

# **Algorithm of Calculations of Helix-Coil Transition Parameters in Heteropolymeric Biopolymers on the Bases of Generalized Model of Polypeptide Chain**

A.V. Asatryan

*Department of Physics Yerevan State University, 1 Alex Manoogian, 0025 Yerevan, Armenia*

E-mail: asatryan.arevik@ysu.am

Received 10 February 2018

**Abstract.** Algorithm of calculations of free energy, helicity degree and other parameters of helix-coil transition theory in the model of heterogen biopolymers with binary heterogeneity has been built. The theory is based on Generalized Model of Polypeptide Chain (GMPC). In the basics of algorithm is laying direct calculation of free energy with the help of random product of transfer matrices of homopolymeric GMPCs.

**Keywords:** algorithm, helix-coil transition, heteropolymer, GMPC

## **1. Introduction**

Significant amount of articles about helix-coil transition in biopolymers have been published since 1960s [1-7] until nowadays [8-15]. Traditionally, theoretical models of helix-coil transition are based on one or another modification of Ising model [16-19] and are mean field models. However, some theoretical non-mean field models for polypeptides [20-22] and DNA [9-13] have been represented before. In addition, a model describing helix-coil transition based on Potts model modification with multi-particle interactions has been introduced in our previous publications. This model is called generalized model of polypeptide chain (GMPC) [24-28]. Furthermore, we have discussed helix-coil transition problem in biopolymers with heterogeneity by energies of base pair formation [34] and by number of conformations [35] with application of constrained annealing method [30].

The current publication is intended for analytic description of transition curves in heterogenic biopolymers with two types of repeated units. GMPC, which has been applied multiple times [24-28, 34-35] is used not only to describe homopolymeric problem but also for heteropolymeric case with application of microcanonical distribution [29]. In our previous papers, all the calculations for heterogenic biopolymers have been done with the application of GMPC and constrained annealing method. In contrary, in this case calculations have been done by straight multiplication of transfer-matrices, which differ from each other according to the type of repeated unit. This method is unique in one respect. Although, these calculations are in the framework of GMPC, there are no approximations and importation of special parameters to simplify the calculations. Hence, the results obtained according to this method can be used as patterns for other theoretical results to compare with. The algorithm can be applied not only for base model, but also for model in solvent and at the presence of ligand.

The main goal of current article is to work out detailed scheme for the theoretical calculations mentioned above and make it possible to calculate such characters as helicity degree, average length of helix region, free energy, etc. Moreover, we expect to gain 15000 for length of biopolymers, with little additions to the following algorithm in the future. The last operation will be possible due to exporting and collecting data while we calculate characteristics for shorter biopolymers. All these steps may lead us to outcomes comparable with experimental data.

## 2. Calculation of Parameters Used in the Algorithm

Let us consider heteropolymer which contains two types of repeated units (ex. *AT* and *GC* pairs for DNA and like GO model [23] for proteins). Thus, we will present heteropolymer as a random sequence of *A* and *B* types repeated units. At the same time it is assumed that types of repeated units are being chosen statistically independent. In the frame of GMPC, the Hamiltonian of heteropolymeric system appears as:

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

where  $\beta = T^{-1}$  is inverse temperature,  $J$  is energy of hydrogen bonds' formation within molecules,  $N$  is number of repeated units and  $\delta_i^{(\Delta)} = \prod_{k=\Delta-1}^0 \delta(\gamma_{j-k}, 1)$ , where  $\delta(\gamma_i, 1)$  is Kronecker symbol:  $\gamma_i = 1; 2; \dots; Q$ ,  $Q$  is number of conformations for  $i$ -th repeated unit.

This differs from homopolymeric model by hydrogen bond's energy, which depends on type of repeated unit in this case. Free energy of such system expresses itself as:

$$F = -k_B T \langle \ln \text{Tr} \prod_{i=1}^N G_i \rangle. \quad (2)$$

Here  $k_B$  is Boltzmann constant and  $G_i$  is transfer-matrix for each component:

$$G_i = \begin{pmatrix} e^{J_i} & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & e^{K_i} \\ 1 & 1 & 1 & 1 & \dots & 1 & e^{K_i} \end{pmatrix} \quad (3)$$

