Thermodynamics of local DNA openings

Tomasz Lipniacki*
Institute of Fundamental Technological Research, Świętokrzyska Street 21, 00-049 Warsaw, Poland
(Received 11 July 2001; published 29 October 2001)

A mechanism connecting the local untwisting and opening of DNA double helix is proposed. The presented thermodynamical approach is based on two models: the Peyrard-Bishop model that describes the denaturation of DNA due to thermal fluctuations and the model developed by the author describing solitary torsional waves, which propagate along the DNA molecule forced by advancing RNA polymerase. The torsional wave implies that the DNA untwists locally causing a local decrease in the stacking interaction between adjacent base pairs. Molecular dynamics simulations have shown that thermal fluctuations (which are too small at physiological temperatures to denaturate the twisted DNA) may lead to the formation of a denaturation bubble placed in the untwisted region.

DOI: 10.1103/PhysRevE.64.051919 PACS number(s): 87.14.Gg

I. INTRODUCTION

The basic form of DNA is a double helix, consisting of two sugar-phosphate backbones and a base pairs chain inside. Since the four bases composing DNA are hydrophobic substances, they tend to stay together, rather than let the surrounding water fill gaps between them. If the adjacent base pairs in the sequence are in contact, the distance between their centers is $A \approx 3.3$ Å, smaller than the distance between the adjacent sugars in backbones $B \approx 6$ Å. This is because the sugar-phosphate backbones must wrap around the base pair chains in order to preserve their length. The base sequence codes the genetic information, but from the mechanical point of view DNA is an aperiodic chain. The four bases composing DNA have different masses. This poses a problem for all models. Fortunately, the masses of two base pairs composing DNA, namely, the adenine-thymine (A-T) and guanine-cytosine (C-T) are almost equal. Only in an approximation in which one neglects internal degrees of freedom of the base pairs, and treats each base pair as a rigid body, can DNA be regarded as a periodic structure.

The interest in the nonlinear dynamics of DNA started when Englander *et al.* [1] suggested that the existence of solitons propagating along the DNA molecule may be important in a process called ''RNA transcription.'' In the last two decades several models were proposed in order to substantiate this idea in quantitative terms (see Gaeta *et al.* [2] and Yakushevich [3] for the review).

II. THE MODEL

The approach to DNA dynamics presented in the current paper is based on the nonlinear model of DNA dynamics proposed in Ref. [4] and further developed in Ref. [5] and on the Peyrard-Bishop model describing thermal denaturation of DNA molecule.

Within the framework of the model [5] we proved the existence of solutions corresponding to torsional traveling waves propagating along the DNA molecule. The pulselike solutions describe the propagation of an untwisted (or oppositely twisted) region along the molecule. This solution pertains to the mechanical aspects of transcription of the messenger RNA—one of the most important processes of DNA evolution. To initiate the transcription process the DNA must untwist and open locally to let one strand serve as a template for synthesis of new RNA strand. Then the untwisted open region, 15–20 base pairs long, moves together with RNA polymerase along the DNA, as long as transcription proceeds. It is known that the RNA polymerase ''works'' as a processive motor capable of generating forces of 25–30 pN. This force is due to the free energy of hydrolysis of nucleotides as they are incorporated into the growing RNA chain (see Ref. [6] for experimental data and Ref. [7] for theory). It was shown [5] that even if the energy dissipation (damping) is present (it seems obvious that the motion of DNA leads to some energy dissipation due to nonelasticity of the DNA molecule or to interaction with the solvent) such solutions may still exist. The dissipated energy can be balanced by energy pumped by advancing polymerase. The physical explanation of this energy transfer is the following. In the vicinity of the polymerase the hydrophobic forces between base pairs are weaker than those in the rest of the DNA molecule. This is due to the fact that the RNA polymerase is accompanied by flat oily amino acids, which can insert themselves between the base pairs. Thus, close to polymerase it is easier to separate base pairs and so to untwist the DNA strands. The motion of the untwisted region is then induced by the motion of the RNA polymerase.

Within the framework of our model we were unable to describe the opening of DNA since we assumed that base pairs are rigid bodies. Nevertheless, one can expect that the untwisting is the prerequisite of opening. First, it may be easier for the untwisted sugar-phosphate backbones to be separated. Second, in the untwisted region the separation between the adjacent base pair is larger and the gaps are filled by oily aminoacids that accompany the RNA polymerase. This implies significantly smaller coupling between base pairs. Peyrard and Bishop (PB) [8] found that the denaturation temperature strongly depends on the assumed coupling (stacking interaction) between neighboring base pairs. The smaller the coupling the lower the denaturation temperature.

