Biophysical Journal Volume 105 October 2013 1904–1914

Temperature Dependence of the DNA Double Helix at the Nanoscale: Structure, Elasticity, and Fluctuations

Sam Meyer,[†‡] Daniel Jost,[†‡] Nikos Theodorakopoulos,[§¶] Michel Peyrard,[†] Richard Lavery,[||*] and Ralf Everaers[†‡]
[†]Université de Lyon, Laboratoire de Physique, Ecole Normale Supérieure de Lyon, Lyon, France; [‡]Centre Blaise Pascal, Ecole Normale Supérieure de Lyon, Lyon, France; [§]Theoretical and Physical Chemistry Institute, National Hellenic Research Foundation, Athens, Greece; [¶]Fachbereich Physik, Universität Konstanz, Konstanz, Germany; and [||]Bases Moléculaires et Structurales des Systemes Infectieux, University Lyon I/Centre National de la Recherche Scientifique, Institut de Biologie et Chimie des Protéines, Lyon, France

ABSTRACT Biological organisms exist over a broad temperature range of −15°C to +120°C, where many molecular processes involving DNA depend on the nanoscale properties of the double helix. Here, we present results of extensive molecular dynamics simulations of DNA oligomers at different temperatures. We show that internal basepair conformations are strongly temperature-dependent, particularly in the stretch and opening degrees of freedom whose harmonic fluctuations can be considered the initial steps of the DNA melting pathway. The basepair step elasticity contains a weaker, but detectable, entropic contribution in the roll, tilt, and rise degrees of freedom. To extend the validity of our results to the temperature interval beyond the standard melting transition relevant to extremophiles, we estimate the effects of superhelical stress on the stability of the basepair steps, as computed from the Benham model. We predict that although the average twist decreases with temperature in vitro, the stabilizing external torque in vivo results in an increase of ~1°/bp (or a superhelical density of $\Delta\sigma \simeq +0.03$) in the interval 0–100°C. In the final step, we show that the experimentally observed apparent bending persistence length of torsionally unconstrained DNA can be calculated from a hybrid model that accounts for the softening of the double helix and the presence of transient denaturation bubbles. Although the latter dominate the behavior close to the melting transition, the inclusion of helix softening is important around standard physiological temperatures.

INTRODUCTION

DNA is the common substrate of genetic information in all living organisms. The mechanical properties of the DNA double helix play a crucial role in the molecular processes related to the replication and the regulated transcription of this information: examples include the tight wrapping of DNA around histone (1) and histonelike (2) proteins in prokaryotes, and sequence recognition by other molecules such as the TATA-box binding protein (3). With biological organisms living at very different temperatures (so-called extremophiles thrive over a temperature range of −15°C to +120°C (4,5)), the question arises of how the properties of DNA vary with temperature.

To discuss the basic ideas of this article regarding the temperature dependence of the DNA double-helix elasticity at different length scales, it is useful to briefly consider generic springlike degrees of freedom. In the simplest example of a harmonic spring, the excitation free energy has the form $F(x) = 1/2\ kx^2$, where k is the spring stiffness. In a molecular mechanics force field, this functional form applies to bond lengths and suitable bond angles. Deformations lead to a purely energetic or enthalpic response. The stiffness of such springs is independent of temperature, $k(T) = k_h$, whereas the amplitude of the corresponding thermal fluctuations, $\langle x^2 \rangle = k_B T/k$, is directly proportional to the temperature. A first complication arises if the mechanical force field is anharmonic. In this case, the linear response of the system to external forces and torques described by a harmonic approximation becomes temperature-dependent. The opposite extreme from enthalpic springs is entropic springs, with $k(T) = -k_s T$ and $\langle x^2 \rangle = k_B/(-k_s)$ independent of temperature. Such springs represent the behavior of soft matter at much larger scales. The best-known examples are the entropic springs substituting random-walk-like polymer chains in the theory of rubber elasticity (6). In this case, $k_s<0$ as the entropy decreases with the extension of the polymer chain. As a result, a rubber band under a mechanical load contracts when its temperature is raised. The nanoscale elasticity of DNA combines all these complications. The double helix has a well defined average shape and its local behavior is not dominated by fluctuations. However, since the description results from integrating out more microscopic degrees of freedom, nanoscale force fields necessarily represent temperature-dependent deformation free energies. In the harmonic approximation and over a finite temperature range, the temperature dependence can always be written in the form $k(T) = k_h - Tk_s$ and can be inferred from the amplitude of thermal fluctuations observed at different temperatures, $k(T) = k_B T/\langle x^2 \rangle_T$. For $k_s \leq = 0$, the results can, at least formally, be extrapolated to a temperature higher than those used to calibrate the harmonic model. If, as turns out to be the case, $k_s>0$, the results cannot be extrapolated beyond a spinodal temperature $T_s = k_h/k_s$, where the spring

Submitted July 24, 2013, and accepted for publication September 9, 2013.

