

UDC: 577.323

## Trajectories of the DNA Kinks in the Sequences Containing CDS Regions

Yakushevich L.V.<sup>\*1</sup>, Krasnobaeva L.A.<sup>\*\*2,3</sup>

<sup>1</sup>*Institute of Cell Biophysics, Pushchino, Moscow region, Russia*

<sup>2</sup>*Siberian State Medical University, Tomsk, Russia*

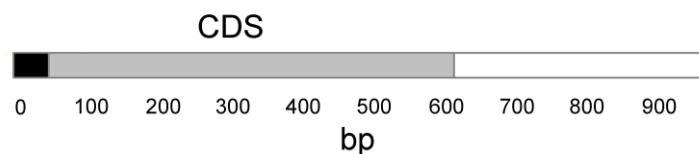
<sup>3</sup>*Tomsk State University, Tomsk, Russia*

**Abstract:** Coding regions (CDS) being an integral part of any gene sequence, play an important role in the process of transcription. One of the tasks associated with the CDS regions, consists in the modeling of the passage of transcription bubbles named also open states or DNA kinks through the coding regions. In this paper, we present a simple and convenient approach to the modeling of the passage. It includes the calculation of the energy profile of the sequence and reducing the initial task to the modeling of the movement of a quasi particle in the field with this energy profile. To illustrate the method, we present the results of the calculations of the trajectories of the DNA kinks moving in the sequence of gene coding interferon alpha 17 (IFNA17) that consists of the three regions: the coding region and the two regions with unknown functional properties. To analyze the kink dynamics, we apply approximation where the DNA parameters are being averaged separately over each of the three regions. In the absence of dissipation, the total kink energy is constant. At the same time the kink velocity is constant only inside each of the regions. In the presence of dissipation, the total kink energy decreases. It is shown that the greater the total initial energy of the kink, the faster the energy decrease. It is suggested that the proposed approach could be useful in finding the ways to govern the movement of transcription bubbles at the first stage of the process of transcription.

**Key words:** DNA dynamics, kink trajectories, energy profile, CDS, IFNA17.

### INTRODUCTION

Coding regions (CDS) being an integral part of any gene sequence, play an important role in the process of transcription. The sequence of the gene coding interferon alpha 17 (IFNA17), is a simple example of the sequence containing only one CDS region [1, 2]. Indeed, according to GenBank [3], this sequence contains three different regions: the CDS region (50..619) and the two regions (1..49 and 620..980) with unknown functional properties (Fig. 1).



**Fig. 1.** Three regions in the sequence of the gene coding interferon alpha 17 (IFNA17). The coding region (CDS) is shown by grey color, the region to the left of the coding region is shown by black color, and the region to the right – by white.

\*yakushev@icb.psn.ru

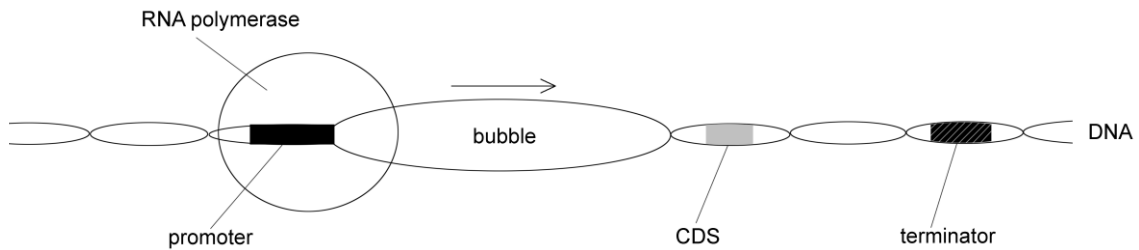
\*\*kla1983@mail.ru

The details of the gene structure are presented in Table 1. There  $N_j$  is the total number of nitrous bases in the  $j$ -th region,  $N_{j,A}$ ,  $N_{j,T}$ ,  $N_{j,G}$  and  $N_{j,C}$  are the numbers of adenines, thymines, guanines and cytosines in the  $j$ -th the region,  $j = 1, 2, 3, 4$ .

**Table 1.** Details of the structure of the sequence of the gene coding alpha interferon 17

Region	Coordinates of the region	$N_{j,A}$	$N_{j,T}$	$N_{j,G}$	$N_{j,C}$	$N_j$
1	1..49	15	12	7	15	49
2 (CDS)	50..619	157	145	130	138	570
3	620..980	110	146	44	61	361

One of the most interesting tasks associated with the CDS regions, is the modeling of the passage of transcription bubble which is named also open state [4, 5], local unwound region [6, 7] and DNA kink [8, 9], through the CDS region (Fig. 2). Further we shall use the latter term.



**Fig. 2.** Schematic picture of the DNA double chain with the moving transcription bubble.

Surprisingly, the passage of the DNA kinks through the CDS region was not modeled by investigators, though the necessary mathematical formalism has been already developed. We can point out the model of Peyrard and Bishop [10] that takes into account the transverse motions of nucleotides, the Y-model [11–13] that takes into account the angular displacements of nitrous bases, the combined model [14] that takes into account both the angular and the transverse displacements, and numerous modifications of these models [15–20]. Any of them could be used for the purpose. In this paper, we use one of the recent modifications of the Y-model [21, 22]. To model the movement of the DNA kink, we apply the method of trajectories that includes the calculation of the energy profile of the gene.