*Email address: tlipnia@ippt.gov.pl

Thus, one can expect that the thermal fluctuations (which are too small at physiological temperatures to denaturate the twisted DNA) lead to the formation of the denaturation bubble placed in the untwisted region. To show this we apply molecular dynamics simulations to nonhomogenous DNA, such that the stacking interaction is smaller at a given part of the chain, using the method proposed by Dauxois, Peyrard, and Bishop (DPB) [9].

III. MOLECULAR DYNAMICS SIMULATION IN PB MODEL

In the PB model each of the two DNA strands is represented by a set of point masses that correspond to the bases. The main characteristics of the model are as follows:

(i) Only the transversal displacements are considered. The displacement of nth base is denoted by w_n for one chain and v_n for the other.

(ii) To keep the model as simple as possible the adjacent bases of the same strand are assumed to be connected by a harmonic potential. On the other hand, the bonds connecting the two bases belonging to different strands are extremely stretched when the double helix opens locally so that their nonlinearity must not be ignored. The Morse potential will be used to represent the transverse interaction of the bases in a pair. It describes not only the hydrogen bonds but also the repulsive interactions of the phosphate groups, partly screened by the surrounding solvent action as well. In convenient variables $x_n=(w_n+v_n)/\sqrt{2}$ and $y_n=(w_n-v_n)/\sqrt{2}$ the Hamiltonian of the system reads [9]

$$H=\sum_n \left[\frac{1}{2}m\dot{x}_n^2+\frac{K}{2}(x_n-x_{n-1})^2\right]+\left[\frac{1}{2}m\dot{y}_n^2+\frac{K}{2}(y_n-y_{n-1})^2 + D(e^{-ay_n}-1)^2\right]. \quad (1)$$

The part of the Hamiltonian that depends on the variable x_n is decoupled from the stretching part. Following DPB we ignore this term in the statistical mechanics of the model. Therefore we confine ourselves only to the part of H that depends on y_n,

$$H_y=\sum_n \left[\frac{1}{2}m\dot{y}_n^2+\frac{K}{2}(y_n-y_{n-1})^2+D(e^{-ay_n}-1)^2\right] \quad (2)$$

and describes the dynamics of the stretching motions we are interested in.

The dynamics of the chain in contact with a thermal bath may be investigated by molecular dynamics simulation using the Nose [10] method. Beginning with the Hamiltonian H_y and the $2N$-dimensional phase space of chain of N base pairs with periodic boundary conditions, the fixed temperature canonical ensemble can be simulated by addition of a single variable s, which regulates the energy flow, its conjugate momentum p_s, and an additional parameter M, which fixes the scale of the temperature fluctuations. The augmented Hamiltonian [8] is then introduced in the form

$$H'=H'_y+\frac{p_s^2}{2M}+(N+1)k_BT\ln s, \quad (3)$$

with

$$H'_y=\sum_n \left[\frac{m\dot{y}_n^2}{2s^2}+\frac{K}{2}(y_n-y_{n-1})^2+D(e^{-ay_n}-1)^2\right], \quad (4)$$

where T denotes the temperature of the system and k_B is the Boltzmann's constant. In Hoover's [11] reformulation of Nose's method the variable s is removed, the equations of motion for s and $\dot{s}$ are replaced by a single equation, and the factor $(N+1)$ is replaced by N. Defining a thermodynamical friction coefficient $\xi=p_s/M$ and redefining the time interval $dt\rightarrow dt/s$, the Hoover formulation of equations of motion is

$$m\ddot{y}_n=K(y_{n+1}+y_{n-1}-2y_n)+2aD(e^{-ay_n}-1)e^{-ay_n} - \xi m\dot{y}_n, \quad (5)$$

where

$$\dot{\xi}=\frac{1}{M}\left[\sum_n m\dot{y}_n^2-Nk_BT\right]. \quad (6)$$

The appropriate units for the problem are electron volts, Angströms, and atomic mass units (amu). The resulting time unit is t.u.$=$(amu)$^{1/2}$ Å eV$^{-1/2}=1.0214\times10^{-14}$ s. DPB have performed simulations of Eqs. (5) and (6) mostly for a chain of 256 base pairs with periodic boundary conditions for temperatures varying from 150 K to over 500 K. They use (as most realistic) the following parameters: dissociation energy $D=0.04$ eV, a spatial scale factor of the Morse potential $a=4.45$ Å^{-1}, a coupling constant $K=0.06$ eV/Å^2, and mass $m=300$ amu. The constant of the Nose thermostat has been set to $M=1000$.