*Correspondence: richard.lavery@ibcp.fr

Sam Meyer's present address is LIRIS, CNRS UMR 5205, INRIA, INSA Lyon, Université de Lyon, F-69621 Lyon, France.

Editor: Michael Levitt.

0006-3495/13/10/1904/11 $2.00

http://dx.doi.org/10.1016/j.bpj.2013.09.004

ceases to resist extension and thus becomes mechanically unstable. Consistency requires that the spinodal temperature be much higher than the actual melting temperature, T_m, where the two DNA strands separate. In this case, the nanomechanical description of the double helix remains valid for $T_m<T \ll T_s$, even though this state now only represents a metastable local free-energy minimum and no longer the global free-energy minimum. If DNA only undergoes partial melting, the mechanical properties of a given basepair step can be described by two-state spring models, which account for the different structures and elastic properties in the helix and coil sections of the molecule (7–9). For the calculation of suitable averages, the relative statistical weights of the two states need to be inferred from sequence-dependent models of DNA melting (10–13). Interestingly, external mechanical forces and torques may change the relative stability of the helix and coil states and hence modify the melting temperature (14).

In this article, we address the temperature-dependent properties of DNA over a wide span of length scales. We start from an atomistic model (15), with which we investigate DNA oligomers at different temperatures, thereby extending previous molecular dynamics (MD) results of the ABC consortium (16,17). The central part of our work deals with the nanoscale structure, elasticity, and stability of the double helix. In particular, we propose and parameterize a temperature-dependent generalization of the rigid-base and rigid-basepair models of DNA (18). Although our simulations were performed for torsionally unconstrained DNA, we estimate the effects of superhelical stress by introducing temperature-dependent torques estimated from the Benham model (14). Biological control of such a torque can serve to regulate spontaneous DNA opening at transcription start sites (19). In particular, the introduction of positive superhelical stress makes it possible to stabilize the double helix in a temperature interval beyond the standard melting transition (14), which is comparable to the conditions under which extremophiles exist. In the final step, we consider DNA on the wormlike chain level. We coarse-grain the nanoscale model to determine the temperature-dependent persistence length of the double helix (20) and reevaluate a recent kinked wormlike chain model (8) to account for the presence of transient denaturation bubbles (7) in the estimation of the apparent persistence length of DNA (9).

The article is organized as follows. The Models and Methods section describes the temperature-dependent rigid-base and rigid-basepair models, the MD simulations, the coupling with the Benham model, the coarse-graining to the wormlike chain model, and the inclusion of transient denaturation bubbles in the calculation of an apparent bending persistence length. Details of the analysis of the simulation data are described in the Supporting Material. In the Results section, we present our findings for the entropic contribution to the DNA nanoscale elasticity. This contribution is particularly strong for the internal basepair elasticity. At the basepair step level, the effect is weaker but detectable, and it results in a softening of the large-scale stiffness of the molecule. In the Discussion section, we address the temperature dependence of internal fluctuations of the double helix, the path to the melting transition, the effect of superhelical stress on the properties of the double helix in vivo, and the large-scale bending rigidity of DNA, including the temperature dependence of the apparent persistence length over the whole experimentally studied temperature interval. The article closes with a brief conclusion.

MODELS AND METHODS

DNA elasticity in the rigid-base and rigid-basepair models

We consider the fluctuations of the DNA double helix at two successive nanoscale levels, 1), inside the basepair (intra parameters), and 2), between adjacent basepairs (step parameters). In both cases, the conformation is described by a six-dimensional vector, $\underline{q} = (q_1, \ldots, q_6)$, corresponding to the relative orientation and position of the relevant objects (bases or basepairs), in conventional notation: buckle, propel, opening, shear, stretch, and stagger in the first case, and tilt, roll, twist, shift, slide, and rise in the second (18).