### MODEL AND METHODS

In the Y-model, angular displacements of nitrogenous bases are modeled by the system of 2N coupled nonlinear differential equations [21]:

$$I_{1,n} \frac{d^2 \varphi_{n,1}(t)}{dt^2} - K'_{1,n} [\varphi_{n+1,1}(t) - 2\varphi_{n,1}(t) + \varphi_{n-1,1}(t)] + k_{1-2,n} R_{1,n} (R_{1,n} + R_{2,n}) \sin \varphi_{n,1} - k_{1-2,n} R_{1,n} R_{2,n} \sin(\varphi_{n,1} - \varphi_{n,2}) = -\beta_{1,n} \frac{d\varphi_{n,1}(t)}{dt}, \quad (1)$$

$$I_{2,n} \frac{d^2 \varphi_{n,2}(t)}{dt^2} - K'_{2,n} [\varphi_{n+1,2}(t) - 2\varphi_{n,2}(t) + \varphi_{n-1,2}(t)] + k_{1-2,n} R_{2,n} (R_{1,n} + R_{2,n}) \sin \varphi_{n,2} - k_{1-2,n} R_{1,n} R_{2,n} \sin(\varphi_{n,2} - \varphi_{n,1}) = -\beta_{2,n} \frac{d\varphi_{n,2}(t)}{dt}. \quad (2)$$

Here  $\varphi_{i,n}(t)$  and  $I_{i,n}$  are the angular displacement and the moment of inertia of the  $n$ -th base of the  $i$ -th chain,  $R_{i,n}$  is the distance between the center of mass of the  $n$ -th base of the  $i$ -th chain and the nearest sugar-phosphate chain,  $K'_{i,n}$  is the coefficient of torsion rigidity of the sugar-phosphate chain in the vicinity of the  $n$ -th base of the  $i$ -th chain,  $\beta_{i,n} = \alpha R_{i,n}^2$ ,  $\alpha$  is the coefficient of dissipation,  $k_{1-2,n}$  is the force constant that characterizes interactions between bases in pairs;  $i = 1, 2$ ;  $n = 1, 2, \dots, N$ . The values of the coefficients of these equations were first estimated in the paper [4]. Later a more complete and accurate set of parameters was presented in [22]. Here we use even more precise set of parameters (Table 2) that has been recently presented in [23].

**Table 2.** Model parameters

Type of the base	$I \times 10^{-44}$ (kg·m <sup>2</sup> )	$R$ (Å)	$K' \times 10^{-18}$ (J)	$k_{1-2} \times 10^{-2}$ (N/m)	$\beta \times 10^{-34}$ (J·s)
A	7.61	5.80	2.35	6.20	4.25
T	4.86	4.80	1.61	6.20	2.91
G	8.22	5.70	2.27	9.60	4.10
C	4.11	4.70	1.54	9.60	2.79

### Continuum approximation

Assuming that angular displacements  $\varphi_{i,n}(t)$  are smooth functions, we rewrite equation (1)–(2) in the continuum approximation:

$$I_1(z)\varphi_{1tt} - K'_1(z)a^2\varphi_{1zz} + k_{1-2}(z)R_1(z)(R_1(z) + R_2(z))\sin\varphi_1 - k_{1-2}(z)R_1(z)R_2(z)\sin(\varphi_1 - \varphi_2) = -\beta_1(z)\varphi_{1t}, \quad (3)$$

$$I_2(z)\varphi_{2tt} - K'_2(z)a^2\varphi_{2zz} + k_{1-2}(z)R_2(z)(R_1(z) + R_2(z))\sin\varphi_2 - k_{1-2}(z)R_1(z)R_2(z)\sin(\varphi_2 - \varphi_1) = -\beta_2(z)\varphi_{2t}. \quad (4)$$

Here the functions  $\varphi_{i,n}(t)$  have been transformed to  $\varphi_i(z, t)$ . The coefficients have become the functions of the variable  $z$ :  $I_{i,n} \rightarrow I_i(z)$ ,  $K'_{i,n} \rightarrow K'_i(z)$ ,  $k_{1-2,n} \rightarrow k_{1-2}(z)$ ,  $R_{i,n} \rightarrow R_i(z)$ ,  $\beta_{i,n} \rightarrow \beta_i(z)$ , and the expression  $[\varphi_{n+1,1}(t) - 2\varphi_{n,1}(t) + \varphi_{n-1,1}(t)]$  has taken the form:  $a^2\varphi_{izz}$ .

### Average field approximation

We consider the angular fluctuations of nitrous bases of one of the two polynucleotide chains in the average field induced by the second polynucleotide chain. Then equation can be written as:

$$I_1(z)\varphi_{1tt} - K'_1(z)a^2\varphi_{1zz} + k_{1-2}(z)R_1(z)(R_1(z) + R_2(z))\sin\varphi_1 - k_{1-2}(z)R_1(z)R_2(z)\sin(\varphi_1 - \langle\varphi_2\rangle) = -\beta_1(z)\varphi_{1t} \quad (5)$$

If we take into account that

$$\sin(\varphi_1 - \langle\varphi_2\rangle) = \sin\varphi_1 \cos(\langle\varphi_2\rangle) - \cos\varphi_1 \sin(\langle\varphi_2\rangle) = \sin\varphi_1,$$

then equation (5) can be rewritten in the form:

$$I_1(z)\varphi_{1tt} - K'_1(z)a^2\varphi_{1zz} + k_{1-2}(z)R_1^2(z)\sin\varphi_1 = -\beta_1(z)\varphi_{1t}. \quad (6)$$

Acting in the same way, we can rewrite equation (4) in the form:

$$I_2(z)\varphi_{2tt} - K'_2(z)a^2\varphi_{2zz} + k_{1-2}(z)R_2^2(z)\sin\varphi_2 = -\beta_2(z)\varphi_{2t}. \quad (7)$$

In further consideration, it is enough to consider the problem (6). The problem (7) can be obviously resolved in a similar manner.