Recall that energy of 1 eV corresponds to 11 600 K. Hence at physiological temperature of 310 K the average energy for one degree of freedom is 0.027 eV, i.e., it is of the order of the assumed dissociation energy, 0.04 eV. In absence of stacking interaction the bases will in a relatively short time escape (one by one) from the Morse potential well. However, because of the fact that the adjacent bases are strongly coupled by stacking potential one may expect that for DPB parameters choice at 310° the bases will remain in the potential well. In the zero temperature limit all base pairs will be in the lowest energy state for which $y_n=0$. Obviously for finite temperatures $\langle|y_n|\rangle>0$ (where $\langle\,\rangle$ denotes average over time or ensemble) and moreover because the asymmetry of the Morse potential well one may expect that $\langle y_n\rangle>0$.

We will consider the DNA dynamics in the following three cases.

(1) The fixed DNA parameters ($D=0.04$ eV, $a=4.45$ Å^{-1}, $K=0.06$ eV/Å^2, and $m=300$ amu) and the temperature varied from 150 K to 450 K. For $T=450$ K we found that after a relatively short time (10^5 t.u. or 1 ns) the base separation for all base pairs grows over 5 Å that corre-

sponds to the distant plateau of the assumed Morse potential. In next 250 ns we observe only small fluctuations in which all particles remain on Morse plateau. We consider this state as DNA denaturation. For extensions larger than 5 Å the force F_m resulting from the Morse potential is very small $F_m < 7.7\times10^{-11}$ eV/Å, i.e., it is much smaller than within the Morse potential hole where it is of order $Da = 0.18$ eV/Å. In view of the above for so large average extensions the simulations are no longer reliable. Moreover from a physical point of view it is not very important whether the final base separation is 7 or 12 Å. For $T = 430$ K the transition between ''chaotic'' motion within the Morse potential well, and small fluctuations on the Morse plateau, took place after 40 ns. It is possible that in this case the transition was influenced by the fact that a relatively short chain of 300 base pairs with periodic boundary conditions has been considered. At $T = 420$ K we observe no transition during the 250-ns long run. We expect that the denaturation temperature is somewhere between 430 and 450 K for the DPB parameters choice. One should note that DPB considered even larger extension up to 15 Å (see Fig. 1, Ref. [9]); nevertheless for extensions larger than 5 Å they obtained very large scattering of results.

(2) The physiological temperature of 310 K and some DNA parameter are fixed as follows: $D = 0.04$ eV, $a = 4.45$ Å^{-1}, and $m = 300$ amu but the coupling constant K is varied from 0.02 to 0.06 eV/Å^2. For $K = 0.02$ eV/Å^2 we found that the average base pairs separation grows over 5 Å, which is regarded as the DNA denaturation.

(3) The physiological temperature of 310 K parameters: $D = 0.04$ eV, $a = 4.45$ Å^{-1}, and $m = 300$ amu, fixed and the coupling constant varying along the chain. Namely, we assume that $K = K_o = 0.06$ eV/Å^2 except the part of the chain of length 25 base pairs. Between the middle 21 base pairs from these $25K = K_{min}$, where K_{min} takes values from 0.01 to 0.05 eV/Å^2. Between the rest 4 base pairs we assume $K = K_1 = 2K_{min}/3 + K_o/3$ for the inner two and $K = K_2 = K_{min}/3 + 2K_o/3$ for the outer two, to avoid the sharp transition.

In these three cases we analyze the dynamics of chains of length of at least 250 base pairs (with periodic boundary condition) to calculate the mean base pair stretching $\langle y \rangle$ and the opening probabilities of a single base pair and of a segment of DNA chain. The mean stretching $\langle y \rangle$ is calculated by averaging over evolution time and ensemble. Obviously in the case of a uniform chain to calculate the average its suffices to average over time or ensemble since these two averages must be equal, in practice however, to have a faster convergence it is better to average over both time and ensemble. In the first two cases the ensemble is all the chain, while in the third case we average over the 21-base pairs long segment for which $K = K_{min}$.