In the harmonic approximation, the Gibbs free energy reduces to a quadratic form:

$$G(\underline{q},\sigma,T) = \frac{1}{2}(\underline{q}-\underline{q}_0(\sigma,T))^t \underline{\underline{k}}(\sigma,T)(\underline{q}-\underline{q}_0(\sigma,T)), \quad (1)$$

where T is the temperature, σ is the DNA sequence, and we have introduced the six-dimensional equilibrium conformation $\underline{q}_0(\sigma,T)$ and the 6×6 stiffness matrix $\underline{\underline{k}}(\sigma,T)$. The latter describes the couplings between the fluctuations of the different degrees of freedom and is proportional to the inverse of the covariance matrix,

$$\underline{\underline{C}}(\sigma,T) = k_B T\, \underline{\underline{k}}(\sigma,T)^{-1}. \quad (2)$$

For the rigid-base(pair) model of DNA, previous studies have extracted the sequence-dependent elastic parameters at room temperature, $T_0 = 300$ K, from the analysis of DNA and DNA-protein crystallographic structures (21), from MD simulations of DNA oligomers (17,22), or from combinations of both approaches (23). Note that for an ensemble of crystal structures, the passage from the observed covariance matrix to the stiffness (Eq. 2) relies on the existence of an effective temperature governing the ensemble of experimentally observed conformations (22,23). This approach can therefore not be used to study the actual temperature dependence of the mechanical properties of DNA. Instead, we extend previous MD simulations of DNA oligomers (16,17) to a broad range of temperatures

to study the sequence and temperature dependence of DNA structure and elasticity.

Inclusion of temperature effects

The knowledge of the elastic parameters at T_0 gives no information on their temperature dependence around T_0, which depends on the relative importance of the enthalpic and entropic contributions:

$$G(\underline{q},\sigma,T) = H(\underline{q},\sigma,T_0) - T\,S(\underline{q},\sigma,T_0). \quad (3)$$

Here, H and S are quadratic forms:

$$H(\underline{q},\sigma,T_0) = \frac{1}{2}(\underline{q}-\underline{q}_0^h(\sigma))^t\,\underline{\underline{k}}_h(\sigma)(\underline{q}-\underline{q}_0^h(\sigma)) \quad (4)$$

$$S(\underline{q},\sigma,T_0) = \frac{1}{2}(\underline{q}-\underline{q}_0^s(\sigma))^t\,\underline{\underline{k}}_s(\sigma)(\underline{q}-\underline{q}_0^s(\sigma)). \quad (5)$$

From Eq. 3, the temperature-dependent elastic parameters $\underline{\underline{k}}(\sigma,T)$ and $\underline{q}_0(\sigma,T)$ of Eq. 1 can be written in terms of enthalpic and entropic contributions. For the stiffness, the relation is simply

$$\underline{\underline{k}}(\sigma,T) = \underline{\underline{k}}_h(\sigma) - T\underline{\underline{k}}_s(\sigma) = \underline{\underline{k}}_0(\sigma) - (T-T_0)\,\underline{\underline{k}}_s(\sigma), \quad (6)$$

where $\underline{\underline{k}}_h$ and $\underline{\underline{k}}_s$ are the enthalpic and entropic contributions to the stiffness, and $\underline{\underline{k}}_0(\sigma)\equiv\underline{\underline{k}}(\sigma,T_0)$ is the stiffness at room temperature.

Although the covariance matrix for the fluctuations (Eq. 2) takes the simple form $\underline{\underline{C}} = k_BT(\underline{\underline{k}}_h - T\underline{\underline{k}}_s)^{-1}$, the equilibrium conformation $\underline{q}_0(\sigma,T)$ has a more complex behavior, and in particular does not vary linearly with temperature:

$$\underline{q}_0(\sigma,T) = (\underline{\underline{k}}_h - T\underline{\underline{k}}_s)^{-1}(\underline{\underline{k}}_h\underline{q}_0^h - T\underline{\underline{k}}_s\underline{q}_0^s).$$

To facilitate the estimation of the temperature-dependent contribution in the following numerical study, we consider the linear expansion of the latter expression around T_0:

$$\underline{q}_0(\sigma,T) \simeq \underline{q}_0^0(\sigma) - (T-T_0)\,\underline{q}_0'(\sigma), \quad (7)$$

where $\underline{q}_0^0(\sigma)\equiv\underline{q}_0(\sigma,T_0)$ is the equilibrium conformation at room temperature, and $\underline{q}_0'(\underline{q}_0^h,\underline{q}_0^s,\underline{\underline{k}}_h,\underline{\underline{k}}_s)$ is the first-order coefficient.

