Sequence-Dependent Basepair Opening in DNA Double Helix

Andrew Krueger, Ekaterina Protozanova, and Maxim D. Frank-Kamenetskii
Center for Advanced Biotechnology and Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

ABSTRACT Preservation of genetic information in DNA relies on shielding the nucleobases from damage within the double helix. Thermal fluctuations lead to infrequent events of the Watson-Crick basepair opening, or DNA "breathing", thus making normally buried groups available for modification and interaction with proteins. Fluctuational basepair opening implies the disruption of hydrogen bonds between the complementary bases and flipping of the base out of the helical stack. Prediction of sequence-dependent basepair opening probabilities in DNA is based on separation of the two major contributions to the stability of the double helix: lateral pairing between the complementary bases and stacking of the pairs along the helical axis. The partition function calculates the basepair opening probability at every position based on the loss of two stacking interactions and one basepairing. Our model also includes a term accounting for the unfavorable positioning of the exposed base, which proceeds through a formation of a highly constrained small loop, or a ring. Quantitatively, the ring factor is found as an adjustable parameter from the comparison of the theoretical basepair opening probabilities and the experimental data on short DNA duplexes measured by NMR spectroscopy. We find that these thermodynamic parameters suggest nonobvious sequence dependent basepair opening probabilities.

INTRODUCTION

Thermal motion within the double helix plays a critical role in DNA structure and function. Basepair openings present the most dramatic deviations from the double helix ground state. Although rare, the events of Watson-Crick basepair opening make the active groups of DNA bases (which are otherwise buried within the double helix) accessible for interaction with proteins and chemicals (1–6). Fluctuational basepair opening implies the disruption of hydrogen bonds between the complementary bases and flipping of the base out of the helical stack (Fig. 1). In addition to being a spontaneous event, the base flipping constitutes a necessary mechanistic step of the enzyme-catalyzed DNA modifications, such as selective methylation by DNA methyltransferase family enzymes (7) and glucosylation of T4 DNA by β-glucosyl transferase (8). DNA damage repair enzymes, such as uracil DNA glycosylase (9,10), 8-oxoguanine DNA glycosylase I (11,12), and alkyl adenine DNA glycosylase (13) also utilize the base-flipping mechanism. When these enzymes interact with DNA, the substrate base is completely flipped out of the helix and is placed in a damaged base-specific pocket where catalysis takes place (14–16).

Kinetic and thermodynamic parameters of individual basepair opening in nucleic acids have been measured in NMR studies of the imino protons exchange as a function of the catalyst (3–6). In the case of the DNA double helix, the exchange of imino protons of guanines and thymines proceeds only from the open state of the basepair. The data on various DNA structures have been reported including B-DNA (17–21), B′-DNA (22–25), Z-DNA (26), drug-DNA complexes (27,28), protein-DNA complexes (9–11,13), DNA with mismatched basepairs (29), and DNA with modified bases (30).

In canonical B-DNA, basepairs predominantly open one at a time. The estimated basepair opening probabilities of A·T and G·C pairs are 10^{-5}–10^{-6} and 10^{-6}–10^{-7}, respectively. Early studies concur that the opening of basepairs is determined primarily by the nature of the basepair, and the neighboring pairs affect it only weakly (17). In special cases, however, basepair openings have been found to depend on the DNA sequence and structure. Five-times-smaller opening probabilities were observed in G·C basepairs within Z-DNA molecules (26). A·T pairs within the runs of four or more continuous adenines, which adopt a B′ conformation, exchange up to an order of magnitude slower (22–25). Sequence-dependent basepair opening dynamics in the DNA double helix has been also addressed in theoretical studies (31–36).

In this report, we use a statistical mechanics approach to calculate sequence-dependent opening probabilities ($K_{\mathrm{op}}^{(\mathrm{k})}$) for individual basepairs. Our approach relies on separation of two major contributions to DNA stability: lateral pairing between the complementary bases and stacking of pairs along the helical axis (37,38). Note that the separation of the two contributions is a prerequisite for our calculations; DNA stability parameters themselves present a cumulative quantity of stacking and basepairing factors (39,40). The probability of an open state of the basepair is calculated assuming that the event of basepair opening involves disruption of hydrogen bonds between the complementary bases and two stacking contacts, leading to the release of the bases positioning them extrahelically (Fig. 1). A major novel element of our treatment is an introduction of a new parameter, the ring factor, which accounts for the entropy penalty associated

Submitted December 1, 2005, and accepted for publication January 25, 2006.

Address reprint requests to Maxim D. Frank-Kamenetskii, Tel.: 617-353-8498; Fax: 617-353-8501; E-mail: mfk@bu.edu.

© 2006 by the Biophysical Society

0006-3495/06/05/3091/09 $2.00

doi: 10.1529/biophysj.105.078774

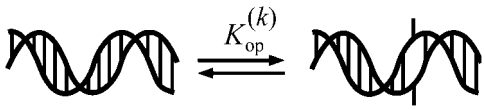

FIGURE 1 Fluctuational opening of one basepair in duplex DNA involves loss of pairing between complementary bases, disruption of stacking with neighboring pairs, and flipping the bases out of the helix leading to formation of a small highly constrained ring.

with the formation of a small constrained loop at the site of an open basepair. The value of the ring factor is estimated from the comparison of the theoretically obtained opening probability profiles for several sequences with the profiles measured for these sequences in imino proton exchange experiments (17–21). We argue that the ring factor we introduce leads to reevaluation of the physical nature of cooperativity of the DNA melting.

METHODS

DNA base-stacking and basepairing parameters

DNA stacking parameters, ΔG^{ST}_{KL}, for all $KL/K'L'$ DNA dinucleotide stacks (where K and K' (L and L') are complementary bases) were obtained using gel electrophoresis of DNA fragments carrying solitary nicks and gaps (37,38). ΔG^{ST}_{KL} values at different temperatures were calculated from parameters reported in Protozanova et al. (37) (corresponding to 37°C, 15 mM [NaCl]) by linear extrapolation with $d\Delta G^{ST}_{KL}/dT = 0.026$ kcal/mol·K determined experimentally by conducting gel electrophoresis at different temperatures (38). Analogously, we adjusted the ΔG^{ST}_{KL} values to the salt concentration used in NMR experiments using $d\Delta G^{ST}_{KL}/d\ln[Na^+] = -0.200$ kcal/mol determined by running gel electrophoresis under different salt conditions (38).

