# Ensemble of DNA Kinks

*Ludmila V.* Yakushevich[1,*] and *Larisa A.* Krasnobaeva[2,3]

[1]Institute of Cell Biophysics of the Russian Academy of Sciences, RU-142290, Pushchino, Moscow region, Russia
[2]Siberian State Medical University, RU-634050, Tomsk, Russia
[3]Tomsk State University, RU-634050, Tomsk, Russia

**Abstract.** The DNA open states, which are locally unwound regions of the double helix within which hydrogen bonds between complementary nitrous bases are broken, are often modeled as quasiparticles – DNA kinks. Most of the works on the DNA kinks are devoted to the studies of their dynamic properties, as well as their role in the functioning of the molecule. However, if not one but *N* open states are formed in the DNA molecule it is reasonable to consider the problem of the statistics of the ensemble of *N* DNA kinks. The statistical properties of such an ensemble are still poorly understood. In the present work, we study these properties applying new data on the dynamic characteristics of DNA kinks.

## 1 Introduction

When studying the internal conformational mobility of the DNA molecule by various experimental methods and, in particular, by the method of hydrogen-tritium exchange [1], the researchers found that locally unwound regions called open states or bubbles [2], can be formed in the DNA double helix. The bubbles arise as a result of exposure to temperature, collisions with solution molecules, radiation and interactions with proteins [3-6]. It is assumed that open states play an important role in the functioning of the molecule. The most striking example is the transcriptional bubble, which is formed at the initial stage of the transcription process as a result of the interaction of RNA polymerase with the promoter DNA region [7]. It is assumed that the velocity of the bubble movement along the DNA sequence is closely related with the velocity of the transcription process.

To describe the open states mathematically, the instrument of theoretical physics and nonlinear mathematics are often used including representations about kinks, antikinks, and breathers. In this paper, following the work of Englander and coauthors [1] as well as the works of their followers [8-13], we shall use the sine-Gordon equation to model the internal dynamics of DNA, and shall apply the one-soliton solutions of the sine-Gordon equation – kinks, to describe the DNA open states.

If the DNA molecule is long enough, it can form not one, but several open states, and it is legitimate to raise the question of the statistical properties of such an ensemble. This question was raised in the monograph [3]. A possibility to apply statistical characteristics of kinks in the calculations the dynamic form factor of neutron scattering on DNA was considered also in [14]. However, at that time the lack of a complete set of data on the DNA dynamic parameters did not make it possible to apply these preliminary developments to real DNA sequences.

In the present work, the question of the statistics of the DNA kinks ensemble is investigated taking into account new data on the DNA dynamic parameters. For definiteness, all calculations are performed for kinks activated in the pBR322 plasmid which is a small circular DNA widely applied in gene research, and its components are used to create new instrumental plasmids [15].
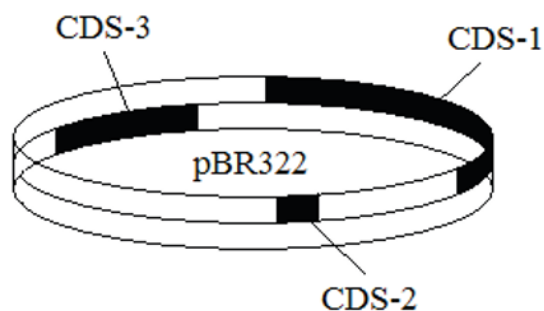


**Fig. 1.** Schematic picture of plasmid pBR322.. CDS-1, CDS-2 and CDS-3 are coding regions.

## 2 DNA internal dynamics

### 2.1 Model

Let us use the sine-Gordon equation to simulate the DNA internal dynamics:

$$I \frac{\partial^2 \phi}{\partial t^2} = K' a^2 \frac{\partial^2 \phi}{\partial z^2} - V \sin \phi \cdot \qquad (1)$$

---

Here $\phi(z,t)$ is the angular displacement of the nitrous base, $I$ is the moment of inertia of the base, $K'$ is the constant characterizing the torsion rigidity of the sugar-phosphate chain, $V$ is the constant that characterizes the interaction between complementary bases within pairs, $a$ is distance between the neighboring pairs of bases.

