  $x_1 = 0.4$ ,  $\Delta = 4$ ,  $U_A = 1$ ,  $U_A = 0.8$ ,  $Q_A = 71$ ,  $Q_B = 51$ . (The temperature is reduced to the value  $t = J_A^{-1}$ .)

Figure 3 shows denaturation curves for different  $N$ . It can be seen that the position of the curves depends on  $N$ , however, even at  $N = 5000$ , when the reduced free energy no longer depends on the primary structure (realization), the denaturation curve depends.

Figure 4 shows that the general properties of denaturation curves do not depend on the realization, but the details of the profile of the curves vary with the realization.

Figure 5 shows the curves of denaturation at different  $x$ . The result of our computations is  $t_m(x) = xt_{mA} + (1 - x)t_{mB}$ , where  $t_m$  is the transition point.

Let us consider the details of the melting of the HelixI structure. When melting, the macromolecule breaks down into HelixI fragments separated by coil-like sections. The proportion of such fragments is determined by the proportion of joints  $(h, \bar{h})$ , that is, in the GMPC language  $p(\gamma_i = 1, \gamma_{i+1} = k), k \neq 1$ , which is expressed by formula (15).

Figure 6 shows the dependence of the fraction of joints  $\eta$  on the degree of helicity at different  $x$ , including the homopolymer 'A' and 'B'. This dependence was chosen to compare the behavior at different  $x$  because  $\eta(t)$  for different  $x$  will be in different temperature regions. So, we chose the dependence  $\eta(\theta)$ .

As it was clear and as seen in Fig. 6, the curves  $\eta(\theta)$  are the curves with the maximum in the region of  $\theta = 0.5$ . The curve at  $x = 0$  (less stable homopolymer polyB) is located above the curve at  $x = 1$  (the more stable polyA). However, the heteropolymer is located above both homopolymers. This is clear because the heteropolymers have more joint variants. It can be seen from the figure that a higher maximum is obtained

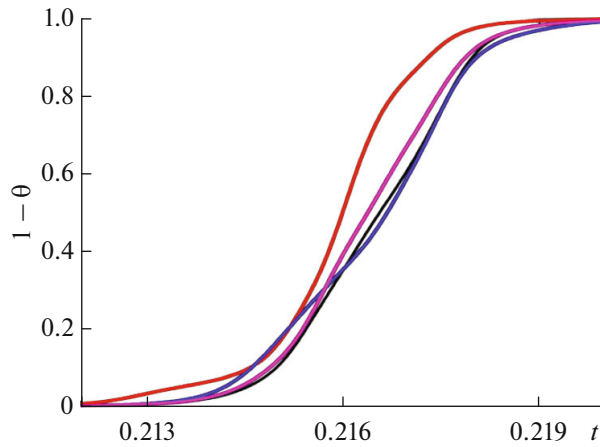


Fig. 4. Temperature dependence of  $1 - \theta$  for different implementations at  $x_1 = 0.4$ ,  $\Delta = 4$ ,  $U_A = 1$ ,  $U_A = 0.8$ ,  $Q_A = 71$ ,  $Q_B = 51$ ,  $N = 5000$ . (The temperature is reduced to the value  $t = J_A^{-1}$ .)

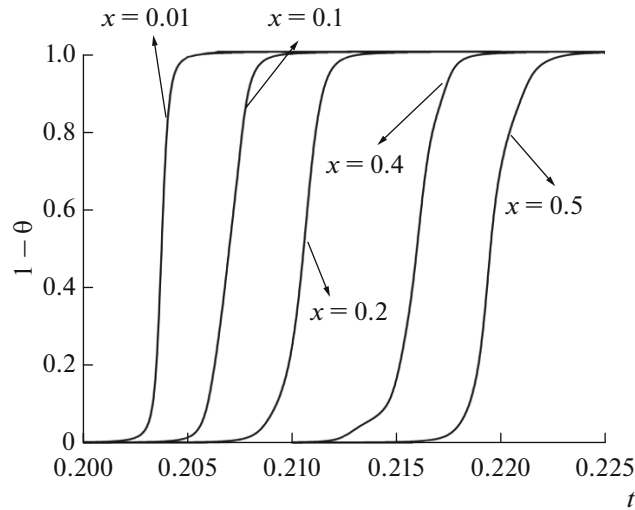


Fig. 5. Temperature dependence of  $1 - \theta$  for sequences with different  $x$  in the thermodynamic limit  $\Delta = 4$ ,  $U_A = 1$ ,  $U_A = 0.8$ ,  $Q_A = 71$ ,  $Q_B = 51$ ,  $N = 5000$ . (The temperature is reduced to the value  $t = J_A^{-1}$ .)