In this paper, we are going to examine application of GMPC on biopolymers with bimodal heterogeneity by both energies of helix structure formation and by number of conformations of repeated units. Let us introduce parameter  $\sigma_i$ , which takes value 1 with probability  $X$ , and  $-1$  with probability  $(1-X)$ . So, we are going to deal with two-component heteropolymer with various values of intra-molecular hydrogen bond's energies:  $J = J_0 + \Delta J \sigma$ ,  $J_A = J_0 + \Delta J$  and  $J_B = J_0 - \Delta J$ . We will analogically deal with  $Q$ . Let us introduce  $K = \ln(Q-1)$  for the convenience. Consequently,  $K = K_0 + \Delta K \sigma$ ,  $K_A = K_0 + \Delta K$ ,  $K_B = K_0 - \Delta K$ . As a result, primary structure of biopolymer will appear as sequence of numbers  $\pm 1$ . According to (2), reduced free energy calculated as:

$$f = -\frac{F}{NkT} = \frac{\ln \text{Sp} \prod_{i=1}^N G_i}{N}, \quad (4)$$

where  $G_i$  defines according to primary sequence as transfer-matrix  $G_i = G(J = J_0 + \Delta J \sigma, K = K_0 + \Delta K \sigma)$ .

For these parameters -  $X, N, J_0, \Delta J, K_0, \Delta K$  for each realization of random sequence reduced free energy can be reached. In case of product of many matrices, we use the following technique to avoid very big numbers. Since, in case of quite big lengths of homopolymer reduced free energy approaches to value  $\ln \lambda_1$ , then we will introduce reduced matrix as follows:

$$G = \lambda_1 g \quad (5)$$

Consequently:

$$f = X \ln \lambda_{1A} + (1-X) \ln \lambda_{1B} + \frac{\ln Sp \prod_{i=1}^N g_i}{N}, \quad (6)$$

where  $\lambda_{1A}$  and  $\lambda_{1B}$  are the first eigenvalues for homopolymers with type  $A$  and type  $B$  repeated units, respectively.

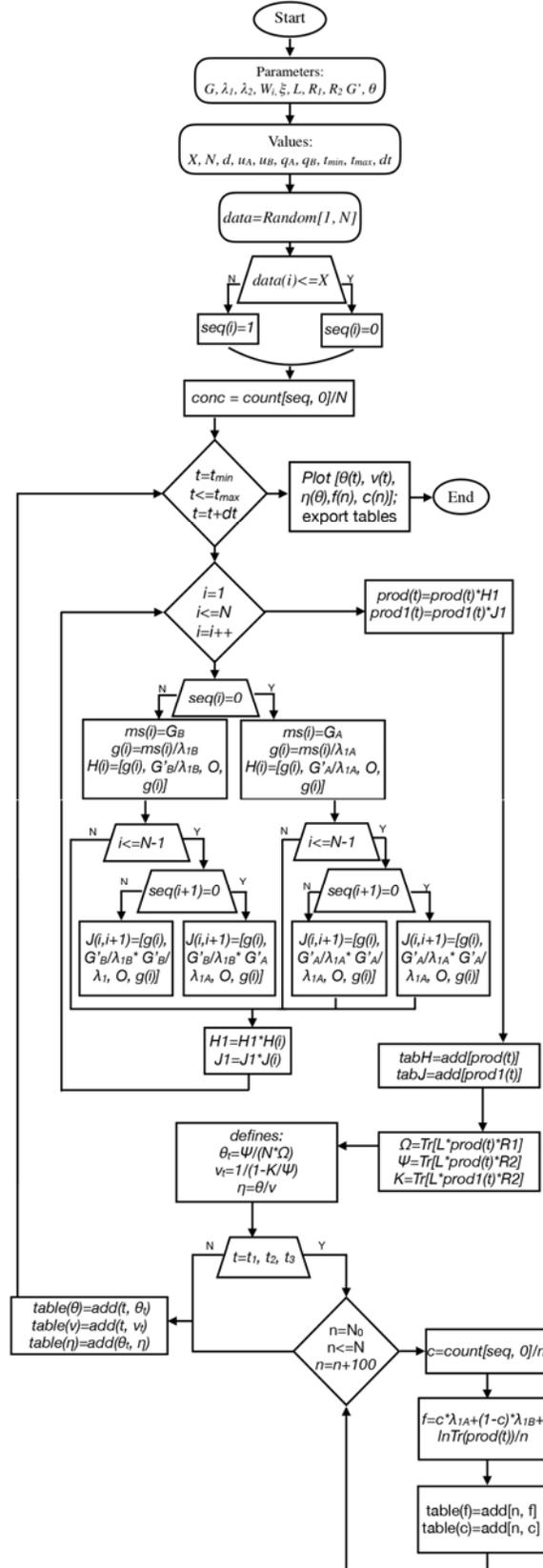
In order to calculate helical fraction  $\langle \delta_i^{(\Delta)} \rangle$ , we use the following expression:

$$\theta_N = \langle \delta_i^{(\Delta)} \rangle = \frac{1}{NZ} \sum_i \delta_i^{(\Delta)} \exp(-\beta H) = \frac{\partial \ln Z}{\partial J_0}. \quad (7)$$