### Method of McLaughlin and Scott

In the particular case of artificial homogeneous sequence, equation (6) transforms to the sine-Gordon equation with additional term that simulates effects of dissipation:

$$I_1 \varphi_{1tt} - K_1' a^2 \varphi_{1zz} + k_{1-2} R_1^2 \sin \varphi_1 = -\beta_1 \varphi_{1t}. \quad (8)$$

If dissipation is small, equation (8) can be solved by method of McLaughlin and Scott [24]. In the frameworks of the method the solution has the form of kink:

$$\varphi_{1,k}(z, t) = 4 \operatorname{arctg} \{ \exp[(\gamma_1 / d_1) \cdot (z - v_{1,k}(t)t)] \}, \quad (9)$$

moving with the velocity  $v_k(t)$  that is determined by the equation:

$$\frac{dv'_{1,k}(t)}{dt} = -\frac{\beta_1}{I_1} v'_{1,k}(t) (1 - v_{1,k}'^2(t)), \quad (10)$$

where  $v'_{1,k}(t) = \frac{v_{1,k}(t)}{C_{01}}$  is the relative velocity of the DNA kink,  $C_{01} = (K_1' a^2 / I_1)^{1/2}$  is the sound velocity in DNA,  $d_1 = (K_1' a^2 / V_1)^{1/2}$  is the size of the kink. The total energy of the kink is determined by the formula:  $E_1(t) = E_{01} (1 - v_{1,k}'^2)^{-1/2}$ , where  $E_{01} = 8(K_1' V_1)^{-1/2}$  is the rest energy of the kink.

The solution of equation (10) can be written explicitly [25]:

$$v'_{1,k}(t) = \frac{v'_{01} \gamma_{01} \exp\left(-\frac{\beta_1}{I_1}(t-t_0)\right)}{\sqrt{1 + (v'_{01} \gamma_{01})^2 \exp\left(-\frac{2\beta_1}{I_1}(t-t_0)\right)}}, \quad (11)$$

where  $v'_{01} = v'_{1,k}(t_0)$  is the relative kink velocity at the initial time  $t_0$ ,  $\gamma_{01} = (1 - v_{01}'^2)^{-1/2}$ .

### Method of concentrations

To make it possible to apply the approach of McLaughlin and Scott for inhomogeneous sequences, the method of concentrations can be used [26, 27]. According to the method, the coefficients of the initial model equations are averaged in the following way:

$$\begin{aligned} \bar{I}_i &= I_A C_{A,i} + I_T C_{T,i} + I_G C_{G,i} + I_C C_{C,i}, \\ \bar{R}_i &= R_A C_{A,i} + R_T C_{T,i} + R_G C_{G,i} + R_C C_{C,i}, \\ \bar{K}'_i &= K'_A C_{A,i} + K'_T C_{T,i} + K'_G C_{G,i} + K'_C C_{C,i}, \\ \bar{k}_{1-2} &= k_{A-T}(C_{A,1} + C_{T,2}) + k_{G-C}(C_{G,1} + C_{C,2}), \\ \bar{\beta}_i &= \beta_A C_{A,i} + \beta_T C_{T,i} + \beta_G C_{G,i} + \beta_C C_{C,i}, \end{aligned} \quad (12)$$

where  $C_{j,i} = N_{j,i}/N$  is the concentration of nitrous bases of the  $j$ -th type ( $j = A, T, G, C$ ) in the  $i$ -th chain ( $i = 1, 2$ );  $N_{j,i}$  is the numbers of nitrogenous bases of the  $j$ -th type in the  $i$ -th chain;  $N$  is the total number of bases in the gene coding interferon-alpha 17.

Then initially inhomogeneous problem is reduced to homogeneous equation (8) but with the new coefficients determined by formulas (12). It is obviously that the solution in this case takes the form:

$$\bar{v}'_{1,k}(t) = \frac{\bar{v}'_{01} \bar{\gamma}_{01} \exp\left(-\frac{\bar{\beta}_1}{\bar{I}_1}(t-t_0)\right)}{\sqrt{1 + (\bar{v}'_{01} \bar{\gamma}_{01})^2 \exp\left(-\frac{2\bar{\beta}_1}{\bar{I}_1}(t-t_0)\right)}}, \quad (13)$$

where  $\bar{v}'_{1,k}(t) = \frac{v_{1,k}(t)}{C_{01}}$ ,  $\bar{C}_{01} = (\bar{K}'_1 a^2 / \bar{I}_1)^{1/2}$ ,  $\bar{v}'_{01} = \bar{v}'_{1,k}(t_0)$ ,  $\bar{\gamma}_{01} = (1 - \bar{v}_{01}^2)^{-1/2}$ .

### Method of blocks

Formula (13) takes into account composition of the gene sequence but do not consider the arrangement of the nitrous bases in the sequence. The method of blocks permits approximately to take into account the arrangement. In the case of a gene encoding interferon alpha 17, the sequence contains three regions (or blocks). In the method of blocks, coefficients of the model equations are averaged not over all length of the sequence, but over the length of each of the three blocks.

**Table 3.** Averaged model parameters

Region	$\tilde{I} \times 10^{-44}$ (kg·m <sup>2</sup> )	$\tilde{K}' \times 10^{-18}$ (N·m)	$\tilde{V} \times 10^{-20}$ (N/m)	$\tilde{\beta} \times 10^{-34}$ (J·s)
1	5.95	1.91	2.08	3.45
2 (CDS)	6.20	1.95	2.16	3.52
3	5.98	1.90	1.95	3.44