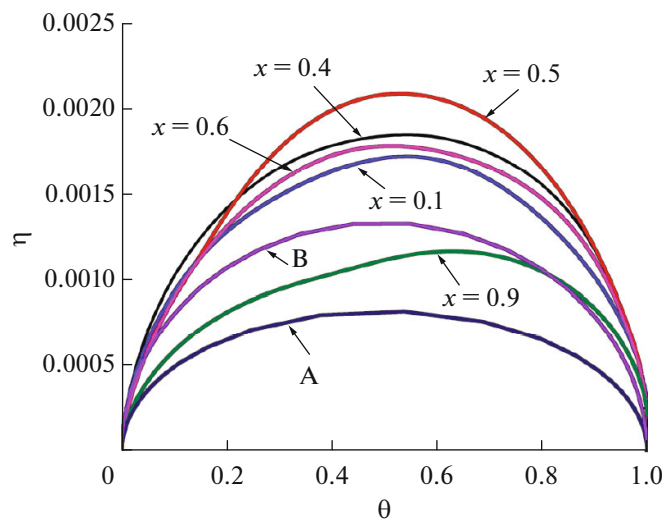
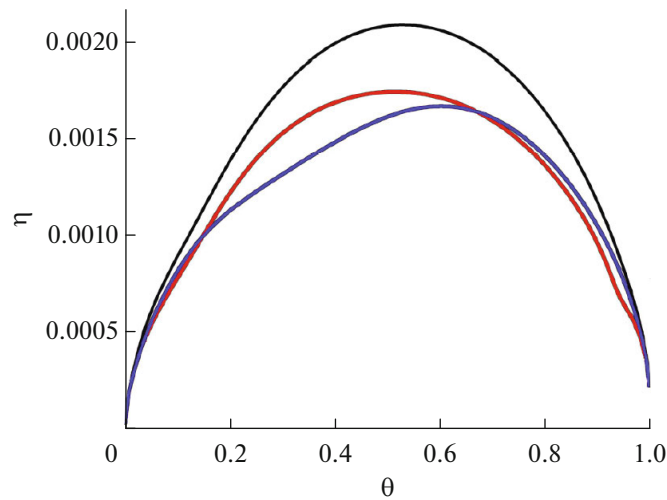
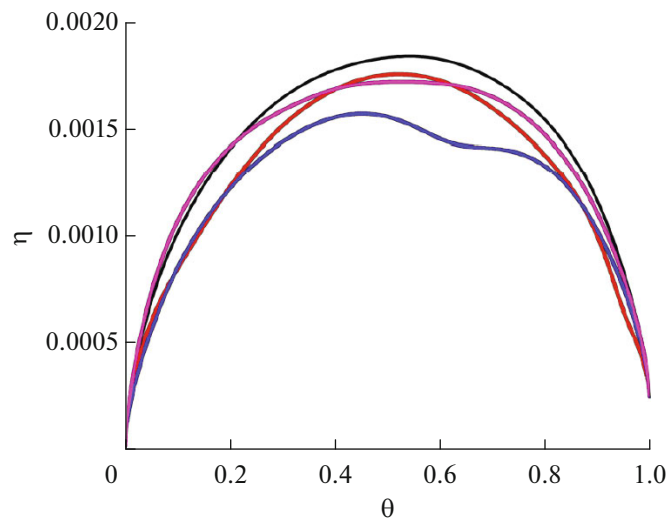


Fig. 6. Dependence  $\eta(\theta)$  for different  $x$ ,  $\Delta = 4$ ,  $U_A = 1$ ,  $U_A = 0.8$ ,  $Q_A = 71$ ,  $Q_B = 51$ ,  $N = 5000$ .



**Fig. 7.** Dependence  $\eta(\theta)$  for three realizations  $\eta(\theta)$ ,  $x = 0.5$ ,  $\Delta = 4$ ,  $U_A = 1$ ,  $U_B = 0.8$ ,  $Q_A = 71$ ,  $Q_B = 51$ ,  $N = 5000$ .



**Fig. 8.** Dependence  $\eta(\theta)$  for four realizations,  $x = 0.5$ ,  $\Delta = 4$ ,  $U_A = 1$ ,  $U_B = 0.8$ ,  $Q_A = 71$ ,  $Q_B = 51$ ,  $N = 5000$ .

at  $x = 0.5$ , that is, at the highest heterogeneity. Nevertheless, the case with  $x = 0.5$  also exhibits a large scatter for different realizations of Fig. 7. The scatter at  $x = 0.4$  is lesser (Fig. 8), and at  $x = 0.1$  and  $x = 0.9$ , the scatter is no longer visible and is not shown in the figures. As for the maximum value, it varies from  $2.2 \times 10^{-3}$  ( $x = 0.5$ ) to  $6 \times 10^{-4}$  ( $x = 1$ ), which corresponds to 11 to 3 joints.