Basepairing parameters, $\Delta G^{BP}_{A\cdot T}$and $\Delta G^{BP}_{G\cdot C}$, were assessed as a difference between the stability parameters of DNA polymers consisting of A·T and G·C pairs, $\Delta G_{A\cdot T}$and $\Delta G_{G\cdot C}$, and the stacking contribution as follows (37,38):

$$\Delta G^{BP}_{A\cdot T} = \Delta G_{A\cdot T} - \frac{1}{4}\sum_{AT,TA,AA,TT}\Delta G^{ST}_{KL} \text{ and}$$
$$\Delta G^{BP}_{G\cdot C} = \Delta G_{G\cdot C} - \frac{1}{4}\sum_{GC,CG,GG,CC}\Delta G^{ST}_{KL}. \quad (1)$$

DNA stability parameters in terms of melting temperatures $T^{A\cdot T}_M$ and $T^{G\cdot C}_M$ are given by the following empirical equations for A·T- and G·C-containing polymers (41):

$$T^{A\cdot T}_M = 355.55 + 7.95\ln[Na^+] \text{ and}$$
$$T^{G\cdot C}_M = 391.55 + 4.89\ln[Na^+], \quad (2)$$

which translate into free-energy difference stability parameters using $\Delta S = -24.85$ cal/mol K and Eq. 3:

$$\Delta G = \Delta S(T_M - T). \quad (3)$$

Base-stacking and basepairing parameters for ''standard'' conditions of 37°C, 0.1 M NaCl are listed in Table 1. Parameters at other conditions were estimated as described above. Note that recently obtained stacking parameters for only contacts involving A·T and G·C basepairs at the chosen conditions (38) are very close, though not identical, to parameters in Table 1 for corresponding contacts, which were obtained from parameters determined in Protozanova et al. (37) via extrapolation as described above.

TABLE 1 Stacking ΔG^{ST}_{KL} and basepairing ΔG^{BP} parameters used in our calculations

ΔG^{ST}_{KL}	$5'$KL	A	T	G	C
$5'$	A	−1.49	−1.72	−1.44	−2.19
	T	−0.57	−1.49	−0.93	−1.81
	G	−1.81	−2.19	−1.82	−2.55
	C	−0.93	−1.44	−1.29	−1.82
ΔG^{BP}	A·T		0.64		
	G·C		0.12		

Stacking and basepairing parameters in kcal/mol correspond to 37°C, 0.1 M NaCl. They were obtained from the data in Protozanova et al. (37) via extrapolation as described in the Methods section.

Theoretical basepair opening probabilities

We apply a standard statistical mechanical approach to calculate the probabilities of basepair opening in dsDNA (31–33) using stacking and basepairing parameters found by us (37,38). The basepair opening event involves disruption of the hydrogen bonding between the complementary bases and disruption of the stacking interactions with both immediate neighbors (Fig. 1). The total partition function, Z_{Total}, of the N-bp-long DNA molecule is given by Eq. 4:

$$Z_{Total} = \sum_{\sigma_1=0,1}\sum_{\sigma_2=0,1}\cdots\sum_{\sigma_N=0,1}\prod_{i=1}^{N}(\delta_i)^{\sigma_i} \times(\alpha_i)^{\sigma_i}(\delta_{i+1})^{f(\sigma_i,\sigma_{i+1})}(\xi)^{f(\sigma_i,\sigma_{i+1})}, \quad (4)$$
$$\delta_1 = 1,\ \delta_{N+1} = 1, \quad \text{and} \quad \sigma_{N+1} = 1,$$

where $\sigma_i = 0$ corresponds to the closed state and $\sigma_i = 1$ corresponds to the open state of the basepair in the position, and

$$f(\sigma_i, \sigma_{i+1}) = \begin{cases} \sigma_i \text{ if } \sigma_{i+1} = 0 \\ 0 \text{ if } \sigma_{i+1} = 1 \end{cases},$$

and α_i, δ_i, and ξ are basepairing, stacking, and the entropic penalty parameter, respectively. The values of α_i and δ_i are calculated from the basepairing and base-stacking parameters as follows:

$$\alpha_i = \exp\left(\frac{\Delta G^{BP}_i}{RT}\right) \text{ and } \delta_i = \exp\left(\frac{\Delta G^{ST}_{i-1,i}}{RT}\right), \quad (5)$$

where $\Delta G^{BP}_i$is the basepairing parameter ($\Delta G^{BP}_{A\cdot T}$ or $\Delta G^{BP}_{G\cdot C}$) of the ith basepair and $\Delta G^{ST}_{i-1,i}$is the stacking parameters of the contact between the (i–1)th and ith positions (ΔG^{ST}_{KL}). Each term of partition function in Eq. 4 corresponds to a specific microscopic state of the DNA molecule with respect to the basepair opening/closing situation. For instance, the left-hand state in Fig. 1 (with all basepairs closed) contributes the term equal to 1 into the partition function, whereas the right-hand state in Fig. 1 with only the kth basepair open contributes the $\delta_k\alpha_k\delta_{k+1}\xi$ term.

Stacking and basepairing parameters used in our calculations correspond to ambient conditions of the NMR experiments. The entropic penalty ξ is an adjustment parameter. It is assumed to be the same for all small open regions independent of their size and sequence.

Conditional partition function of this molecule with one basepair in the kth position open is

$$Z_k = \sum_{\sigma_1=0,1}\sum_{\sigma_2=0,1}\cdots\sum_{\sigma_{k-1}=0,1}\sum_{\sigma_{k+1}=0,1}\cdots\sum_{\sigma_N=0,1}\prod_{i=1}^{N}(\delta_i)^{\sigma_i} \times(\alpha_i)^{\sigma_i}(\delta_{i+1})^{f(\sigma_i,\sigma_{i+1})}(\xi)^{f(\sigma_i,\sigma_{i+1})}, \quad (6)$$
$$\delta_1 = 1,\ \delta_{N+1} = 1,\ \sigma_{N+1} = 1, \quad \text{and} \quad \sigma_k = 1.$$