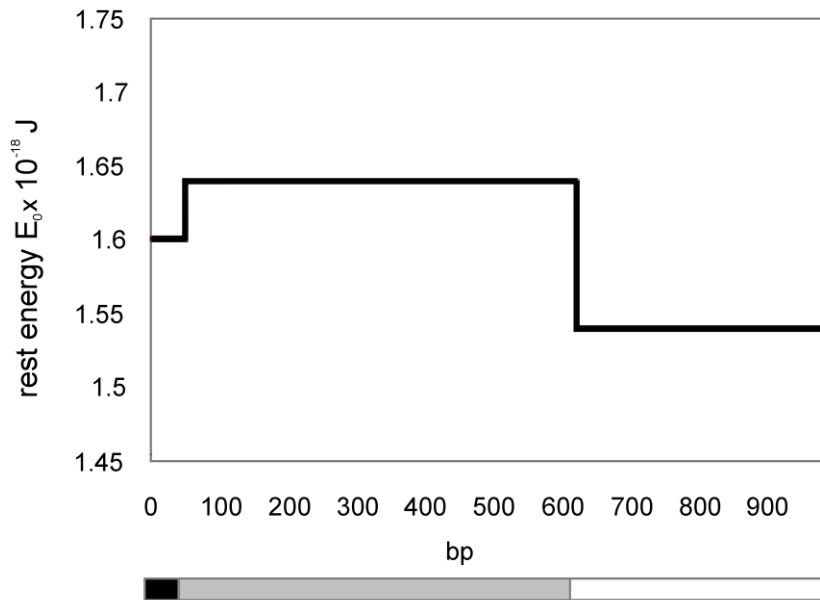
Inside each of the blocks the problem can be easily solved by the method of concentrations. To “stitch” the solutions at the boundary between the neighboring blocks, it is suggested that in the vicinity of the boundary the total kink energy on the left and on the right from the boundary are equal. This condition is valid, however, only in the case of the absence of the energy loss when crossing the boundaries between the blocks.

### Results and discussion

In Fig. 3 we present the energy profile of the gene coding interferon alpha 17. To construct the profile, we calculated the rest energy of the DNA kink according to the formula  $\tilde{E}_0 = 8\sqrt{\tilde{K}'\tilde{V}}$  for each of the three regions (Table 4). Taking into account that the kink energy is constant inside each of the regions we obtained the desired energy profile.

**Table 4.** Kink rest energy

Region	$\tilde{E}_0 \times 10^{-18}$ (J)
1	1.60
2 (CDS)	1.64
3	1.54



**Fig. 3.** Energy profile of the gene coding interferon alpha 17 and schematic picture of the sequence (at the bottom).

As can be seen from Fig. 3, the energy profile of the gene coding interferon alpha 17 includes one barrier corresponding to CDS region. So, the problem of the kink movement in the gene can be reformulated now as the problem of the movement of a quasi particle in the potential field with one energy barrier.

### Results obtained in the absence of dissipation

In this case, the total energy of the DNA kink remains constant during the movement along the whole gene sequence. Assuming that at the initial moment the velocity of kink is equal to  $v_{01}$ , the total energy can be written as:

$$\tilde{E}(t) = \frac{\tilde{E}_{01}}{\sqrt{1 - \left(\frac{v_{01}}{\tilde{C}_{01}}\right)^2}} = \text{const}, \quad (14)$$

where  $\tilde{E}_{01} = 8\sqrt{\tilde{K}_1} \tilde{V}_1$  is the rest energy of the kink moving in the first region and  $\tilde{C}_{01} = (\tilde{K}' a^2 / \tilde{I})^{1/2}$  is the sound velocity in the first region.

Since the energy profile of the gene includes the barrier, it is interesting to estimate the minimum value of the initial kink velocity ( $v_{01}^{crit}$ ) required to overcome the barrier. We obtained the estimation by equating the total energy of the kink moving with the velocity  $v_{01}^{crit}$  in the first region to the rest energy of the kink in the second region:

$$\frac{\tilde{E}_{01}}{\sqrt{1 - \left(\frac{v_{01}^{crit}}{\tilde{C}_{01}}\right)^2}} = \tilde{E}_{02}. \quad (15)$$

From (15) we found:

$$v_{01}^{crit} = \sqrt{\frac{\tilde{K}'_1 a^2}{\tilde{I}_1}} \sqrt{1 - \left(\frac{\tilde{E}_{01}}{\tilde{E}_{02}}\right)^2}. \quad (16)$$

Inserting into (16) the values of the kink rest energies presented in Table 4, as well as the values of the averaged model parameters from Table 3, we found:  $v_{01}^{crit} = 453.80$  m/s.

In view of this value for further analysis we choose three model values of the initial kink velocity ( $v_{01}$ ): 500 m/s, 800 m/s and 1500 m/s. They all exceed the obtained critical value  $v_{01}^{crit}$ . Thus in all three cases, the DNA kink necessarily reaches the CDS region.

To construct the kink trajectories, we applied the following simple algorithm. First of all we calculated the kink velocity in the second region ( $v_2$ ) by equating the total energies of the kink in the first and second regions:

$$\frac{\tilde{E}_{01}}{\sqrt{1 - \left(\frac{v_{01}}{\tilde{C}_{01}}\right)^2}} = \frac{\tilde{E}_{02}}{\sqrt{1 - \left(\frac{v_2}{\tilde{C}_{02}}\right)^2}}. \quad (17)$$

From (17) we found the required velocity:

$$v_2 = \tilde{C}_{02} \sqrt{1 - \frac{\tilde{E}_{02}^2}{\tilde{E}_{01}^2} \left(1 - \left(\frac{v_{01}}{\tilde{C}_{01}}\right)^2\right)}. \quad (18)$$

Acting in a similar way, we obtained the kink velocity in the third region:

$$v_3 = \tilde{C}_{03} \sqrt{1 - \frac{\tilde{E}_{03}^2}{\tilde{E}_{02}^2} \left(1 - \left(\frac{v_{02}}{\tilde{C}_{02}}\right)^2\right)}. \quad (19)$$

Inserting into (18) and (19) the values of the model parameters, we obtained the values of the kink velocities in all three regions and for each of the three model values of the initial kink velocity. The results of the calculation are presented in Table 5.