The presence of two parameters in both Eqs. 6 and 7 signifies that the temperature-dependent elastic model involves twice as many parameters as the model at a single temperature, $(\underline{\underline{k}}_0,\underline{q}_0^0)$. In the following sections, we estimate these new parameters $(\underline{\underline{k}}_s,\underline{q}_0')$ from MD simulations of DNA oligomers. For different temperatures and sequences, $\underline{q}_0(\sigma,T)$ is estimated from the mean value of the conformational distribution, and $\underline{\underline{k}}(\sigma,T)$ is estimated by inverting the covariance matrix (Eq. 2). Because Eqs. 6 and 7 have a linear temperature dependence, the parameters can then be computed by linear regression.

Simulations of DNA oligomers

The protocol for the MD simulations was chosen to be as close as possible to that used by the ABC consortium (16,17). The oligomers used in the simulation were 18-mers built from tetranucleotide repeats: an oligomer termed xyzw has the sequence GCzw xyzw xyzw xyzw GC (uppercase letters are basepairs conserved in all oligomers). Thus, the oligomer AAAC has the sequence GCACAAACAAACAAACGC. To eliminate possible end effects, we excluded the four terminal basepairs at either end of the oligomers from our analysis.

We simulated four 18-mers of dsDNA (AAAC, AGAT, GCGC, GGGG) at five different temperatures (273 K, 283 K, 300 K, 325 K, and 350 K) for 50 ns each. This data set contains all unique dinucleotide sequences and includes the influence of different flanking sequences for a limited sample (see Table S1 in the Supporting Material). The two types of basepair (AT and GC) appear in nine different trinucleotides and are used for studying intra parameters. The 10 unique dinucleotides appear within single tetranucleotides, except for AA, which appears in two contexts (AAAC and CAAA, both in the oligomer AAAC). Our data set is therefore less comprehensive than the ABC study at room temperature but more varied than the previous generation of MD simulations (22).

Each oligomer was constructed in the B-DNA conformation and simulated in Amber (24) for 50 ns in 150 mM KCl. The parameters and protocol of the simulations can be found in Lavery et al. (17). In particular, the DNA force field includes the parmbsc0 backbone parameters (15). We assumed that these force fields remain valid for double-helical DNA in the temperature interval under consideration. Water was modeled with the TIP4P/Ew model (25), which was designed to reproduce the structural and dynamical properties of water in a broad temperature range (273–350 K), in contrast to the SPC/E model used in most ABC runs. The results obtained with both solvents were compared at room temperature and exhibited no significant difference (17).

The temperature and pressure, $P = 1$ atm, during the simulations were controlled with the Berendsen algorithm (26). The simulations were stable at all temperatures with the same time step, $\tau = 2$ fs. Note that 350 K is beyond the experimental melting temperature of short oligomers; melting does not occur in the simulations, due to the limited sampling time and/or to limitations in the force fields.

Analysis of the trajectory

DNA conformations were analyzed with the program Curves+ (27), which uses commonly accepted conventions for the definitions of the helical parameters (18,28). From

TABLE 1 Sequence-averaged standard deviations of the different degrees of freedom

Buckle	Propel	Opening	Shear	Stretch	Stagger
11.62°	9.37°	4.53°	0.30 Å	0.12 Å	0.43 Å
Tilt	Roll	Twist	Shift	Slide	Rise
4.46°	7.10°	6.76°	0.71 Å	0.71 Å	0.35 Å

All values given are at 300 K, $\sigma_i = \sqrt{\langle (q_i - q_i^0)^2 \rangle_{T_0}}$. These reference values were used to express all covariance/stiffness elements in dimensionless units.

the distributions of the basepair and basepair-step parameters, we checked the validity of the harmonic approximation of the free energy (Eq. 1) at all sampled temperatures (see Figs. S2 and S3, and the detailed discussion in the Supporting Material). For all sequences, we then computed the mean values and covariance matrices, which were inverted to compute the stiffness matrices (Eq. 2). The entropic and enthalpic parameters were estimated by linear regression of 1), each element of the 6 × 6 stiffness matrix (Eq. 6) and 2), each degree of freedom of the equilibrium values (Eq. 7). We included a systematic estimation of statistical errors for the computed quantities by the block-averaging method (29). We used the weighted-least-squares fitting algorithm (30), which provides error estimates on the fitted model: the latter is therefore depicted by a shaded area corresponding to one standard-error deviation rather than a single line (see, for instance, Fig. 1 *B*). To discriminate the cases where the effect of temperature can be reliably estimated, we introduced a criterion based on the f-test (30), a statistical test that compares the accuracy of the two-parameter enthalpic + entropic model with that of a one-parameter enthalpic model. The different steps of this analysis procedure, as well as its validation on artificially generated data, are described in more detail in section II of the Supporting Material.