The only difference between Eqs. 4 and 6 is that in the latter summation over σ_k is omitted, whereas $\sigma_k = 1$ whenever it is met in the expression. This corresponds to the kth basepair being fixed in the open state whereas allowing all other basepairs to occupy both open and closed states. The opening probability of the basepair in the kth position, $K_{op}^{(k)}$, is the ratio of the two partition functions

$$K_{op}^{(k)} = Z_k / Z_{Total}. \quad (7)$$

The partition functions in Eqs. 4 and 6 can be calculated directly, using the well-known matrix representation (42,44,45) or via recursion equations (31,42,45). We used a MATLAB (The MathWorks, Natick, MA) program to calculate the partition function employing matrix representation as in Lazurkin et al. (42), Zimm (44), and Vedenov et al. (45). Corresponding equations for Z_{Total} and Z_k are as follows:

$$Z_{Total} = (1 \quad 1) \prod_{i=1}^{N-1} \begin{pmatrix} 1 & 1 \\ \delta_i\alpha_i\delta_{i+1}\xi & \delta_i\alpha_i \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \delta_N\alpha_N & \delta_N\alpha_N \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (8)$$

$$Z_k = (1 \quad 1) \prod_{i=1}^{k-1} \begin{pmatrix} 1 & 1 \\ \delta_i\alpha_i\delta_{i+1}\xi & \delta_i\alpha_i \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \delta_k\alpha_k\delta_{k+1}\xi & \delta_k\alpha_k \end{pmatrix} \times \prod_{i=k+1}^{N-1} \begin{pmatrix} 1 & 1 \\ \delta_i\alpha_i\delta_{i+1}\xi & \delta_i\alpha_i \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \delta_N\alpha_N & \delta_N\alpha_N \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (9)$$

Sequence-dependent basepair opening profiles were calculated for 10–17-bp-long molecules for which the profiles were determined in NMR experiments (Table 2). In our calculations of fluctuational basepair opening, we neglect the contribution stemming from the opening of duplex from the ends (fraying) by flanking the molecule with 10 G·C basepairs on both ends.

For each molecule, we determine the value of the adjustable parameter ξ, which gives the best correlation between theoretical $K_{op}^{(k)}$ values and the probabilities $K_{NMR}^{(k)}$ determined in NMR experiments for this sequence. Specifically, ξ (the ring factor) is assigned the value from minimization of $\langle D \rangle$ given by Eq. 10:

$$\langle D \rangle = \left[\frac{1}{N} \sum_k \left(\log \frac{K_{NMR}^{(k)}}{K_{op}^{(k)}} \right)^2 \right]^{1/2}, \quad (10)$$

where the summation is performed over all N unique basepairs for which the opening probability was measured in NMR experiments.

TABLE 2 DNA molecules considered in this study*

	Duplex sequence	Reference
I	CC **T3 T4 T5 C6 G6 A5 A4 A3** GG	(17)§
II	GG **A3 A4 A5 G6 C6 T5 T4 T3** CC	(17)§
III	CG **C3 G4 A5 T5 C4 G3** CG	(18)‡§
IV	CG **C3 A4 C5 A6 T6 G5 T4 G3** CG	(19)†
V	CG **C3 A4 G5 A6 T6 C5 T4 G3** CG	(19)†
VI	GC **G3 A4** T **C6** T **A8 T9 T10 T11 A12 T13 T14 T15** GC	(20)¶, (21)
VII	GC **G3** A **T5 C6** T **G8 T9 T10 C11 T12 A13** T **T15** GC	(20)¶, (21)

*Opening probabilities were measured in NMR experiments for basepairs shown in bold and numbered.

†Opening probabilities at 20°C were calculated using ΔH_{op} and ΔS_{op} values reported in Folta-Stogniew and Russu (19).

‡Opening probabilities were estimated as a ratio of the slope of an experimental dependence of exchange time on $1/[NH_3]$ (Fig. 8 of Leroy et al. (18)) to the imino proton transfer rate from the free mononucleotide per mole of the added base, which equals 2×10^8 M^{-1} s^{-1} (4).

§NMR measurements were performed at 15°C.

¶NMR measurements were performed at 20°C.

RESULTS

Basepair opening profiles

The event of fluctuational basepair opening involves disruption of the complementary basepair, disruption of stacking interactions with two neighboring basepairs, and formation of a small loop comprised of one open basepair (Fig. 1).

In our analysis, we consider seven dsDNA molecules for which the data on basepair opening were obtained in imino proton exchange experiments using NMR spectroscopy (Table 2). These molecules adopt a canonical B-DNA conformation for which stacking and basepairing parameters have been determined (37,38). Here we do not consider molecules with long runs of contiguous A·T pairs, which have been shown to display considerably different basepair opening kinetics and thermodynamics (22–25,29). This limits the pool of molecules for which the data on basepair fluctuational openings are available to seven sequences listed in Table 2.

Five out of seven molecules have palindromic sequences, thus reducing the number of unique basepairs studied in NMR experiments to four in molecules I, II, IV, and V and three in molecule III. Molecules VI and VII have non-palindromic sequences with 11 and 10 unique basepairs characterized in NMR experiments. Experimentally determined basepair opening profiles for internal sequences (shown in bold in Table 2) of molecules I–VII are presented in Figs. 2 and 3. The numbering scheme of basepairs corresponds to that in Table 2. Fig. 2 presents the data from Gueron and co-workers (17,18); these studies were performed at 15°C in the presence of 0.1 M NaCl. In Fig. 3 we show more recent results reported by Russu and co-workers (19,20); these experiments were carried out at 20°C, 0.1 M NaCl.

Sequence-dependent theoretical probability of the kth basepair opening event is obtained as a ratio of the conditional partition function Z_k (basepair in the kth position is open) to the total partition function of this molecule Z_{Total} (Eqs. 4–7). Partition functions are calculated using stacking and basepairing parameters adjusted to the conditions of the NMR experiments, i.e., 0.1 M NaCl and 15°C for molecules I–III or 20°C for molecules IV–VII (see Methods).

An additional parameter, ξ, or the ring factor, is introduced to represent the entropic penalty stemming from the structural constraints of the small loop (a ring) at the site of the open basepair in the double helix; we dub this parameter the ring factor. The introduction of this parameter proved to be absolutely necessary to reconcile theory and experiment. Indeed, let us consider the event of one basepair opening as shown in Fig. 1. Without the ring factor, the opening probability $K_{op}^{(k)}$ will be equal to $\exp(\Delta G/RT)$, where $\Delta G = \Delta G_{k-1,k}^{ST} + \Delta G_{k,k+1}^{ST} + \Delta G_k^{BP}$. According to the data we base our treatment upon Protozanova et al. (37) and Yakovchuk et al. (38); each stacking contributes, roughly, between 0 and −2.5 kcal/mol, whereas basepairing contributes very little. So if we assumed an average contribution from stacking as

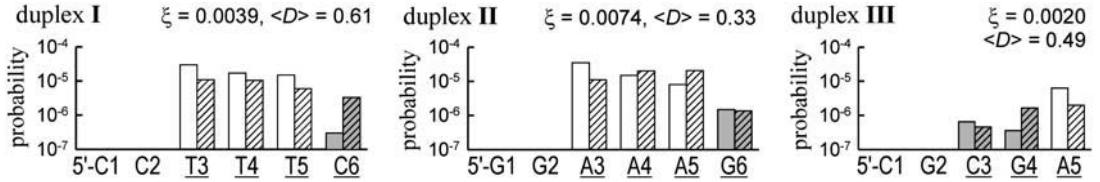

FIGURE 2 Comparison of the opening probabilities obtained in NMR experiments (bars with *open* or *shaded fill* for A·T or G·C pairs, respectively) with the results of theoretical calculations (bars with *patterned fill*) for duplexes I–III. The opening probabilities correspond to unique internal basepairs of palindromic sequences. Calculations were carried out with stacking and basepairing parameters corresponding to 15°C and 0.1 M [NaCl]. The ring factor ξ is indicated at each panel along with the least-square parameter $\langle D \rangle$ given by Eq. 10.