The sequence of the plasmid pBR322 is heterogeneous. Therefore, the values of the coefficients $I$, $K'$ and $V$ are not constants and change along the DNA sequence. To simplify calculations, let us use the quasi-homogeneous approximation, in which the coefficients of equation (1) become constants because of the averaging over the entire length of the plasmid sequence [16,17]:

$$\bar{I} = I_A c_A + I_T c_T + I_G c_G + I_C c_C,$$
$$\overline{K'} = K'_A c_A + K'_T c_T + K'_G c_G + K'_C c_C, \quad (2)$$
$$\overline{V} = V_A c_A + V_T c_T + V_G c_G + V_C c_C,$$

where $I_j$, $K'_j$ and $V_j$ are the values of the coefficients of equation (1) for different types of bases ($j$ = A, T, G, C), $c_j = n_j / n$ is the concentration of nitrous bases of the $j$-th type; $n_j$ is the number of nitrous bases of the $j$-th type; $n$ is the total number of bases in the plasmid sequence. The values of the coefficients $I_j$, $K'_j$ and $V_j$ are presented in the Table 1.

**Table 1**. Coefficients of equations (2) [18].

| Type of the base | $I_j \times 10^{-44}$ (kg·m²) | $K'_j \times 10^{-18}$ (J) | $V_j \times 10^{-20}$ (N/m) |
|---|---|---|---|
| A | 7,61 | 2,35 | 2,09 |
| T | 4,86 | 1,61 | 1,43 |
| G | 8,22 | 2,27 | 3,12 |
| C | 4,11 | 1,54 | 2,12 |

Considering relations (2) the model equation (1) takes the form:

$$\bar{I}\frac{\partial^2 \phi}{\partial t^2} = \overline{K'}a^2 \frac{\partial^2 \phi}{\partial z^2} - \overline{V}\sin\phi. \quad (3)$$

We calculated the values of the coefficients $\bar{I}$, $\overline{K'}$ and $\overline{V}$ for the sequence of the plasmid pBR322. The results of the calculations are presented in Table 2.

**Table 2**. Coefficients of equations (3).

| Type of the sequences | $\bar{I} \times 10^{-44}$ (kg·m²) | $\overline{K'} \times 10^{-18}$ (J) | $\overline{V} \times 10^{-20}$ (N/m) | $a \times 10^{-10}$ (m) |
|---|---|---|---|---|
| pBR322 | 6,14 | 1,93 | 2,21 | 3,4 |

## 2.2 Dynamic characteristics of DNA kinks

The kink-like solution of equation (3) is determined by the formula:

$$\phi_k(z,t) = 4\arctg\left\{\exp\left[\frac{\gamma}{d_k}\left(z - \upsilon_k t - z_0\right)\right]\right\}, \quad (4)$$

where $\upsilon_k$ is the velocity of the kink, $d_k = (\overline{K'}a^2/\overline{V})^{1/2}$ is the kink size, $\gamma = (1 - \upsilon_k^2/C^2)^{-1/2}$, $C = (\overline{K'}a^2/\bar{I})^{1/2}$ is the sound velocity in DNA.

Hamiltonian corresponding to equation (3) has the form:

$$H = \int_{-\infty}^{\infty}\left[\frac{\bar{I}}{2}\left(\frac{\partial \phi_t}{\partial t}\right)^2 + \frac{\overline{K'}a^2}{2}\left(\frac{\partial \phi_t}{\partial z}\right)^2 + \overline{V}(1-\cos\phi)\right]\frac{dz}{a}. \quad (5)$$

Inserting the solution (4) into the formula (5), we obtain the total energy of the DNA kink:

$$E_k = 8\gamma\sqrt{\overline{K'}\overline{V}}. \quad (6)$$

In the "non-relativistic" limit, when the kink velocity, $v_k$, is small compared to the sound velocity, $C$, the formula (6) takes the form:

$$E_k \cong 8\sqrt{\overline{K'}\overline{V}}\left(1 - \frac{1}{2}\frac{\upsilon_k^2}{C^2}\right) = E_{0k} + \frac{8\sqrt{\overline{K'}\overline{V}}}{C^2}\frac{\upsilon_k^2}{2}, \quad (7)$$

where $E_{0k} = 8\sqrt{\overline{K'}\overline{V}}$ is the rest energy of the kink.