Effects of superhelical stress in biological DNA

We augment a thermodynamic model that efficiently predicts the local opening properties of superhelical DNA (19) by adding the temperature and basepair step dependencies of torsional energetics, the computation of which is included in this article. Here, we briefly describe the model, and we refer the reader to a study by Jost and co-workers (19) for more details on the formalism.

Local opening of DNA basepairs can be described by the Benham model, a thermodynamic model of DNA under superhelical stress that couples the standard thermodynamic description of basepairing with the torsional stress energetics (31). In the limit of long sequences, this model, defined in a superhelical density-imposed ensemble, is equivalent to a similar model in the torque-imposed ensemble where the constant applied torque can be computed self-consistently for a given superhelical density (19,32). In this Ising-like model, the energy of a DNA configuration constrained by a torque, Γ, is given by

$$H = H_{ZB} - \sum_i \left[\frac{\Gamma^2}{2C}\theta_i + \left(\frac{\Gamma^2}{2K_i} + \Gamma w_i \right)(1 - \theta_i) \right], \quad (8)$$

where H_{ZB} is the Zimm-Bragg Hamiltonian describing the denaturation of unconstrained DNA (13), $\theta_i = 1$ if the basepair step i is open and 0 if it is closed, and $C = 3.1 k_B T$ is the torsional stiffness of unpaired basepair steps. The sequence-dependent torsional stiffness, K_i, of double-stranded steps and the natural helical twist, w_i, given in Tables S9 and S10, are temperature-dependent and have been computed from Tables S5–S8. Computations of equilibrium properties for a given sequence (like the local opening probabilities $\langle \theta_i \rangle$) are then performed using the transfer-matrix method (19).

Coarse-graining the nanoscale elasticity to the persistence length

Starting from the computed nanoscale stiffness parameters, we neglect the influence of the intrabasepair deformations on the large-scale elasticity and consider only the step fluctuations. In an ideal B-DNA helix, only the two bending angles (tilt, τ, and roll, ρ) contribute to the large-scale bending of the molecule: $l_p = 2b/[\langle (\tau - \tau_0)^2 \rangle + \langle (\rho - \rho_0)^2 \rangle]$, where $b \simeq 0.34$ nm is the average rise. This relation is, however, not true for the real molecule, where all six step degrees of freedom contribute to this bending because of local distortions. We use a coarse-grain calculation (20) that takes these effects into account to compute the persistence length from the nanoscale parameters. The simpler tilt-roll relation is used to compute error bars. Note that because the deviations from the ideal B-DNA helix remain small, the discrepancy between the two methods is typically <5% (20).

From the sequence-dependent parameters obtained from the simulations, the sequence-neutral persistence length is computed by averaging the mean rise and the angle fluctuations (covariance matrix) over the 16 possible dinucleotides. The computed values are in the same order of magnitude as the experimental values, but significantly lower (~40 nm instead of ~50 nm, or 120 instead of 150 bp). Similar deviations have already been observed by Becker and Everaers (20), and it is unclear whether this effect is a consequence of the MD-estimated microscopic fluctuations or the calculation method. We assumed that this issue does not influence the computed temperature dependence and therefore rescaled them by their value at 278 K, following Theodorakopoulos and Peyrard (9).

Hybrid model

We construct a hybrid model by including the temperature dependence of the double-helical stiffness, computed in the previous paragraph, in a recently proposed model (9)