-1.25 kcal/mol and neglected the basepairing contribution, we would arrive at the conclusion that the basepair opening probability must be in the range of 10^{-2}, which is in an irreconcilable contradiction with the observed probabilities of $\sim 10^{-5}$ (see Introduction). Hence the absolute necessity of the introduction of the ring factor, ξ, which leads to an amended equation of the opening probability as

$$K_{\mathrm{op}}^{(\mathrm{k})} \approx \xi \exp(\Delta G/RT). \tag{11}$$

We assume that the ring factor is independent of the DNA sequence. We also assume that ξ is independent of the size of the open region. The partition functions Z_{Total} and Z_{k} given by Eqs. 4 and 6 include one penalty parameter for each open region. By varying ξ, we find the value of this adjustable parameter that gives the best correlation between the theoretical basepair opening probabilities and the values measured in the imino proton exchange experiments for these molecules.

Calculated sequence-dependent basepair opening probability profiles for molecules I–VII are presented in Figs. 2 and 3 along with the opening probabilities measured in NMR experiments. Note, that molecules I–V are palindromic, and opening probabilities of the unique half of each molecule are shown. The ring factor values obtained for each molecule are indicated by each panel. The first set of data (duplexes I and II) shows the distinct difference between the probabilities for G·C and A·T basepairs; variations between the A·T pairs is unremarkable. Our calculations are in a good agreement with the NMR data with the exception of the C6·G pair in duplex I, where the theoretical opening probability exceeds experimental data by almost an order of magnitude. A similar result is obtained for molecule III, where the difference in opening probabilities of G4·C and A5·T pairs observed experimentally is not reproduced by theoretical calculations. By contrast, in the case of molecule II, the experimental difference in openings of A·T pairs and the central G6·C pair is nicely reproduced by calculations. In all, the ring factor ranges from 7.4×10^{-3} to 2.0×10^{-3} for molecules II and III, respectively. The average value of $\xi = 4.4 \times 10^{-3}$ is obtained for this set of molecules.

The second set of data presents a much more interesting case since the variations of the opening probabilities for the A·T basepair depending on the immediate neighbors are observed in the NMR experiment. Indeed, there is an order of magnitude difference in the opening probabilities for A4·T and A6·T in duplex V. In this case, A4·T displays a ''normal'' behavior, whereas A6·T is apparently stabilized by the stronger stacking interaction with neighboring basepairs. This neighbor-dependent behavior is nicely reproduced by our calculations. In the case of duplex IV, the difference between A4·T and A6·T is insignificant both in NMR experiment and in our calculations. The average value of the

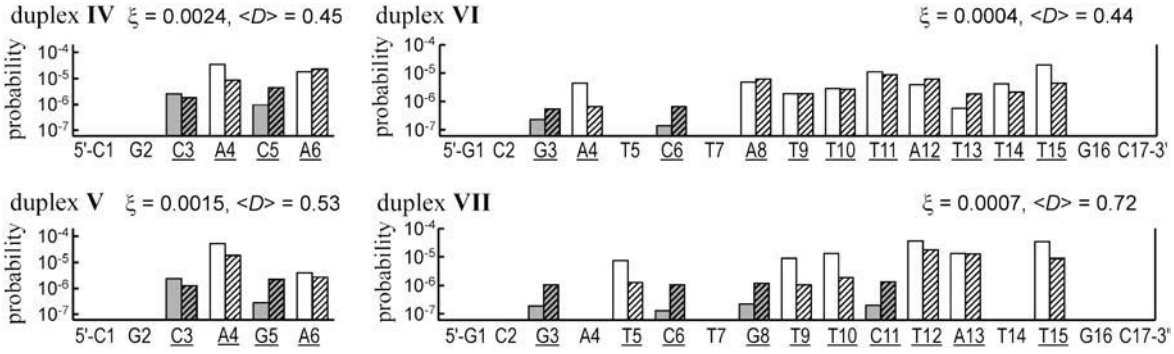

FIGURE 3 Comparison of the opening probabilities obtained in NMR experiments (bars with *open* or *shaded fill* for A·T or G·C pairs, respectively) with the results of theoretical calculations (bars with *patterned fill*) for duplexes IV–VII. The opening probabilities for four internal basepairs of palindromic sequences IV and V are shown. For duplexes VI and VII, experimental values are available only for basepairs that are underlined. Calculations were carried out with stacking and basepairing parameters corresponding to 20°C and 0.1 M [NaCl]. The ring factor ξ is indicated at each panel along with the least-square parameter $\langle D \rangle$ given by Eq. 10.

ring factor obtained in the analysis of duplexes IV-V is $\xi = 2.0 \times 10^{-3}$.

The most striking correlation between the experimental data and theoretical predictions is obtained for duplex VI. In this case, sequence-dependent variations in opening probabilities for A·T pairs are predicted remarkably well by our calculations.

In the case of duplex VII, the theoretical profile shows only slight variations of opening probabilities along the sequence of the molecule from the G3·C pair to the C11·G pair and an order of magnitude increase in opening probabilities for pairs T12·A through T15·A. Although relative changes in the opening probabilities for A·T pairs only are predicted well, there is a systematic difference between experimental and theoretical values for G·C pairs in this molecule. The average value of the ring factor for molecules VI–VII is 0.6×10^{-3}.

Temperature dependence of basepair opening probabilities

The experimental data on temperature dependence of fluctuational basepair opening for short DNA duplexes has been obtained for temperatures ranging from 10°C to 35°C (19,21). We used thermodynamic parameters of basepair opening to calculate experimental opening probabilities at different temperatures (ΔH_{op} and ΔS_{op} in the case of duplexes IV and V are given in Table 3 of Folta-Stogniew and Russu (19); ΔG_{op} and ΔH_{op} in the case of duplex VI are given in Tables 1 and 2 of Coman and Russu (21)). Experimental opening profiles corresponding to 37°C, 25°C, and 10°C are shown in Fig. 4.