**Table 5.** Velocities of the DNA kinks

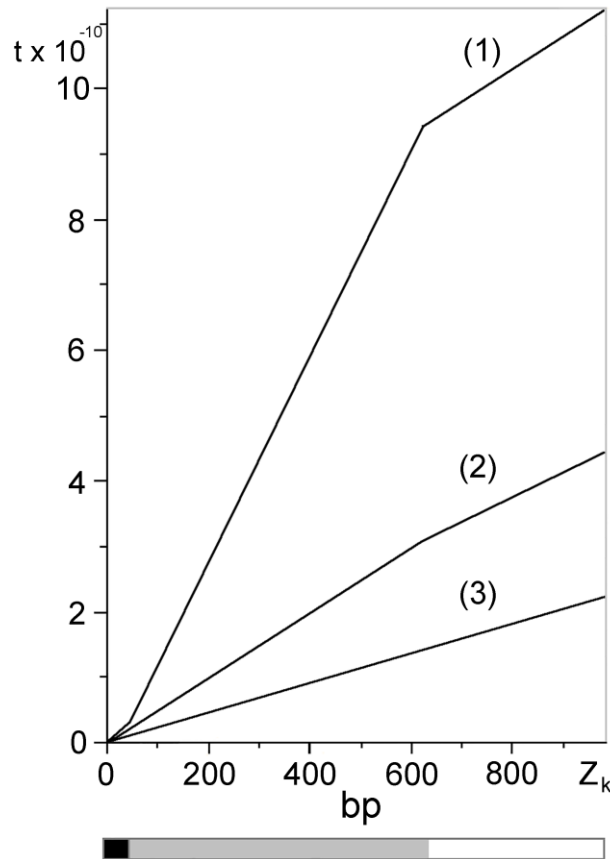
Region	$v_1$ (m/s)	$v_2$ (m/s)	$v_3$ (m/s)
	500	216.69	688.67
2 (CDS)	800	670.64	914.16
3	1500	1455.32	1525.68

Secondly, we used the obtained values to calculate the time of crossing the first and second boundaries ( $t_1$  and  $t_2$ ), as well as the time of reaching the right end of the gene ( $t_3$ ):

$$t_1 = \frac{z_1}{v_1}, \quad t_2 = \frac{(z_2 - z_1) + v_2 t_1}{v_2}, \quad t_3 = \frac{(z_3 - z_2) + v_3 t_2}{v_3}, \quad (20)$$

where,  $z_1$  and  $z_2$  are the coordinates of the first and the second boundaries, and  $z_3$  is the coordinate of the right end of the gene.

Thirdly, we connected the points  $(0, 0)$ ,  $(t_1, z_1)$ ,  $(t_2, z_2)$  and  $(t_3, z_3)$ . The obtained results are presented in Fig. 4.



**Fig. 4.** Trajectories of the DNA kinks in the gene coding interferon alpha 17 and schematic picture of the sequence (at the bottom). Calculations were made in the absence of dissipation and for three values of the initial kink velocity: (1)  $v_{01} = 500$  m/s, (2)  $v_{01} = 800$  m/s and (3)  $v_{01} = 1500$  m/s.

The results presented in Fig. 4 show that with the increasing of the initial kink velocity the trajectories become more independent on the sequence of nitrous bases.

### Results obtained in the presence of dissipation

In this case, the total kink energy and the kink velocity are not constant even inside the regions. The dependence of the DNA kink velocity on time in each of the regions is described by the equation of McLaughlin and Scott [24]. In the first region (1..49) the solution of the equation has the form [25]:

$$v_{1,k}(t) = \frac{\left[ (v_{01} \tilde{\gamma}_{01}) \exp\left(-\frac{\tilde{\beta}_1}{\tilde{I}_1} t\right) \right]}{\sqrt{1 + \left[ \left( \frac{v_{01}}{\tilde{C}_{01}} \tilde{\gamma}_{01} \right) \exp\left(-\frac{\tilde{\beta}_1}{\tilde{I}_1} t\right) \right]^2}}, \quad (21)$$

where  $v_{01}$  is the initial kink velocity,  $\tilde{I}_1$  is the moment of inertia averaged over the first region,  $\tilde{C}_{01}$  is the sound velocity in the first region,  $\tilde{\gamma}_{01} = \left(1 - (v_{01} / \tilde{C}_{01})^2\right)^{-1/2}$ ,  $\tilde{\beta}_1$  is the coefficient of dissipation.

The solutions in the second (50..619) and in the third (620..980) regions are determined by the similar formulas:



$$v_{2,k}(t) = \frac{\left[ (v_{02} \tilde{\gamma}_{02}) \exp\left(-\frac{\tilde{\beta}_2}{\tilde{I}_2}(t-t_1)\right) \right]}{\sqrt{1 + \left[ \left( \frac{v_{02}}{\tilde{C}_{02}} \tilde{\gamma}_{02} \right) \exp\left(-\frac{\tilde{\beta}_2}{\tilde{I}_2}(t-t_1)\right) \right]^2}}, \quad (22)$$

$$v_{3,k}(t) = \frac{\left[ (v_{03} \tilde{\gamma}_{03}) \exp\left(-\frac{\tilde{\beta}_3}{\tilde{I}_3}(t-t_2)\right) \right]}{\sqrt{1 + \left[ \left( \frac{v_{03}}{\tilde{C}_{03}} \tilde{\gamma}_{03} \right) \exp\left(-\frac{\tilde{\beta}_3}{\tilde{I}_3}(t-t_2)\right) \right]^2}}, \quad (23)$$

where  $v_{02}$  and  $v_{03}$  are the initial velocities of the DNA kink in the second and the third regions,  $\tilde{I}_2$  and  $\tilde{I}_3$  are the moments of inertia of nitrous bases averaged over the second and third regions, respectively,  $\tilde{C}_{02}$  and  $\tilde{C}_{03}$  are the sound velocities in the second and in the third regions,  $\tilde{\gamma}_{02} = (1 - (v_{02} / \tilde{C}_{02})^2)^{-1/2}$ ,  $\tilde{\gamma}_{03} = (1 - (v_{03} / \tilde{C}_{03})^2)^{-1/2}$ ,  $\tilde{\beta}_2$  and  $\tilde{\beta}_3$  are the coefficients of dissipation,  $t_1$  and  $t_2$  are moments of time of crossing the first and the second boundaries.