From a biological point of view the opening probabilities are more important than the average value of y. Let $P_l(K)$ denote the probability that the single base pair ($l = 1$), or the DNA segment of prescribed length l is ''open.'' We said that (at time t) the nth base pair is open when $y_n(t)$ exceeds a threshold of $y_o = 1$ Å that corresponds to the base separation of $\sqrt{2}$ Å. This (rather arbitrary) threshold corresponds roughly to the ''edge'' of the plateau of the Morse potential. Similarly, we consider a DNA segment as open at time t when the average value of $y(t)$ in this segment exceeds $y_o = 1$ Å. The probability P_l is defined as

$$P_l = \frac{t_o}{t_T} \quad \text{for} \quad t_T \to \infty, \tag{7}$$

where t_o is the sum of time periods for which the segment is open and t_T is the total evolution time. In Ref. [9] another definition is used, namely, DPB defined denaturation rate counting a fraction of base pair that have a mean stretching exceeding a given threshold (see Fig. 7, Ref. [9]). Unfortunately, this simple definition produces numerical artefacts. This is due to the fact that when the evolution of uniform ''sufficiently long'' chain is considered over ''sufficiently long'' time the mean stretching of all base pairs should be almost equal. This follows from the principle that in equilibrium of a homogenous system consisting of identical particles, the average over time is equal to the average over ensemble. Thus after sufficiently long time the fraction of base pairs having the mean stretching larger than a given threshold should be 1 or 0, except for the situation in which the threshold is equal to the mean base pair stretching. The curves presented in Fig. 7, Ref. [9] results from the fact that the simulation time was relatively short and base pairs had different mean stretching. For the above reason we use here another definition, and calculate the opening probabilities, both for uniform and nonuniform chains.

In the simulations performed, the probabilities P_l have been estimated by checking at a large number ($N_T = 5000$) of time points if the given segment is open. Then

$$P_l \approx \frac{N_o}{N_T}, \tag{8}$$

where N_o is the number of time points for which the considered segment is open. The ''checking'' points were placed every 50 000 time steps.

IV. CONCLUSIONS

We have shown within the limits of the Peyrard-Bishop model that a denaturation bubble is likely to arise in low stacking region. From the data collected in Table I one can conclude that the drop of stacking interaction in a given segment of DNA leads to significant growth of the opening probability of that region. When, for example, the stacking interaction drops by 33% from its reference value the opening probability grows by a factor of four. The data collected in Tables I and II also indicated that the rest of the chain plays a stabilizing role. In the 21-base pair long, low stacking region placed within the chain of higher stacking constant, the fluctuations are smaller than in the chain for which stacking is low everywhere. This effect is especially significant for $K \leqslant 0.02$ when the average base separation in uniform chains grows over 5 Å, which can be considered a DNA denaturation. In the case of the nonuniform chain the

TABLE I. The nonuniform chain simulations. In the first column the value of K_{min} in 21-base pairs long, low stacking region is given; $\langle y^+\rangle(K_{min})$ is the average value of y in this region, while $P_{21}(K_{min})$ is the opening probability of the region. $K=0.06$ eV/Å^2 in the rest of the chain (see text for details).

K_{min} (eV/Å^2)	$\langle y^+\rangle(K_{min})$ (Å)	$P_{21}(K_{min})$
0.06	0.41	0.020
0.05	0.47	0.038
0.04	0.59	0.091
0.03	0.75	0.20
0.02	1.13	0.48
0.01	2.1	0.89

opening bubble is restricted to the low stacking region and even for $K_{min}=0.01$ eV/Å the average value of y in the region is 2.1 Å.

The proposed mechanism of local DNA opening, which accompanies RNA transcription, is the following: The RNA polymerase moves along the DNA chain due to chemical reactions and hence a torsional traveling wave may propagate, forced by advancing polymerase. In the untwisted region the coupling between neighboring base pairs becomes smaller than in the rest of the DNA chain. Thus, the thermal fluctuations (which are too small at physiological temperatures to denaturate the twisted DNA) lead to the formation of the denaturation bubble placed in the untwisted region.

ACKNOWLEDGMENT

This work was supported by KBN Grant No. 8 T07A 045 20.

APPENDIX: NUMERICAL SIMULATIONS

Without the last term in the right hand side, Eq. (5) describes the conservative dynamics of the chain. Parameter ξ in the last term regulates the energy flows between the Nose thermostat and the system. Positive ξ implies positive friction and leads to the decrease of the mechanical energy of the system. Negative ξ results in the system gaining energy from the thermostat. The parameter ξ is controlled by Eq. (6); it becomes smaller when the kinetic energy of the system is smaller than Nk_BT, and becomes larger in the opposite case. This mechanism connects the average value of H_y

TABLE II. Results from uniform chains simulations. P_1 is the single-base pair opening probability, while P_{21} (and P_{61}) are the 21-base (61-base) pairs long segment opening probability (see text for details).