### 3. Mechanisms and Applications of Algorithm

The flowchart of algorithm is presented below. Here we are going to discuss the entire algorithm step-by-step. In the first step we define parameters  $G_i, \lambda_1, \lambda_2, W_i, \xi, L, R_1, R_2, DW, \theta$  giving their formulas of calculation. Where  $\lambda_1$  and  $\lambda_2$  are the first and the second eigenvalues, respectively,  $W_i = \exp(u_i/T)$  is the reduced energy of hydrogen bond formation, which depends on type of repeated unit in case of heteropolymer and on temperature,  $\xi = \frac{1}{\ln \frac{\lambda_1}{\lambda_2}}$  is correlation length,

$L, R_1, R_2$  are the left and right matrices respectively,  $G' = \frac{\partial G_i}{\partial J_0}$  is the transfer matrix of a repeated unit in fixed helix state and finally  $\theta$  is helicity degree as it is already known. Secondly, all the numeric parameters:  $X, N, d, u_A, u_B, q_A, q_B, t_{\min}, t_{\max}, dt$  are given their values. Here  $X$  is a probability of  $A$  type repeated unit's concentration,  $N$  is a number of repeated units of a heteropolymer,  $d$  is a topological parameter which shows the number of repeated units needed to form hydrogen bond,  $u_A$  and  $u_B$  are hydrogen bonds' formation energies for type  $A$  and type  $B$  repeated units respectively,  $q_A$  and  $q_B$  are correspondingly the numbers of conformations of type  $A$  and type  $B$  repeated units,  $t_{\min}$  and  $t_{\max}$  are limits for temperature scale and  $dt$  is the step of the cycle by temperature. After having all parameters set up, we are going right to the functional part of the algorithm. Firstly, we build an array with  $N$  numbers randomly chosen from  $[0, 1]$  scale. Next, with the help of *if* function and probability  $X$  we get the sequence of heteropolymer with bimodal heterogeneity and calculate the concentration of type  $A$  repeated unit. After these steps,



we start cycle of temperature and continue all the calculations for each temperature in that cycle. Moreover, we use another cycle in the mentioned one, but this cycle is for sequence. Calculations in this cycle are for each repeated unit. In the first step of this cycle, we introduce transfer matrix, reduced transfer matrix and super-matrix according to the type of repeated unit. This filtration of parameters by repeated units is done with help of function *if*. As we are going to form another super-matrix for the calculation of average length of helix state, where we will use not only the transfer-matrix of *i*-th repeated unit, but also the one for *i*+1-th repeated unit in helix state, then in the next step we will use again the function *if* to check if the *i*-th repeated unit is not the last one. Moreover, we will check the type of *i*+1-th repeated unit and form the super-matrix according to that type. Later, we calculate the product one and super-matrices separately and come back to the beginning of the cycle. Then, we repeat all the steps and afterwards, there is a product of last and next super-matrices instead of product of one and super-matrix. Subsequently, when the cycle is over and from now on all the operations are going to take place in the cycle by temperatures, we reach the next step. In this step, we multiply identity matrix and the product of all super-matrices we have got from cycle. This operation refers both to product of super-matrices for helicity degree:  $H1$ , and to product of super-matrices for average length of helix state:  $J1$ . These two products we will call  $prod(t)$ -for helicity degree and  $prod1(t)$ -for average length of helix state.

For each temperature we add  $prod(t)$  to some table called  $tabH$ , and  $prod1(t)$  to a table called  $tabJ$ . In the next two steps, we calculate helicity degree, average length of helix state and number of junctions. Then we use function *if* to check if the temperature is equal to the chosen ones, and if it is we start another cycle by number of repeated units:  $n$ . In this cycle we calculate concentration and free energy for given number of repeated units and add the results in  $table(c)$  and  $table(f)$ , respectively. The formula of free energy -  $f$  appears in flowchart and for concentration we have  $c = n_A/n$ , where  $n$  is the total number of repeated units and  $n_A$  is the number of A type repeated units which varies from one iteration to another one. The step for this cycle is  $dn = 100$ . And we repeat this calculations as long as  $n \leq N$ . When the cycle is finished, we lead to the next step, which is also the step after last function *if* in case of temperature is not one of the chosen ones. In this step, we add  $(t, \theta_t)$  to  $table(\theta)$ ,  $(t, \nu)$  to  $table(\nu)$  and  $(\theta, \eta)$  to  $table(\eta)$ . Then we repeat all the mentioned calculations until  $t \leq t_{max}$ . Finally, when this cycle is over too, we make graphs of  $\theta(t), \nu(t), \eta(\theta), f(n), c(n)$  with the help of tables and export these result as arrays.