From (7) we obtain the rest mass ($m_k$), kinetic energy ($T_k$) and potential energy ($U_k$) of the kink:

$$m_k = \frac{8\sqrt{\overline{K'}\overline{V}}}{C^2}, \quad (8)$$

$$T_k = \frac{8\sqrt{\overline{K'}\overline{V}}}{C^2}\frac{\upsilon_k^2}{2}, \quad (9)$$

$$U_k = 8(\sqrt{\overline{K'}\overline{V}}). \quad (10)$$

The momentum of the kink can be written then in the form:

$$p_k = m_k \upsilon_k. \quad (11)$$

These dynamic characteristics make it possible to propose a simple model of the DNA kinks ensemble in the form of a conventional classical system consisting of $N$ interacting particles having size $d_k$, mass $m_k$, momentum $p_k$ and rest energy $E_{0k}$.

We calculated the values of the main dynamic parameters of the kinks activated in pBR322 plasmid and presented them in the Table 3.

**Table 3**. Dynamic parameters of the kinks activated in pBR322 plasmid.

| Type of the sequences | $m_k \times 10^{-24}$ (kg) | $E_{0k} \times 10^{-17}$ (J) | $d_k$ (bp) |
|---|---|---|---|
| pBR322 | 0,224 | 0,165 | 9,34 |

# 3 Statistical characteristics of an ensemble of DNA kinks

## 3.1 Model

If the number of DNA kinks is fixed, interactions between them are negligible and all kinks are identical, then the total energy of an ensemble of the DNA kinks can be written as:

$$E(P,Q) = NE_{0k} + \sum_i^N \frac{p_{ki}^2}{2m_{ki}} = 8N\sqrt{K'V} + \sum_i^N \frac{8\sqrt{K'V}}{C^2}\frac{\upsilon_{ki}^2}{2}, \qquad (12)$$

and the statistical sum (or partition function) $Z$ [19] and the Gibbs distribution $\rho(P,Q)$ [20] take the form:

$$Z = \frac{L^N}{N!}\left(e^{-\frac{E_{0k}}{k_BT}}\sqrt{\frac{m_k k_B T}{2\pi\hbar^2}}\right)^N = \frac{L^N}{N!}\left(\frac{2}{\hbar C}e^{-\frac{8\sqrt{K'V}}{k_BT}}\sqrt{\frac{\sqrt{K'V}k_BT}{\pi}}\right)^N, (13)$$

$$\rho(P,Q) = \frac{1}{\left(L\sqrt{2\pi m_k k_B T}\right)^N}\exp\left[-\frac{\sum_1^N p_{ki}^2}{2m_k k_B T}\right], \quad (14)$$

where $L$ is the length of the DNA sequence, $N$ is the number of kinks, $T$ is the absolute temperature, $k_B$ is the Boltzmann constant, $Q=\{q_1, q_2, ...., q_N\}$ and $P = \{p_{k1}, p_{k2}, ..., p_{kN}\}$ are the sets of coordinates and momentums of the kinks.

### 3.2 Velocity distribution function of DNA kinks

It follows from the form of formula (14) that, in the approximation under consideration, the kinks are statistically independent, and the distribution function of the individual $i$-th kink can be written as:

$$\rho_{ki}(p_{ki}) = \frac{e^{-\left(\frac{p_{ki}^2}{2m_k k_B T}\right)}}{L\sqrt{2\pi m_k k_B T}}. \qquad (15)$$