To calculate theoretical opening probabilities at a given temperature, we used a set of base-stacking and basepairing parameters obtained at this temperature (see Methods). The theoretical basepair opening profiles are plotted in Fig. 4; corresponding values of the ring factor are indicated on each panel. In the case of duplex IV, ξ increases insignificantly (from 1.4×10^{-3} to 3.2×10^{-3}) when temperature decreases from 37°C to 10°C; the ring factor is constant at 1.5×10^{-3} for duplex V and at 0.4×10^{-3} for duplex VI throughout the temperature range. In fact, in the case of duplex V, a 25-fold increase in the fluctuational opening probability of the A6·T pair and a 13-fold increase of opening probability of the A4·T pair are observed in NMR experiments upon the rise in temperature from 10°C to 37°C. Theoretical calculations yield a 24-fold and a 16-fold increase for the A6·T and A4·T pairs, respectively, using constant ring factor and temperature-dependent stacking and basepairing parameters. A remarkable correlation between sequence-dependent experimental and theoretical opening probabilities is also observed for the A8·T-T14·A tract of duplex VI for all temperatures. As measured in NMR experiment, opening probabilities of these A·T pairs increase from 20- to 40-fold with the temperature rising from 10°C to 37°C. Profiles calculated with constant ξ closely match the experimental values for these temperatures. These data indicate that the temperature dependence of DNA stacking and basepairing parameters determines the temperature dependence of the basepair opening probabilities.

DISCUSSION

Nature of the open basepair

An algorithm for calculations of sequence-dependent basepair opening probabilities we apply relies on the assumption that the open state looks like in Fig. 1 (see also Fig. 5 *A*). We assume that the event of *k*th basepair opening contributes to the partition function the $\delta_k \alpha_k \delta_{k+1} \xi$ term (see Methods), which involves disruption of basepairing interaction between complementary partners (α_k) and the loss of stacking with both flanking helical interfaces (δ_k and δ_{k+1}) accompanied by flipping of both bases out of the helix with the formation of a constrained ring (ξ).

Strictly speaking, the exact nature of the transiently open state of the basepair is still unknown. A major question is

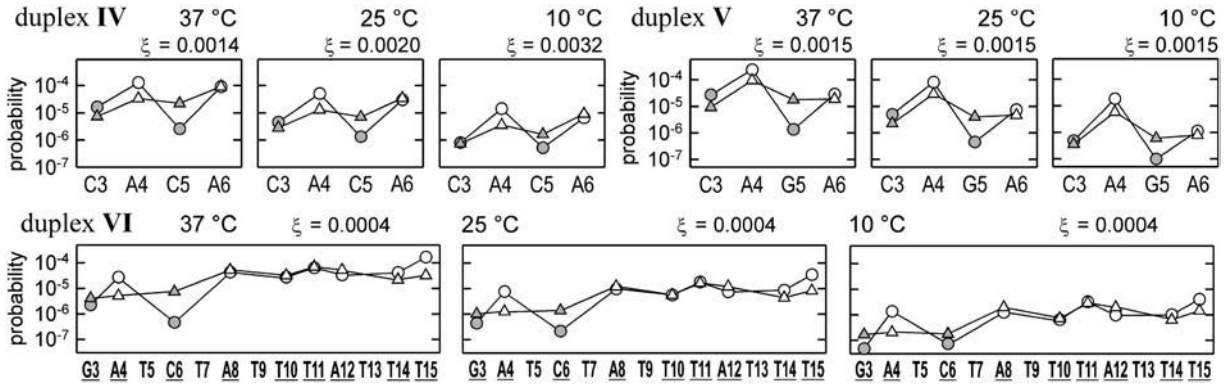

FIGURE 4 Experimental (*circles*) and theoretical (*triangles*) basepair opening probabilities of duplexes IV, V, and VI at different temperatures indicated at the top. In the case of duplex VI, experimental values are available only for basepairs that are underlined. The value of the ring factor are shown above each panel. Open and shaded fills indicate A·T and G·C pairs, respectively.

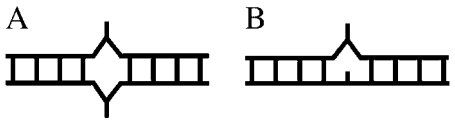

FIGURE 5 Schematic representation of possible open basepair configurations. (*A*) both bases are positioned extrahelically. (*B*) One base remains in the helical stack.

whether the opening is symmetric with both bases displaced from the helix (Fig. 5 *A*) or asymmetric with only one base extruded (Fig. 5 *B*). NMR experiments cannot definitely answer this question because kinetics of imino proton exchange of only one basepairing partner (G or T) can be measured. A prominent exception is the DNA molecules containing the G·T mismatched pair. Although the pair is markedly destabilized as compared to the G·C pair, both G(H1) and T(H3) exchange at similar rates (29,30,43) supporting the symmetric pathway in Figs. 1 and 5 *A*. Molecular dynamic simulations also reveal coupled rotation of both bases out of the helix (43).

On the other hand, an asymmetrically opened basepair has been observed in the case of DNA complexed to base-flipping enzymes (9–13). Evidently, this conformation of the open basepair is induced by the interaction with the protein, which stabilizes the flipped-out base in its catalytic pocket for inspection and excision or modification. There is no indication that asymmetrical configuration of the open basepair is assumed in naked DNA. Our algorithm based on the pathway in Fig. 5 *A* correctly predicts the basepair opening profiles detected by NMR; it is an additional, though indirect, indication that the pathway in Fig. 5 *B* does not take place for naked DNA. But in any case, we are unable to consider the pathway in Fig. 5 *B* because the stacking parameters necessary to describe this pathway are unavailable.

A recent analysis of kinetic and thermodynamic parameters of basepair openings in duplexes VI and VII has revealed a diverse behavior of basepairs depending on not only nearest neighbors but also more distant basepairs (21). Formation of a transition open state of the basepair has been proposed. The nature and energetics of this state depend on local sequence through hydrogen bonding interactions between the extrahelical base placed in the DNA groove and neighboring basepairs (35). Detailed characterization of this effect awaits further investigation; we are unable to consider this pathway quantitatively since its parameters are unknown.

The ring factor

In this report, we analyze sequence-dependent basepair opening profiles for dsDNA molecules existing in the B form. Based on the comparison of theoretical predictions with the experimentally obtained values for the opening probabilities, we determined the value of the entropic adjustable parameter for the formation of a small highly constrained ring (the ξ value).