K (eV/Å^2)	$\langle y\rangle$ (Å)	P_1	P_{21}	P_{61}
0.06	0.41	0.12	0.020	0.002
0.05	0.49	0.17	0.049	0.015
0.04	0.64	0.24	0.12	0.71
0.03	0.91	0.35	0.31	0.30
0.02	Denaturates			

TABLE III. Results from uniform chain simulations for fixed DNA parameters ($D=0.04$ eV, $a=4.45$ Å^{-1}, $K=0.06$ eV/Å^2, and $m=300$ amu) and temperatures varied from 150 K to 450 K. P_1 is the single-base pair opening probability, while P_{21} (and P_{61}) are the 21-base (61-base) pairs long segment opening probability (see text for details).

T (K)	$\langle y\rangle$ (Å)	P_1	P_{21}	P_{61}
150	0.093	0.002	$<10^{-4}$	$<10^{-4}$
250	0.24	0.04	$<10^{-3}$	$<10^{-4}$
310	0.41	0.12	0.020	0.002
350	0.59	0.21	0.096	0.044
380	0.79	0.31	0.22	0.20
420	1.25	0.46	0.48	0.58
450	Denaturates			

with temperature. During the typical simulation of Eqs. (5) and (6) the energy flows in both directions, once from thermostat to the system and once from the system to the thermostat. The presence of the thermostat has such nice property that even if the numerical scheme does not strictly conserve the mechanical energy, it may give reliable results. The simulations of systems (5) and (6) have been performed using the simplest first order scheme, with time step $t_s=0.1$ t.u. Such a numerical scheme does not strictly conserve the energy. In our case the total energy H', Eq. (3), has been slightly growing during the run. However, with a time step so chosen, the average absolute value of energy flow (between the system and the thermostat) when calculated per time step, has been four orders of magnitude larger than the total energy surplus. One may then expect that in such a case the numerical errors are absorbed by the thermostat. This conjecture has been confirmed by performing simulations with smaller t_s. This causes the energy surplus to become smaller, but does not alter the dynamics of the system. Obviously one cannot expect that using a simple numerical scheme will produce, after millions of time steps, the correct values of system variables y_i, one may expect only that the average values such as $\langle y\rangle$ and P_l are correct.

The initial velocities $\dot{y}_n$ were set to zero, while the positions y_n were chosen randomly. In a typical simulation of (at least) 250-base pair long chain the systems (5) and (6) have been integrated during 2.5×10^8 time steps or 250 ns. The data were collected after 2×10^6 time steps. Such a typical run takes roughly 10 h on average PC (Celeron 533 MHz) using a free Pascal compiler (see www.freepascal.org). Note that the integration time in our simulations is 50 times longer than reported by DPB. This longer time was needed because we were interested in averaging over relatively short segments of the DNA chain. For most sets of parameters for uniform chains we make one run, while for some sets two runs (more for $T=310$ K and $K=0.06$ eV/Å^2—with various chain lengths and integration times) have been performed to estimate the stochastic errors. For nonuniform chains two runs have been performed for each set of parameters. The magnitudes of expected stochastic errors of data

obtained for uniform chains and collected in Tables II and III are of order 3% for $\langle y \rangle$ and up to 10% for P_{21} and P_{61}—the errors are relatively small when the probabilities are large but significant in the cases when P_{21} and P_{61} are small. For nonuniform chains (Table I) the expected errors of $\langle y \rangle$ are of order of 5% up to 10% for P_{21}. Those values have been deduced by comparing the data collected in two runs or from comparing two halves of each run.

[1] S.W. Englander *et al.*, Proc. Natl. Acad. Sci. U.S.A. **777**, 7222 (1980).
[2] G. Gaeta, C. Reiss, M. Peyrard, and T. Dauxois, Nuovo Cimento D **17**, 1 (1994).
[3] L.V. Yakushevich, *Nonlinear Physics of DNA* (Wiley, New York, 1998).
[4] T. Lipniacki, Nuovo Cimento Soc. Ital. Fis., D **20D**, 833 (1998).
[5] T. Lipniacki, Phys. Rev. E **60**, 7253 (1999).
[6] H. Yin *et al.*, Science **270**, 1653 (1995).
[7] H.-Y. Wang *et al.*, Biophys. J. **74**, 1186 (1998).
[8] M. Peyrard and A.R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
[9] T. Dauxois, M. Peyrard, and A.R. Bishop, Phys. Rev. E **47**, 684 (1993).
[10] S. Nose, J. Chem. Phys. **81**, 511 (1984).
[11] W. Hoover, Phys. Rev. A **31**, 1695 (1985).