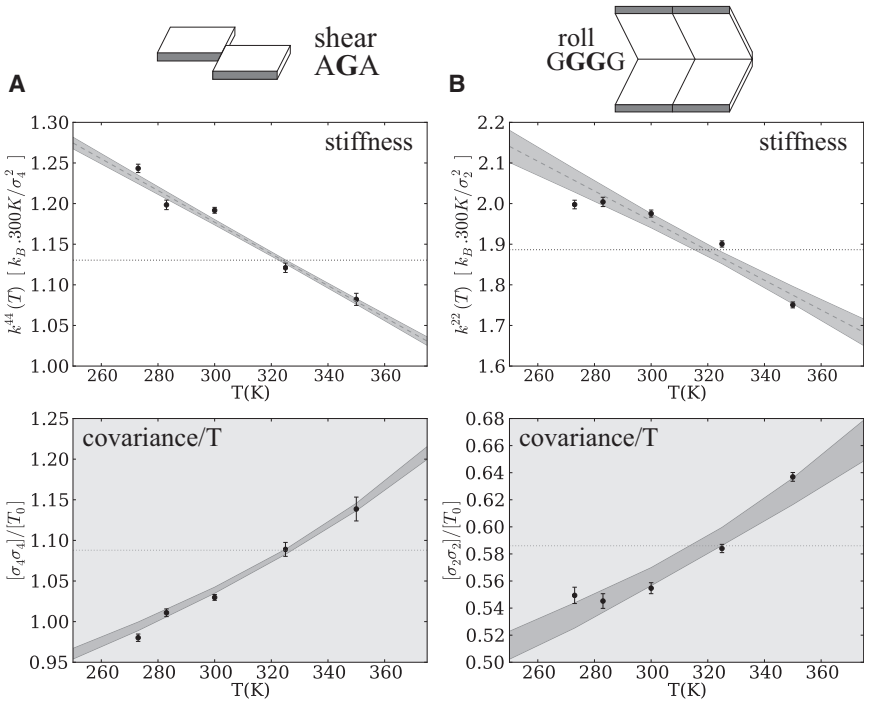

FIGURE 1 Linear fitting of the stiffness (*upper*) and corresponding temperature dependence of the variance (*lower*) for two typical diagonal matrix elements. (*A*) Intrabasepair parameter shear (sequence AGA). (*B*) Step parameter roll (sequence GGGG). In both cases, the data points rule out a purely enthalpic model (dotted line). The temperature dependence of the distribution is compatible with a linear entropic contribution, within the statistical errors of the simulations (*shaded area*). This contribution is stronger within the basepair (*A*) than for the step parameters (*B*). Note that we plot the covariance divided by the temperature, which is constant for a purely enthalpic phenomenon. All quantities are expressed in reduced units, defined using the corresponding standard deviation σ_i (see Table 1).

that accounts for the reduction of the effective persistence length of DNA by transient denaturation bubbles. We describe here the main lines of the calculation; more detail can be found in Theodorakopoulos and Peyrard (9).

We model DNA polymer elasticity in terms of a heterogeneous Kratky-Porod (KP) (33) chain of N segments of length a with a configurational energy,

$$H = -\sum_{j=1}^{N} J_{j,j+1}\vec{R}_j \cdot \vec{R}_{j+1}, \tag{9}$$

$N+1$ basepairs (considered as point monomers for simplicity), and contour length $L = Na$. $\vec{R}_j$ is a unit vector joining the jth to the $(j+1)$ th basepair, and the local stiffness constants,

$$J_{j,j+1} = \big(1-\theta_{j+1}\big)J + \theta_{j+1}J',$$

are weighted averages of hard (J) and soft (J') couplings, according to the probability θ_{j+1} that the $(j+1)$ th basepair is in the unbound (open) state. The local melting fractions, $\{\theta_j\}$, are computed in terms of the Peyrard-Bishop-Dauxois model (12,34), using only the sequence information and the molar ion concentration, c, with no further adjustable parameters (35,36).

The average end-to-end distance can be numerically computed for the heterogeneous KP model (9), because the correlation functions factorize. It is then possible to use it to extract an effective persistence length, λ, from the relationship $<r^2> = 2\lambda L - 2\lambda^2(1-e^{-\lambda/L})$, which is valid in the continuum limit of the homogeneous KP chain (known as the wormlike chain), and should be quantitatively adequate as long as $\lambda \gg a$.

The above scheme was used in the Theodorakopoulos and Peyrard model (9), with $c = 0.004, J = 6\times 10^{-12}$erg, and $J' = 0.14 \times 10^{-12}$erg, to compute the effective persistence length of the phage fragment studied in Geggier et al. (37). In this work, we incorporated the temperature dependence of the local stiffness constants arising from harmonic fluctuations, i.e.,

$$J \rightarrow J\frac{k_b(T)}{k_b(278K)}, \tag{10}$$

where $k_b(T)$ is the temperature-dependent double-helical bending stiffness, as estimated from the MD simulations.

RESULTS

We simulated four 18-mers of DNA of different sequences at temperatures of 273 K, 283 K, 300 K, 325 K, and 350 K at fixed pressure. The atomistic trajectories were analyzed with the conformational analysis software Curves+ (27), which provided distributions of intra (internal basepair) and step (basepair step) parameters. Our data set includes all unique mono- and dinucleotide sequences. However, basepairs and basepair steps can be influenced by the flanking sequences and in some cases (17) can lead to bimodal parameter distributions. We observe such effects (see Fig. S3) and deal with them by separately analyzing each trinucleotide or tetranucleotide sequence fragment (see Table S1). In this case, for the oligomers we studied,

the distributions can indeed be approximated as single Gaussians, compatible with the harmonic approximation of the free energy. The covariance matrices are inverted to compute stiffness matrices, from which we estimate the entropic contribution according to Eq. 6. We include a systematic estimation of the statistical errors for the matrix elements, which are fitted independently of each other (see Models and Methods). Fig. 1 shows typical examples of such fits for diagonal matrix elements of intra (Fig. 1 *A*) and step (Fig. 1 *B*) parameters, where the entropic contribution is apparent.