Thus, even with such a superficial consideration, the picture of the transition turns out to be quite paradoxical. With a length of 5000 repeating units, we observe 11 joints for the heteropolymer and on the order of single joints for homopolymers. So, analyzing the behavior of the fraction of joints, we conclude that the length of 5000 repeating units, although it is enough for the free energy of the heteropolymer (and even more so for homopolymer) would go to saturation, and division into sections does not lead to such results. In the recently published work of A. Badasyan [25], using the Zimm-Bragg model as an example, it was shown that even in the homopolymer case, there is an intermediate in length regime associated with the correlation length. In this regard, our next studies will be related to the study of the correlations of conformations in heteropolymers.

#### FUNDING

The research was carried out with the financial support of the State Committee for Science of the Ministry of Education and Science of the Republic of Armenia within the framework of Scientific Project No. 19YR-1F057.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

1. Bloomfield, V.A., *American J. of Phys.*, 1999, vol. 67, p. 1212.
2. Schreck, J.S. and Yuan, J.-M., *Phys. Rev. E*, 2010, vol. 81, p. 061919.
3. Ghosh, K. and Dill. K.A., *J. Am. Chem. Soc.*, 2009, vol. 131, p. 2306.
4. Hausrath, A.C., *Prot. Science*, 2006, vol. 15, p. 2051.
5. Jacobs, D.J. and Wood, G.G., *Biopolymers*, 2004, vol. 75, p. 1.
6. Rayan, G., Tsamaloukas, A.D., Macgregor, R.B., and Heerklotz, H., *J. of Phys. Chem. B*, 2009, vol. 113, p. 1738.
7. Kornilova, S.V., Yasem, P., Grigorev, D.N., et al., *Biofizika*, 1998, vol. 43, no. 1, p. 46.
8. Petrovskaya, E.V., Vasilevskaya, V.V., and Khokhlov, A.R., *Moscow University Physics Bulletin*, 2007, vol. 62, no. 2, p. 99.
9. Benham, C.J., *Phys. Rev. E*, 1996, vol. 53, p. 2984.
10. Einert, T. and Netz, R., *Biophys. J.*, 2010, vol. 99, p. 578.
11. Einert, T., Naeger, P., Orland, H., and Netz, R., *Phys. Rev. Lett.*, 2008, vol. 101, p. 048103.
12. Ananikyan, N.S., Hayryan, Sh.A., Mamasakhlishov, E.Sh., and Morozov, V.F., *Biopolymers*, 1990, vol. 30, p. 357.
13. Hayryan, Sh.A., Mamasakhlishov, E.Sh., and Morozov, V.F., *Biopolymers*, 1995, vol. 35, p. 75.
14. Ananikyan, N.S., Mamasakhlishov, E.Sh., and Morozov, V.F., *Phys. Chem.*, 1993, vol. 27, no. 3, p. 603.
15. Morozov, V.F., Mamasakhlishov, E.Sh., Hayryan, Sh.A., and Chin-Kun, Hu., *Physica A*, 2000, vol. 281, p. 51.
16. Badasyan, A.V., Grigoryan, A.V., Chukhadzhyan, A.Yu., Mamasakhlishov, Y.Sh., and Morozov, V.F., *Izvestiya Natsional'noi Akademii Nauk Armenii, Fizika*, 2002, vol. 37, p. 320.
17. Morozov, V.F., Badasyan, A.V., Grigoryan, A.V., Sahakyan, M.A., and Mamasakhlishov, E.Sh., *Biopolymers*, 2004, vol. 75, no. 5, p. 434.
18. Grigoryan, A.V., Badasyan, A.V., Morozov, V.F., and Mamasakhlishov, E.S., *Izvestiya Natsional'noi Akademii Nauk Armenii, Fizika*, 2004, vol. 39, p. 265.
19. Tonoyan, Sh.A., PhD thesis, Yerevan, Armenia, 2011.
20. Tsarukyan, A.V., PhD thesis, Yerevan, Armenia, 2009.
21. Asatryan, A.V., PhD thesis, Yerevan, Armenia, 2020.
22. Oseledets, V.I. and Mosk, Tr., *Mat. Obs.*, 1968, vol. 19, p. 179.
23. Serva, M. and Paladin, G., *Phys. Rev. Lett.*, 1993, vol. 70, p. 105.
24. Badasyan, A., PhD thesis, Yerevan, Armenia, 2005.
25. Badasyan, A., *Polymers*, 2021, vol. 13, no. 12, p. 1985.

*Translated by V. Musakhanyan*