The ring factor must be introduced for open regions in DNA since basepairing and base-stacking parameters are defined only for basepairs opening at the end of the helix. When the open region is formed within the double helix (surrounded by duplex pairs, as in Figs. 1 and 5 *A*) a very constrained ring is formed. The number of conformations of such a ring is greatly reduced as compared to the number of conformations corresponding to the open basepair at the very end of the double helix. The ring factor we introduce must be distinguished from the well-known loop factor entering the helix-coil transition theory of DNA (42,44–48). The loop factor (known also as the Jacobson-Stockmayer parameter) appears for large open regions, which include many persistent lengths of single-stranded DNA. By contrast, the ring factor is applied to very small open regions, mostly including only one basepair. As it is discussed below, the ring factor contributes into the cooperativity of DNA melting, rather than to the loop factor.

We assume that the ring factor is independent of the sequence of the open region as well as or the size of the ring. The first assumption is quite reasonable, since in our calculations the ring factor is introduced as the penalty stemming from structural constraints accompanying the event of the basepair opening; within the first approximation, these structural constraints should be similar for A·T and G·C basepairs. Further calculations involving dependence of the ring factor on the identity of the open basepairs and the sequence of the open region will be possible when more comprehensive and precise NMR data are available.

Secondly, we assign a single ring factor to every open region regardless of the size of this region. The ring factor introduced this way presents the penalty for the formation of the helix-coil boundary in the event of fluctuational basepair opening. Since the range of ambient conditions we consider is far below the DNA melting temperature, the opening of single pairs dominate; only at higher temperatures do larger open regions start forming (32). Here we neglect the effect arising from the formation of large open regions.

The independence of the ring factor of temperature is a major test of the validity of our algorithm for calculating the sequence-dependent basepair opening probability. Indeed, the ring factor is introduced in our calculations as an entropic penalty parameter arising from the formation of a structurally constrained small loop at the site of the open basepair. The ring factor is an intrinsic structural feature of the DNA backbone, which should be unaffected by temperature. Therefore, enhanced ''breathing'' of DNA basepairs at elevated temperatures is due to the temperature-dependent energetics of the double helix.

Quantitatively, the value of the ring factor ranges from 0.6×10^{-3} to 4×10^{-3} obtained from the comparison with different sets of NMR data; the ring factor does not significantly change with temperature. This range of values for the ring factor is a consequence of variability within the NMR data stemming from experimental imprecision or possibly

significant contributions from other pathways proposed in Coman and Russu (21). Note that our approach allows us to consider only the closed basepair to open basepair pathway shown in Figs. 1 and 5 *A*; we are unable to characterize other pathways due to the lack of parameters describing other configurations of the open state. More precise estimation of ξ is feasible upon accumulation of more accurate and more complete NMR data. Meanwhile, the estimate of $\xi \approx 10^{-3}$ is most reasonable.

The ring factor and DNA melting cooperativity

Temperature-, pressure- (49), or force-induced (50,51) melting of DNA double helix is a highly cooperative process. The cooperativity factor σ defines the free-energy cost for the initiation of the melted region in the helix (42,44,45). The ring factor stems from the formation of a small constrained ring in the double helix and it contributes to the cooperativity factor. The value of σ has been assessed in DNA melting experiments to be $\sigma \sim 10^{-5}$ for high salt conditions (52,53). A traditional interpretation of the DNA melting cooperativity relates the boundary initiation energy, i.e., $RT\ln\sigma = -7$ kcal/mol at 37°C to the loss of $n + 1$ stacking contacts upon melting of n consecutive basepairs (42,44). This figure exceeds greatly, by the absolute value, base-stacking parameters, which range, roughly, from 0 to −2.5 kcal/mol depending on the contact and on ambient conditions (see Protozanova et al. (37), Yakovchuk et al. (38), and Table 1). Thus, a naive view on the cooperativity of DNA melting as a result of disruption of $n + 1$ stacking interactions in the open region consisting of n basepairs definitely fails. Note that when we estimate the cooperativity parameter σ as the ring factor multiplied by the Boltzmann factor for stacking,

$$\sigma = \xi \exp\left(\frac{\Delta G^{ST}}{RT}\right), \tag{12}$$

we obtain a good agreement.

We conclude that the entropic penalty stemming from the structural constraints upon formation of small loops in dsDNA, i.e., the ring factor, is an essential factor in the cooperativity of DNA melting. This view of DNA melting cooperativity has been invoked by a microscopic theory of DNA melting (54). This approach factors in the restrictions on the backbone conformation imposed by the helical state and relies on microscopical characteristics of the molecular chain (the energy associated with the formation of a basepair and the number of conformations of the repeated unit) to describe helix-coil transition in DNA. In this case, melting cooperativity of the double-stranded polymer is defined by two factors: the persistence length of a DNA single strand and the number of available conformations of the repeated unit in the molten state (54). The interpretation of the melting cooperativity offered by the microscopic theory is consistent with the notion of the ring factor, which also stems from the conformational constraints imposed by the rigidity of DNA single strands. Experimentally, backbone rigidity has been shown to increase the distance of cooperativity transfer between two DNA helical domains through the anucleosidic bulge in elegant melting studies reported recently (55).

In energy terms, the ring factor translates into additional $RT\ln(1/\xi) \approx 4$ kcal/mol cost accompanying the event of DNA basepair opening. A small constrained loop formed in the case of fluctuational basepair opening is closely related to the junction between helical interfaces of B- and Z-forms of dsDNA. Under torsional stress of negative supercoiling, $d(CG)_n$ and $d(CA)_n$ tracts in recombinant plasmids undergo a right- to left-handed transition (56–62). The junction between the B- and the Z-form in DNA duplex is small and at least partially single-stranded, which underlies the sensitivity of the junctions to chemical agents and nucleases affecting only single-stranded DNA. Energetic cost of the formation of the B-Z junction has been estimated at 4–5 kcal/mol (60–64). This unfavorable contribution is in a very good agreement with our estimation of the energy cost associated with the ring factor (see above). This agreement strongly indicates that at the site of a B-Z junction, the fluctuationally open state of a base-pair is ''frozen''. This picture is very much consistent with the recent x-ray crystallography data, which have found one basepair to be open at the B-Z junction induced by protein binding (65).

CONCLUSION

Basing on separation of base-stacking and basepairing contributions into the DNA stability and direct computation of the partition function for DNA, we predict the basepair opening probability profiles for dsDNA molecules. An entropic penalty parameter, the ring factor, accompanying formation of small loops, is estimated from the comparison of theoretical probabilities with the probabilities measured for these molecules by NMR experiments. Theory and experiment agree that only individual basepairs open well below DNA melting temperature, and the calculated basepair opening profiles agree reasonably well with NMR data for a variety of sequences. As could be expected on the basis of a wide variation of the stacking parameter values for different stacks, the opening probability depends not only on the nature of the basepair but also on its immediate neighbors.