Note that the different degrees of freedom cannot be analyzed separately. The passage between the stiffness and the covariance elements involves a matrix inversion, which couples the 6×6 degrees of freedom and their respective levels of statistical noise; an example of the complete fitting procedure is shown in Fig. S5. The good agreement between the fitted model and the data points in both representations is therefore nontrivial and validates the approximations used in our analysis. We now discuss separately the results obtained for the intra and step parameters.

Internal basepair flexibility

In this section, we describe the effect of temperature on internal basepair flexibility. As explained previously, we estimate the entropic contribution separately for each sequence fragment (here trinucleotides) where the harmonic approximation is valid. Fig. 2 shows the temperature evolution of the stretch and opening stiffnesses, averaged over the A-T (Fig. 2 *A*) and C-G (Fig. 2 *B*) basepairs. Note that we use reduced units, dividing all coordinates by their average standard deviations, σ_i, at $T_0 = 300$ K (Table 1).

We were not surprised to find that the values depend strongly on the type of basepair, reflecting the different number of hydrogen bonds (three for C-G and two for A-T). The effect of temperature emerges in all cases as a regular linear decrease of the stiffness. We measure the strength of this entropic contribution by comparing it to the enthalpic part (Eq. 6). For stretch (Fig. 2, *upper*), this ratio is ~0.5 at room temperature, i.e., the stiffness has already dropped to half its value at 0 K. For opening (Fig. 2, *lower*), the effect is weaker (~0.3), and still more so for the other degrees of freedom.

Interestingly, the influence of base sequence is slightly different for these two parameters. For stretch, although the stiffness constant for the C-G basepair at $T_0 = 300$ K is ~50% larger than that for A-T, the entropic stiffness (slope) is of the same order: the relative entropic contribution is therefore stronger for A-T. For opening, on the other hand, the stiffness of the C-G basepair at room temperature is three times stronger than that of A-T, but here, the entropic contribution is also approximately three times higher, so that its relative weight is similar.

At this level, the error bars still partly reflect the structural heterogeneity due to the flanking sequences. We therefore look at the results obtained for the different trinucleotides, including the remaining degrees of freedom. These results are summarized in Fig. 3 *A*, which gives the relative entropic contribution (as described in the previous paragraph) plotted against the stiffness at room temperature. In this representation, the behavior noted previously for opening is reflected in the fact that the two groups of points are distant horizontally (different stiffness at T_0) but have approximately the same values on the vertical axis.

We systematically determine which parameters present a detectable entropic contribution by applying a quantitative criterion, comparing the accuracy of the enthalpic + entropic model with that of a purely enthalpic one (see Models and Methods). An entropic contribution is detected for all sequences with respect to the opening angle and the three translational parameters stretch, shear, and stagger.