Formulas (22) and (23) contain unknown values of the initial kink velocities in the second and thirds regions ( $v_{02}, v_{03}$ ), as well as the moments of time of crossing the first and the second boundaries ( $t_1, t_2$ ). To calculate them, we use approximation where the energy loss at the moment of crossing the boundaries is neglected. The results of the calculations are presented in Table 6.

**Table 6.** The initial kink velocities and the moments of time of crossing the boundaries

$v_{01}$ m/s	$t_1$ $\times 10^{-11}$ s	$v_{02}$ m/s	$t_2$ $\times 10^{-10}$ s	$v_{03}$ m/s	$t_3$ $\times 10^{-10}$ s
500	3.68	–	–	–	–
800	2.19	570.31	–	–	–
1500	1.12	1413.13	2.00	913.74	4.20

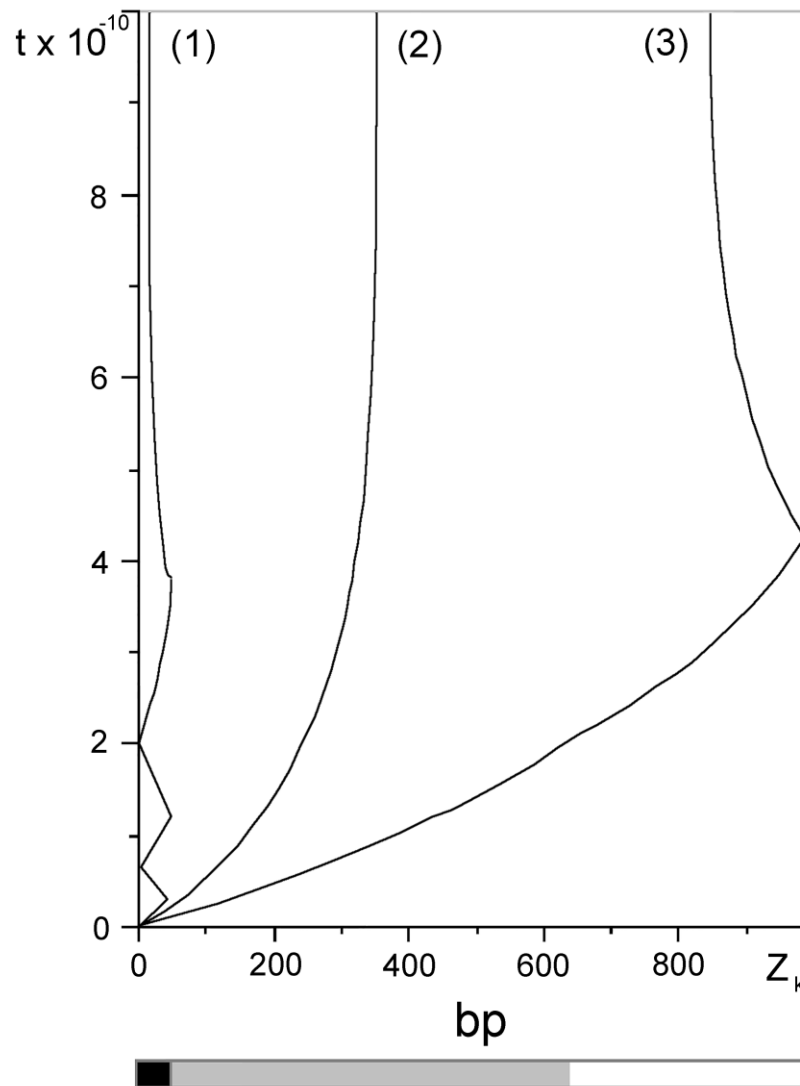
By defining the coordinate of the DNA kink in the  $i$ -th region by the relation:  $v_{ki}(t) = \frac{dz_{ki}(t)}{dt}$ , we find the formulas that determine the dependence of the DNA kink coordinate on time for each of the regions:

$$z_{1,k}(t) = \int_0^t v_{1,k}(\tau) d\tau. \quad (24)$$

$$z_{2,k}(t) = \int_{t_1}^t v_{2,k}(\tau) d\tau \quad (25)$$

$$z_{3,k}(t) = \int_{t_2}^t v_{3,k}(\tau) d\tau \quad (26)$$

With the help of formulas (24)–(26) and of the data of Tables 3 and 6, we construct the trajectory of the kink in the plane ( $t, z$ ) for different values of the initial kink velocity  $v_{01}$  (Fig. 5).



**Fig. 5.** Trajectories of the DNA kinks in the gene coding interferon alpha 17 and schematic picture of the sequence (at the bottom). Calculations were made in the presence of dissipation and for three values of the initial kink velocity: (1)  $v_{01} = 500$  m/s, (2)  $v_{01} = 800$  m/s and (3)  $v_{01} = 1500$  m/s.

Fig. 5 shows that in the first case (curve (1)) the initial total energy of the DNA kink is not large enough to overcome the energy barrier. Therefore when reaching the first boundary, the kink is reflected from the boundary and begins to move to the left end of the gene. After reaching the left end two scenarios are possible.

a) The DNA kink leaves the gene and goes to the left neighboring region. This scenario is possible when the value of the rest kink energy in the left neighboring region is larger than or equal to the total kink energy at the moment of crossing the left end of the gene.

b) The DNA kink is reflected from the left end of the gene. This scenario is possible when the value of the rest kink energy in the left neighboring region is less than the total kink energy at the moment of reaching the left end of the gene. After a few zig-zag motions the DNA kink stops. Just this scenario is shown in Fig. 5 (curve (1)).