We thank Alex Vologodskii and Yevgeni Mamasakhlisov for helpful discussions.

REFERENCES

1. Frank-Kamenetskii, M. D. 1987. How the double helix breathes. *Nature*. 328:17–18.
2. Frank-Kamenetskii, M. D. 1985. Fluctuational motility of DNA. *In* Structure and Motion: Membranes, Nucleic Acids and Proteins. E. Clementi, editor. Adenine Press, New York. 417–432.

3. Gueron, M., M. Kochoyan, and J. L. Leroy. 1987. A single mode of DNA base-pair opening drives imino proton exchange. *Nature*. 328: 89–92.
4. Gueron, M., E. Charretier, J. Hagerhorst, M. Kochoyan, J. L. Leroy, and A. Moraillon. 1990. Applications of Imino Proton Exchange to Nucleic Acid Kinetics and Structures. *In* Structure and Methods. R. H. Sarma and M. H. Sarma, editors. Adenine Press, New York.: 113–137.
5. Gueron, M., and J. L. Leroy. 1995. Studies of base pair kinetics by NMR measurement of proton exchange. *Methods Enzymol.* 261: 383–413.
6. Russu, I. M. 2004. Probing site-specific energetics in proteins and nucleic acids by hydrogen exchange and nuclear magnetic resonance spectroscopy. *Methods Enzymol.* 379:152–175.
7. Cheng, X., and R. J. Roberts. 2001. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.* 29: 3784–3795.
8. Lariviere, L., and S. Morera. 2002. A base-flipping mechanism for the T4 phage beta-glucosyltransferase and identification of a transition-state analog. *J. Mol. Biol.* 324:483–490.
9. Parikh, S. S., G. Walcher, G. D. Jones, G. Slupphaug, H. E. Krokan, G. M. Blackburn, and J. A. Tainer. 2000. Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl. Acad. Sci. USA*. 97:5083–5088.
10. Cao, C., Y. L. Jiang, J. T. Stivers, and F. Song. 2004. Dynamic opening of DNA during the enzymatic search for a damaged base. *Nat. Struct. Mol. Biol.* 11:1230–1236.
11. Bruner, S. D., D. P. Norman, and G. L. Verdine. 2000. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*. 403:859–866.
12. Banerjee, A., W. Yang, M. Karplus, and G. L. Verdine. 2005. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damages DNA. *Nature*. 434:612–618.
13. Lau, A. Y., M. D. Wyatt, B. J. Glassner, L. D. Samson, and T. Ellenberger. 2000. Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc. Natl. Acad. Sci. USA*. 97:13573–13578.
14. Roberts, R. J. 1995. On base flipping. *Cell*. 82:9–12.
15. Stivers, J. T. 2004. Site-specific DNA damage recognition by enzyme-induced base flipping. *Prog. Nucleic Acid Res. Mol. Biol.* 77:37–65.
16. David, S. S., and S. D. Williams. 1998. Chemistry of glycosylases and endonucleases involved in base-excision repair. *Chem. Rev.* 98:1221–1262.
17. Kochoyan, M., J. L. Leroy, and M. Gueron. 1987. Proton exchange and base-pair lifetimes in a deoxy-duplex containing a purine-pyrimidine step and in the duplex of inverse sequence. *J. Mol. Biol.* 196: 599–609.
18. Leroy, J. L., M. Kochoyan, T. Huynh-Dinh, and M. Gueron. 1988. Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *J. Mol. Biol.* 200: 223–238.
19. Folta-Stogniew, E., and I. M. Russu. 1994. Sequence dependence of base-pair opening in a DNA dodecamer containing the CACA/GTGT sequence motif. *Biochemistry*. 33:11016–11024.
20. Coman, D., and I. M. Russu. 2005. Base-pair opening in three DNA-unwinding elements. *J. Biol. Chem.* 280:20216–20221.
21. Coman, D., and I. M. Russu. 2005. A nuclear magnetic resonance investigation of the energetics of the base-pair opening pathways in DNA. *Biophys. J.* 89:3285–3292.
22. Leroy, J. L., E. Charretier, M. Kochoyan, and M. Gueron. 1988. Evidence from base-pair kinetics for two types of adenine tract structures in solution: their relation to DNA curvature. *Biochemistry*. 27:8894–8898.
23. Moe, J. G., and I. M. Russu. 1990. Proton exchange and base-pair opening kinetics in 5′-d(CGCGAATTCGCG)-3′ and related dodecamers. *Nucleic Acids Res.* 18:821–827.
24. Moe, J. G., E. Folta-Stogniew, and I. M. Russu. 1995. Energetics of base pair opening in a DNA dodecamer containing an A3T3 tract. *Nucleic Acids Res.* 23:1984–1989.
25. Chen, C., and I. M. Russu. 2004. Sequence-dependence of the energetics of opening of at basepairs in DNA. *Biophys. J.* 87:2545–2551.
26. Kochoyan, M., J. L. Leroy, and M. Gueron. 1990. Processes of base-pair opening and proton exchange in Z-DNA. *Biochemistry*. 29: 4799–4805.
27. Leroy, J. L., X. L. Gao, M. Gueron, and D. J. Patel. 1991. Proton exchange and internal motions in two chromomycin dimer-DNA oligomer complexes. *Biochemistry*. 30:5653–5661.
28. Leroy, J. L., X. L. Gao, V. Misra, M. Gueron, and D. J. Patel. 1992. Proton exchange in DNA-luzopeptin and DNA-echinomycin bisintercalation complexes: rates and processes of base-pair opening. *Biochemistry*. 31:1407–1415.
29. Moe, J. G., and I. M. Russu. 1992. Kinetics and energetics of base-pair opening in 5′-d(CGCGAATTCGCG)-3′ and a substituted dodecamer containing G.T mismatches. *Biochemistry*. 31:8421–8428.
30. Bohon, J., and C. R. de Los Santos. 2005. Effect of 6-thioguanine on the stability of duplex DNA. *Nucleic Acids Res.* 33:2880–2886.
31. Vologodskii, A. V., and M. D. Frank-Kamenetskii. 1975. Theoretical study of DNA unwinding under the action of formaldehyde. *J. Theor. Biol.* 55:153–166.
32. Lukashin, A. V., A. V. Vologodskii, M. D. Frank-Kamenetskii, and Y. L. Lyubchenko. 1976. Fluctuational opening of the double helix as revealed by theoretical and experimental study of DNA interaction with formaldehyde. *J. Mol. Biol.* 108:665–682.
33. Wartell, R. M., and A. S. Benight. 1982. Fluctuational base-pair opening in DNA at temperatures below the helix-coil transition region. *Biopolymers.* 21:2069–2081.
34. Keepers, J., P. A. Kollman, and T. L. James. 1984. Molecular mechanical studies of base-pair opening in d(CGCGC):d(GCGCG), dG5. dC5, d(TATAT):d(ATATA), and dA5. dT5 in the B and Z forms of DNA. *Biopolymers.* 23:2499–2511.
35. Banavali, N. K., and A. D. MacKerell Jr. 2002. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J. Mol. Biol.* 319:141–160.
36. Giudice, E., and R. Lavery. 2003. Nucleic acid base pair dynamics: the impact of sequence and structure using free-energy calculations. *J. Am. Chem. Soc.* 125:4998–4999.
37. Protozanova, E., P. Yakovchuk, and M. D. Frank-Kamenetskii. 2004. Stacked-unstacked equilibrium at the nick site of DNA. *J. Mol. Biol.* 342:775–785.
38. Yakovchuk, P., E. Protozanova, and M. D. Frank-Kamenetskii. 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* 34:564–574.
39. Vologodskii, A. V., B. R. Amirikyan, Y. L. Lyubchenko, and M. D. Frank-Kamenetskii. 1984. Allowance for heterogeneous stacking in the DNA helix-coil transition theory. *J. Biomol. Struct. Dyn.* 2:131–148.
40. SantaLucia, J., Jr. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*. 95:1460–1465.
41. Frank-Kamenetskii, M. D. 1971. Simplification of the empirical relationship between melting temperature of DNA, its GC content and concentration of sodium ions in solution. *Biopolymers.* 10:2623–2624.
42. Lazurkin, Y. S., M. D. Frank-Kamenetskii, and E. N. Trifonov. 1970. Melting of DNA: its study and application as a research method. *Biopolymers.* 9:1253–1306.
43. Varnai, P., M. Canalia, and J. L. Leroy. 2004. Opening mechanism of G·T/U pairs in DNA and RNA duplexes: a combined study of imino proton exchange and molecular dynamics simulation. *J. Am. Chem. Soc.* 126:14659–14667.
44. Zimm, B. H. 1960. Theory of melting of the helical form in double chains of the DNA type. *J. Chem. Phys.* 33:1349–1356.