Having made a change of variables $p_{ki} = m_k \upsilon_{ki}$ in (15), we get:

$$\rho_{ki}(p_{ki}) = \frac{1}{L}\rho_{ki}(\upsilon_{ki}), \qquad (16)$$

where

$$\rho_{ki}(\upsilon_{ki}) = \sqrt{\frac{m_k}{2\pi k_B T}}e^{-\left(\frac{m_k\upsilon_{ki}^2}{2k_B T}\right)}, \qquad (17)$$

is the kink velocity distribution function that is similar to the Maxwell velocity distribution. We calculated the velocity distribution function of kinks in plasmid pBR322. The calculation result is shown in Fig. 2.
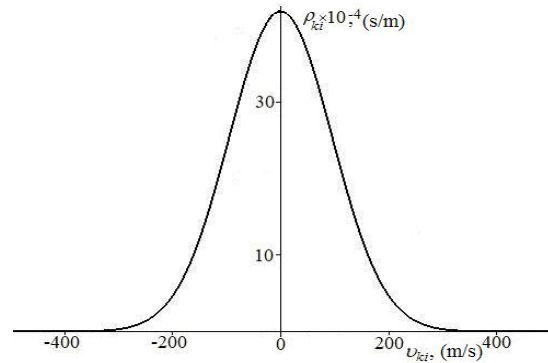


**Fig. 2.** Velocity distribution function of kinks.

### 3.3 Free energy, average energy and entropy

Let us use th In the approximation considered, the free energy of the DNA kinks ensemble is defined as [21]:

$$F = -k_B T \ln Z = N\left[8\sqrt{K'V} - k_B T \ln\left(\frac{2L}{\hbar C}\frac{e}{N}\sqrt{\frac{k_B T}{\pi}\sqrt{K'V}}\right)\right], (18)$$

and the average energy is given as [22]:

$$\varepsilon = (k_B T)^2 \frac{\partial \ln Z}{\partial(k_B T)} = N\left(8\sqrt{K'V} + \frac{k_B T}{2}\right). \qquad (19)$$

Then the entropy and the heat capacity are determined by formulas [23]:

$$S = \frac{\varepsilon - F}{k_B T} = N\left[\frac{1}{2} + \ln\left(\frac{2L}{\hbar C}\frac{e}{N}\sqrt{\frac{k_B T}{\pi}\sqrt{K'V}}\right)\right], \quad (20)$$

$$C_{v,k} = \frac{\partial\varepsilon_k}{\partial T} = \frac{Nk_B}{2}. \qquad (21)$$

We plotted the temperature dependence of the free and average energies, as well as the entropy of the ensemble of kinks in plasmid pBR322.

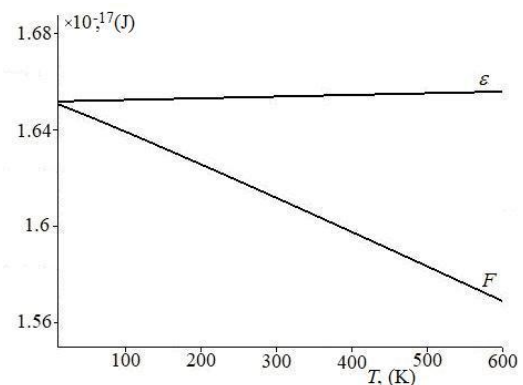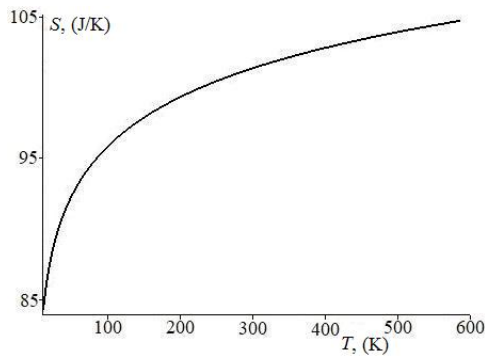The results are presented in Fig. 3 and Fig. 4.



**Fig.3.** Free energy $F$ and average energy $\varepsilon$ of ensemble of kinks of plasmid pBR322.

**Fig.4.** Entropy $S$ of ensemble of kinks of plasmid pBR322.

It can be seen from formulas (18) - (21) and from the graphs presented in Figures 3, 4 that with increasing temperature, the average energy and the entropy increase, the free the energy decreases, and the heat capacity remains constant for any temperature $T$. As expected this behavior is in a good agreement with the statistical properties of an ideal gas of quasiparticles having mass $m_k$, velocity $v_k$ and the rest energy $E_{0k}$.