In the second case (curve (2)), the total energy of the DNA kinks is quite large to cross the first boundary and to get into the CDS region. But later the DNA kink stops inside the CDS region because of the energy losses due to dissipation.

In the third case (curve (3)), the DNA kink passes the whole CDS region, gets into the third region and reaches the right end of the gene. Here again two scenarios are possible. In the first scenario, the DNA kink goes outside the gene. In the second scenario, the kink is

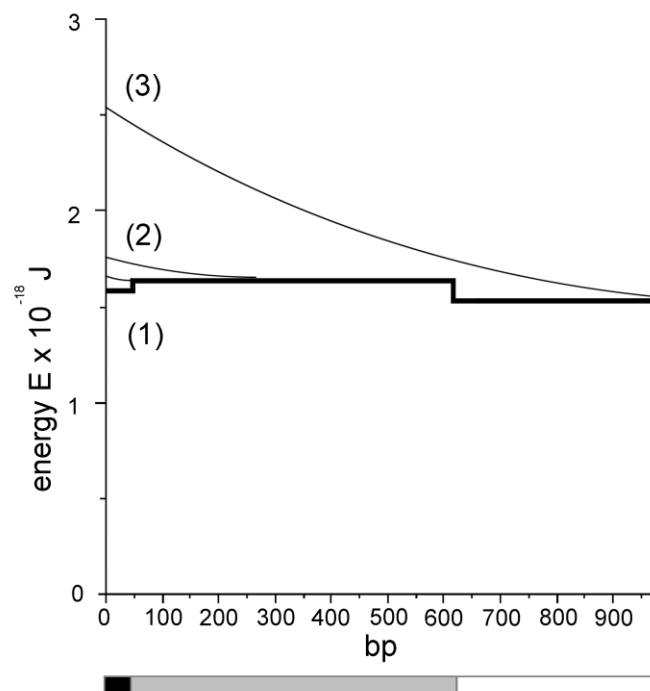
reflected from the right end of the gene and stops inside the third region, as it shown in Fig. 5, curve (3).

The total energy of the DNA kink in each of the regions is the function of time:

$$\tilde{E}_i(t) = \frac{\tilde{E}_{0i}}{\sqrt{1 - \left(\frac{v_{i,k}(t)}{\tilde{C}_{0i}}\right)^2}}, i = 1, 2, 3, \quad (27)$$

where the velocities  $v_{i,k}(t)$  are the functions of time and determined by formulas (21)–(23).

Fig. 6 shows the results of the calculations of the total kink energy made for three model values of the initial kink velocity: 500 m/s, 800 m/s and 1500 m/s. It can be seen that in the all three cases the total energy of the kink decreases due to effects of dissipation. The results show, that the greater the initial total energy of the kink, the faster the energy decrease. The energy decrease is limited, however, from the bottom by the rest energy of the kink shown in Fig. 6 by the thick solid line.



**Fig. 6.** Energy losses of the DNA kink in the gene coding interferon alpha 17 and schematic picture of the sequence (at the bottom). Calculations were made for three values of the initial kink velocity: (1)  $v_{01} = 500$  m/s, (2)  $v_{01} = 800$  m/s and (3)  $v_{01} = 1500$  m/s. The thick solid line shows the energy profile of the gene.

## CONCLUSIONS

In this paper, we presented a simple and convenient method of the analysis of the movement of transcription bubbles which were modeled by kink solutions of the modified sine-Gordon equation. With the help of the method we calculated the trajectories of the bubbles moving through the CDS region in the gene coding interferon alpha 17.

It should be noted, however, that the developed approach is not without of some deficiencies and limitations. One of them is a rather large value of the initial kink velocity required to overcome the CDS region. One more limitation consists in the assumption of nonradioactive crossing the boundaries. Finally, it should be noted that all results have been obtained in the frameworks of the Y-model. But this model takes into account the angular oscillations of the nitrous bases only in one of the two polynucleotide chains of the DNA. The

second chain is modeled as an average field. The model also does not take into account the interaction of the angular displacements with transverse and longitudinal displacements of the nitrous bases. However, we hope that these limitations are not insurmountable, and the method can be generalized and applied to more accurate and complex models of the DNA dynamics.

This work has been partly supported by the Program for increasing the international competitiveness of Tomsk State University for 2013–2020.