#### 4. Results and Discussions

According to this flowchart discussed above, the program has been built on Wolfram Mathematica. We have built  $\theta(t), \nu(t), \eta(\theta), f(n), c(n)$  dependences within this program and process graphs in OriginLab8. The dependence of helicity degree from temperature:  $\theta(t)$ , and its differential:  $d\theta/dt$  are some of the obtained results in that program. Results are presented in following figures.

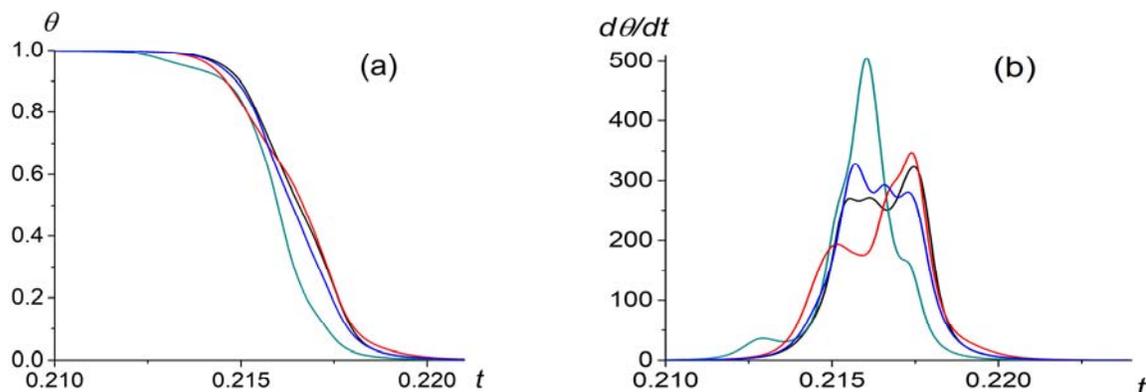


Figure 2. The temperature dependence of (a) helicity degree- $\theta(t)$  and (b) its differential- $d\theta/dt$  for various realizations of repeated units' sequence.

One can see from the figures that curves strongly depend on sequence realizations. It is more obvious from differential curves. To compare theoretical and experimental results we need to calculate average from numerous realizations.

We have used various values for parameters, such as “number of repeated units”, “probability of concentration for type  $A$  repeated unit”, “temperature region” and obtained results according to the given values. Although, some numbers of mentioned dependences have been built, we need more outcomes to get to actual analysis and scientific results. So far, we are going to use other values for the parameters we have not changed before. These are “energies of hydrogen bonds” formation for each type of repeated units”, “numbers of conformations”, etc. Moreover, exporting all the tables from program in Wolfram Mathematica as data files, particularly *tabH* and *tabJ* (which contain the products of super-matrices from the whole cycle and the temperature used to calculate these products during each iteration) will allow us to merge several sequences and calculate these and other parameters for much longer biopolymers quite quickly in other program. Therefore, this algorithm as well as the program will give us an opportunity of getting qualitatively new results in the future.

## 5. Conclusions

As a result, the algorithm will allow taking into account interactions with solvent by redefinition of parameters  $W_i$  and  $Q_i$ . In addition, it will give the opportunity to calculate other transition parameters such as  $\xi$ : correlation length. This program makes possible to analyze helix-coil transition in heteropolymers with correlation in the sequence by corresponding generation. With the help of the program, we are able to merge various sequences and calculate these parameters for longer biopolymers quite quickly. Moreover, it will allow to investigate block-structured biopolymers.