## 4 Discussion and conclusion

In this work, we studied statistical properties of ensemble of the DNA kinks which simulate the open states of the molecule. When calculating the statistical characteristics, we used the data on the dynamic parameters of the kinks of the pBR322 plasmid which is widely applied in genetic engineering.

Formulas for statistical sum, free energy, velocity distribution function, average energy, entropy and heat capacity of the DNA kinks ensemble were obtained under the following three conditions: the number of the DNA kinks is fixed, interactions between them are negligible and all kinks are identical.

In this case, ensemble of the DNA kinks is similar to the ideal gas of identical quasiparticles having mass $m_k$, velocity $v_k$ and the rest energy $E_{0k}$. Removing any of the three conditions will allow one to obtain in the future amendments to the obtained formulas and graphs. We believe, however, that these amendments will be small and will not make fundamental changes to the results obtained above.

It should also be noted that all these results were obtained for the DNA model, which takes into account only one type of the internal DNA motions - angular oscillations of nitrous bases. Another limitation is related to the use of quasihomogeneous approximation. Nevertheless, it can be assumed that the above approach and the results obtained in the studying of the statistical properties of the DNA kinks are quite general and can be used to study the statistical properties of the DNA open states in the framework of more complex and more realistic DNA models.

## References

1. W. Englander, N.R. Kallenbach, A.J. Heeger, J.A. Krumhansl, A. Litwin, PNAS USA **77**, 7222 (1980)

2. A.A. Grinevich, A.A. Ryasik, L.V. Yakushevich, Chaos, Solitons & Fractals **75**, 62 (2015)

3. L.V. Yakushevich, *Nonlinear Physics of DNA* (Wiley, Weinheim, 2004)

4. M. Peyrard, Nonlinearity **17**, R1 (2004)

5. M. Peyrard (ed.), *Nonlinear Excitations in Biomolecules* (Springer, Berlin, 1995)

6. A. Scott (ed.), *Encyclopedia of Nonlinear Science* (Frances and Taylor, New York, 2005)

7. E.S. Severin, T.L. Aleinikova, E.V. Osipov, S.A. Silaeva, *Biological chemistry* (Medical Information Agency, Moscow, 2008) [in Russian]

8. G. Gaeta, Phys. Rev. E **74**, 021921 (2006)

9. M. Daniel, V. Vasumathi, Physica D **231**, 10 (2007)

10. G. Gaeta, L. Venier, Phys. Rev. E **78**, 011901 (2008)

11. V. Vasumathi, M. Daniel, Phys. Lett. A **373**, 76 (2008)

12. M. Cadoni, R. De Leo, S. Demelio, G. Gaeta, J. Nonlin. Math. Phys **17**, 557 (2010)

13. G. Derks, G. Gaeta, Physica D **240**, 1805 (2011)

14. L.V. Yakushevich, J. of Biolog. Phys. **24**, 131 (1999)

15. F. Bolivar, R.L. Rodriguez, P.J. Greene, M.C. Betlach, H.L. Heyneker, H.W. Boyer, J.H. Crosa and S. Falkow, Gene **2**, 95 (1977)

16. L.V. Yakushevich, L.A. Krasnobaeva, A.V. Shapovalov, N.R. Quintero, Biophysics **50**, 450 (2005)

17. L.V. Yakushevich, L.A. Krasnobaeva, Mathematical Biology and Bioinformatics **12**, 1 (2017)

18. L.V. Yakushevich, L.A. Krasnobaeva, Biophysics **61**, 286 (2016)

19. G.F. Karavaev, *Basic Principles of Statistical Physics* (TSU, Tomsk, 1993) [in Russian]

20. I.A. Kvasnikov, *Thermodynamics and Statistical Physics* (Editorial URSS, Moscow, 2010) [in Russian]

21. I.P. Bazarov, *Thermodynamics* (Higher School, Moscow, 1991) [in Russian]

22. A.V. Levanov, E.E. Antipenko, *Determination of thermodynamic properties by statistical methods. Classic perfect gas* (Moscow State University, 2006) [in Russian]

23. A. Sommerfeld, *Thermodynamics and Statistical Physics* (Moscow, 1955) [in Russian]