## REFERENCES

1. Dubois A., Francois C., Descamps V., Fournier C., Wychowski C., Dubuisson J., Castelain S., Duverlie G. Enhanced anti-HCV activity of interferon alpha 17 subtype. *Virology*. 2009. V. 6. P. 70. doi: [10.1186/1743-422X-6-70](https://doi.org/10.1186/1743-422X-6-70)
2. Bychkov V.A., Ryazantseva N.V., Novitsky V.V. Analysis of the combined effect of polymorphisms interferon genes OAS1, OAS3, PKR, IFNA17 and IFNG in susceptibility to chronic viral hepatitis C. *Bulletin of Siberian medicine*. 2011. V. 3. P. 19–23.
3. *GenBank: Homo sapiens interferon alpha 17 (IFNA17), mRNA*. URL: [http://www.ncbi.nlm.nih.gov/nuccore/NM\\_021268.2](http://www.ncbi.nlm.nih.gov/nuccore/NM_021268.2) (accessed December 23, 2016)
4. Englander S.W., Kallenbach N.R., Heeger A.J., Krumhansl J.A., Litwin A. Nature of the open state in DNA structure. *Proc. Natl. Acad. Sci. USA* 1980. V. 77. P. 7222–7226.
5. Shigaev A.S., Ponomarev O.A., Lakhno V.D. Theoretical and Experimental Investigations of DNA Open States. *Mathematical biology and bioinformatics*. 2013. V. 8. P. 553–664. doi: [10.17537/2013.8.553](https://doi.org/10.17537/2013.8.553)
6. Scott A. *Encyclopedia of Nonlinear Science*. New York: Frances and Taylor, 2005.
7. Ohya T. *DNA Conformation and Transcription*. New York: Springer–Science Business Media, 2005.
8. Ivancevic V.G., Ivancevic T.T. Sine-Gordon Solitons, Kinks and Breathers as Physical Models of Nonlinear Excitations in Living Cellular Structures. *arXiv*. 2013. 1305.0613v1 [q-bio.OT].
9. Sanchez D.S., Qu H., Bulla D., Zocchi G. DNA kinks and bubbles: temperature dependence of the elastic energy of sharply bent 10-nm-size DNA molecules. *Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys.* 2013. V. 87. № 2. P. 022710. doi: [10.1103/PhysRevE.87.022710](https://doi.org/10.1103/PhysRevE.87.022710)
10. Peyrard M., Bishop A.R. Statistical Mechanics of a Nonlinear Model for DNA Denaturation. *Phys. Rev. Lett.* 1989. V. 62. P. 2755–2758. doi: [10.1103/PhysRevLett.62.2755](https://doi.org/10.1103/PhysRevLett.62.2755)
11. Yomosa S. Solitary excitations in deoxyribonucleic acid (DNA) double helices. *Phys. Rev. A*. 1984. V. 30. P. 474–480. doi: [10.1103/PhysRevA.30.474](https://doi.org/10.1103/PhysRevA.30.474)
12. Takeno S., Homma S.A. Coupled Base-Rotator Model for Structure and Dynamics of DNA. *Prog. Theor. Phys.* 1984. V. 72. P. 679–693. doi: [10.1143/PTP.72.679](https://doi.org/10.1143/PTP.72.679)
13. Yakushevich L.V. Nonlinear DNA dynamics: a new model. *Phys. Letters A*. 1989. V. 136. P. 413–417. doi: [10.1016/0375-9601\(89\)90425-8](https://doi.org/10.1016/0375-9601(89)90425-8)
14. Barbi M., Cocco S., Peyrard M. Helicoidal model for DNA opening. *Phys. Letters A*. 1999. V. 253. P. 358–369. doi: [10.1016/S0375-9601\(99\)00059-6](https://doi.org/10.1016/S0375-9601(99)00059-6)
15. Yakushevich L.V. *Nonlinear Physics of DNA*. New York: Wiley–Verlag, 2004. 190 p.
16. Vasumathi V., Daniel M. Base-pair opening and bubble transport in a DNA double helix induced by a protein molecule in a viscous medium. *Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys.* 2009. V. 80. P. 061904. doi: [10.1103/PhysRevE.80.061904](https://doi.org/10.1103/PhysRevE.80.061904)

17. Daniel M., Vanitha M. Bubble solitons in an inhomogeneous, helical DNA molecular chain with flexible strands. *Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys.* 2011. V. 84. P. 031928. doi: [10.1103/PhysRevE.84.031928](https://doi.org/10.1103/PhysRevE.84.031928)
18. Di Garbo A. Longitudinal displacements of base pairs in DNA and effects on the dynamics of nonlinear excitations. *Math. Biosci.* 2013. V. 245. P. 70–75. doi: [10.1016/j.mbs.2013.03.010](https://doi.org/10.1016/j.mbs.2013.03.010)
19. Di Garbo A. Anharmonic longitudinal motion of bases and dynamics of nonlinear excitation in DNA. *Biophys. Chem.* 2016. V. 208. P. 76–83. doi: [10.1016/j.bpc.2015.09.006](https://doi.org/10.1016/j.bpc.2015.09.006)
20. Shapovalov A.V., Krasnobaeva L.A. *The solitons of the sine-Gordon*. Tomsk: Tomsk State University, 2009.
21. Krasnobaeva L.A., Yakushevich L.V. Rotational dynamics of bases in the gene coding interferon alpha 17 (IFNA17). *Journal of Bioinformatics and Computational Biology*. 2015. V. 13. P. 1540002. doi: [10.1142/S0219720015400028](https://doi.org/10.1142/S0219720015400028)
22. Yakushevich L.V., Krasnobaeva L.A., Shapovalov A.V., Quintero N.R. One- and Two-Soliton Solutions of the Sine-Gordon Equation as Applied to DNA. *Biophysics*. 2005. V. 50. P. 450–455.
23. Yakushevich L.V., Krasnobaeva L.A. Forced Oscillations of DNA bases. *Biophysics*. 2016. V. 61. P. 286–296. doi: [10.1134/S000635091602024X](https://doi.org/10.1134/S000635091602024X)
24. McLaughlin D.W., Scott A.C. Perturbation analysis of fluxion dynamics. *Phys. Rev. A*. 1978. V. 18. P. 1652. doi: [10.1103/PhysRevA.18.1652](https://doi.org/10.1103/PhysRevA.18.1652)
25. Yakushevich L.V., Krasnobaeva L.A. Effects of dissipation and external fields on the dynamics of conformational distortions in DNA. *Biophysics*. 2007. V. 52. P. 237–243. doi: [10.1134/S0006350907020066](https://doi.org/10.1134/S0006350907020066)
26. Yakushevich L.V., Krasnobaeva L.A. Peculiar Features of Kink Dynamics in Inhomogeneous DNA. *Biophysics*. 2008. V. 53. P. 36–41. doi: [10.1134/S0006350908010041](https://doi.org/10.1134/S0006350908010041)
27. Yakushevich L.V., Krasnobaeva L.A. A new approach to studies of non-linear dynamics of kinks activated in inhomogeneous polynucleotide chains. *Int. J. Nonlin. Mech.* 2008. V. 43. P. 1074–1081. doi: [10.1016/j.ijnonlinmec.2008.05.003](https://doi.org/10.1016/j.ijnonlinmec.2008.05.003)

Received November 24, 2016.

Published January 09, 2017.