## References

- [1] D.C.Poland and H.A. Scheraga. The Theory of Helix-Coil Transition, Academic, New York, 1970.
- [2] Yu. Grosberg and A.R. Khokhlov. Statistical Physics of Macromolecules, AIP, New York, 1994.
- [3] R. Cantor and T.R. Shimmel. Biophysical Chemistry, Freeman, San Francisco, 1980.
- [4] P.J. Flory. Statistical Mechanics of Chain Molecules, Interscience, New York, 1969.
- [5] R.M. Wartell and A.S. Benight. Phys.Rep., 126, 67 (1985)
- [6] A.A. Vedenov, A.M. Dykne, and M.D. Frank-Kamenetski. Usp. Fiz. Nauk (Russian), 105, 479 (1971).
- [7] Wada and A. Suyama. Prog. Biophys. Mol. Biol., 47, 113 (1986).
- [8] T. Chalikian. Biopolymers, 70, 492 (2003).
- [9] T. Garel, C. Monthus, and H. Orland. Europhys. Lett., 55, 132 (2001).
- [10] D. Cule and T. Hwa. Phys. Rev. Lett., 79, 2375 (1997).
- [11] M. Baiesi, E. Carlon, E. Orlandini, and A.L. Stella. e-print cond-mat/0207122.
- [12] M. Peyrard. Nonlinearity 17, R1 (2004).
- [13] M. Barbi, S. Lepri, M. Peyrard, and N. Theodorakopoulos. Phys. Rev. E, 68, 061909 (2003).
- [14] M. Takano, K. Nagayama, and A. Suyama. J.Chem/Phys., 116, 2219 (2001).
- [15] V. Munoz and L. Serrano. Biopolymers, 41, 495 (1997).
- [16] B.H. Zimm, P. Doty, and K. Iso. Proc. Natl. Acad. Sci. USA, 45, 1601 (1959).
- [17] B.H. Zimm and J.K. Bragg. J. Chem. Phys., 31, 526 (1959).
- [18] B.H. Zimm. J. Chem. Phys., 33, 1349 (1960).
- [19] B.H. Zimm and N. Rice. Mol. Phys., 3, 391 (1960).
- [20] S. Lifson and A. Roig. J. Chem. Phys., 34, 1963 (1961).
- [21] S. Lifson and B.H. Zimm. Biopolymers, 1, 15 (1963).
- [22] S. Lifson and J. Allegra. Biopolymers, 2, 65 (1964).
- [23] G. N. and Abe H., *Biopolymers*, 20(5), 991-1011, (1981)
- [24] V.F. Morozov, E. Sh. Mamasakhlisov, Sh. A. Hayryan, and C.-K. Hu. Physica A, 281, 51 (2000).
- [25] V.F. Morozov, A.V. Badasyan, A.V. Grigoryan, M.A. Sahakyan, and E. Sh. Mamasakhlisov. Biopolymers, 75, 434 (2004).
- [26] A.V. Tsarukyan, Sh. A. Tonoyan, A.V. Badasyan, A.V. Grigoryan, E. Sh. Mamasakhlisov, V.F. Morozov. J. Contemp. Phys. (Armenian Ac. Sci.), Physics, 41, 63 (2006).
- [27] Sh. A. Tonoyan, T.Y. Buryakina, A.V. Tsarukyan, E. Sh. Mamasakhlisov, V.F. Morozov. J. Contemp. Phys. (Armenian Ac. Sci.), Physics, 42, 466 (2007).
- [28] Sh. A. Tonoyan, A.S. Mirzakhanyan, G.N. Hayrapetyan, A.V. Tsarukyan, V.F. Morozov. J. Contemp. Phys. (Armenian Ac. Sci.), Physics, 45, 133 (2010).
- [29] A.V. Badasyan, A.V. Grigoryan, E. Sh. Mamasakhlisov, A.S. Benight, and V.F. Morozov. J. Chem. Phys., 123, 194701 (2005).
- [30] M. Serva, and G. Paladin. Phys. Rev. Lett., 70, 105 (1993).
- [31] A.V. Badasyan, Sh. A. Tonoyan, A.V. Tsarukyan, E. Sh. Mamasakhlisov, A.S. Benight, and V.F. Morozov. J. Chem. Phys., 128, 195101 (2008).
- [32] M.D. Frank-Kamenetski, A.V. Vologodski. Nature, 269, 729 (1977).
- [33] S. Cuesta-Lopez. Europhys. Lett., 87, 48009 (2009).
- [34] Sh. A. Tonoyan, A.S. Mirzakhanyan, Y.M. Mamasakhlisov, V.F. Morozov. J. Contemp. Phys. (Armenian Ac. Sci.), Physics, 48, 350, (2013).
- [35] Sh. A. Tonoyan, A.V. Asatryan, Y.M. Mamasakhlisov, V.F. Morozov. J. Contemp. Phys. (Armenian Ac. Sci.), Physics, 43, 209, (2014).