45. Vedenov, A. A., A. M. Dykhne, and M. Frank-Kamenetskii. 1972. The helix-coil transition in DNA. *Sov. Phys. Usp.* 14:715–736.

46. Cantor, C. R., and P. R. Schimmel. 1980. Biophysical Chemistry Part III: The Behavior of Biological Macromolecules. W. H. Freeman, San Francisco.

47. Wartell, R. M., and A. S. Benight. 1985. Thermal denaturation of DNA molecules. *Phys. Rep.* 126:67–107.

48. Wada, A., and A. Suyama. 1986. Local stability of DNA and RNA secondary structure and its relation to biological functions. *Prog. Biophys. Mol. Biol.* 47:113–157.

49. Dubins, D. N., A. Lee, R. B. Macgregor, Jr., and T. V. Chalikian. 2001. On the stability of double stranded nucleic acids. *J. Am. Chem. Soc.* 123:9254–9259.

50. Williams, M. C., and I. Rouzina. 2002. Force spectroscopy of single DNA and RNA molecules. *Curr. Opin. Struct. Biol.* 12:330–336.

51. Weeks, J. D., J. B. Lucks, Y. Kafri, C. Danilowicz, D. R. Nelson, and M. Prentiss. 2005. Pause point spectra in DNA constant-force unzipping. *Biophys. J.* 88:2752–2765.

52. Amirikyan, B. R., A. V. Vologodskii, and Y. L. Lyubchenko. 1981. Determination of DNA cooperativity factor. *Nucleic Acids Res.* 9:5469–5482.

53. Kozyavkin, S. A., S. M. Mirkin, and B. R. Amirikyan. 1987. The ionic strength dependence of the cooperativity factor for DNA melting. *J. Biomol. Struct. Dyn.* 5:119–126.

54. Morozov, V. F., E. S. Mamasakhlisov, S. Hayryan, and C. K. Hu. 2000. Microscopical approach to the helix-coil transition in DNA. *Physica A.* 281:51–59.

55. Becaud, J., I. Pompizi, and C. J. Leumann. 2003. Propagation of melting cooperativity along the phosphodiester backbone of DNA. *J. Am. Chem. Soc.* 125:15338–15342.

56. Singleton, C. K., J. Klysik, S. M. Stirdivant, and R. D. Wells. 1982. Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature.* 299:312–316.

57. Peck, L. J., A. Nordheim, A. Rich, and J. C. Wang. 1982. Flipping of cloned d(pCpG)n.d(pCpG)n DNA sequences from right- to left-handed helical structure by salt, Co(III), or negative supercoiling. *Proc. Natl. Acad. Sci. USA.* 79:4560–4564.

58. Haniford, D. B., and D. E. Pulleyblank. 1983. Facile transition of poly[d(TG) × d(CA)] into a left-handed helix in physiological conditions. *Nature.* 302:632–634.

59. O'Connor, T. R., D. S. Kang, and R. D. Wells. 1986. Thermodynamic parameters are sequence-dependent for the supercoil-induced B to Z transition in recombinant plasmids. *J. Biol. Chem.* 261:13302–13308.

60. Ho, P. S. 1994. The non-B-DNA structure of d(CA/TG)n does not differ from that of Z-DNA. *Proc. Natl. Acad. Sci. USA.* 91:9549–9553.

61. Peck, L. J., and J. C. Wang. 1983. Energetics of B-to-Z transition in DNA. *Proc. Natl. Acad. Sci. USA.* 80:6206–6210.

62. Singleton, C. K., J. Klysik, and R. D. Wells. 1983. Conformational flexibility of junctions between contiguous B- and Z-DNAs in supercoiled plasmids. *Proc. Natl. Acad. Sci. USA.* 80:2447–2451.

63. Frank-Kamenetskii, M. D., and A. V. Vologodskii. 1984. Thermodynamics of the B-Z transition in superhelical DNA. *Nature.* 307: 481–482.

64. Vologodskii, A. V., and M. D. Frank-Kamenetskii. 1984. Left-handed Z form in superhelical DNA: a theoretical study. *J. Biomol. Struct. Dyn.* 1:1325–1333.

65. Ha, S. C., K. Lowenhaupt, A. Rich, Y. G. Kim, and K. K. Kim. 2005. Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature.* 437: 1183–1186.