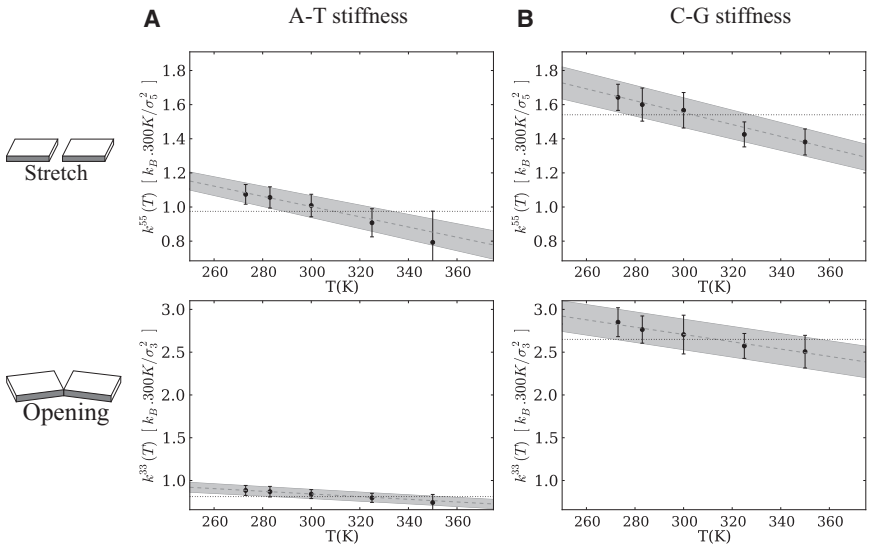

FIGURE 2 Temperature dependence of the stiffness for the stretch-stretch (*upper*) and opening-opening (*lower*) diagonal terms on the A-T (*A*) and C-G (*B*) basepairs. The reduced units are defined using the average standard deviation of each parameter, σ_i, given in Table 1. Here, the error bars partly reflect the structural heterogeneity due to the flanking sequences. The effect of temperature is strongest in these two degrees of freedom: at room temperature, the average stiffness has decreased by ~45% and 30%, respectively, with respect to the values at 0 K (enthalpic stiffness).

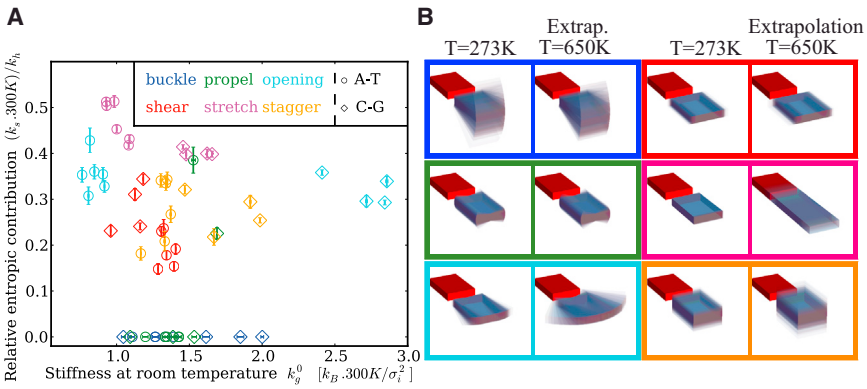

FIGURE 3 Entropic contribution to internal basepair elasticity: (*A*) Relative entropic contribution to the stiffness, compared to the enthalpic stiffness, for different sequences. Note that because the units are parameter-specific, the values of different parameters should not be compared. Although the stiffness of the A-T and C-G basepairs is very different in the opening direction, the relative entropic contribution is similar. (*B*) Illustration of the sequence-averaged range of fluctuations in the different degrees of freedom for $T = 273$ K and extrapolated to $T = 650$ K, i.e., close to the destabilization of the basepair (see text) to emphasize the temperature effect. Note that the colored rectangles in *B* correspond to the colors of the basepair parameters in *A*. In the directions most sensitive to temperature, stretch, and opening, the fluctuations become considerable.

The error bars are small enough to discriminate the individual (trinucleotide) sequences, confirming the influence of flanking sequences on basepairs. The contributions are strongest for stretch and opening, as noted previously; for shear and stagger, the relative entropic contribution is ~0.25 at room temperature. In contrast, the two other angular parameters exhibit very little or no detectable entropic contribution (for details, see Table S2).

We then look at the equilibrium values, i.e., Eq. 7. The parameters for which temperature induces an average displacement are generally also those for which the stiffness changes; the average stretch increases, as does the average opening, but mainly for G-C. The other parameters do not exhibit a systematic effect.

Basepair step flexibility: parameterization of a T-dependent rigid-basepair model

We now focus on the step parameters. Our oligomers contain the 10 unique dinucleotides, each located within a single tetranucleotide except for AA, which is present in two sequence contexts (see Table S1).

The stiffness values for the step parameters are less uniform than those for the intra parameters, as is the effect of temperature. For most sequences, an entropic contribution is found for the two bending angles, tilt and roll, and for the translational-parameter rise. The relative weight of this contribution, defined in the previous section as $(k_s \times 300\ \mathrm{K})/k_h$, spans the intervals 0.25–0.5, 0.15–0.35, and 0.25–0.4, respectively, depending on the sequence. As an example, Fig. 1 *B* shows the roll-roll diagonal elements of the GG dinucleotide: the entropic contribution clearly emerges from the statistical errors but is indeed smaller than for shear. More surprisingly, the twist and shift stiffnesses exhibit a very low sensitivity to temperature.

The temperature dependence of the equilibrium step parameters is a question of considerable interest: a modification of the spontaneous basepair stacking would be an evolutionary challenge to organisms living at high temperatures. Such a modification is not observed in the data. The only parameter where a clear and sequence-independent tendency emerges out of statistical noise is rise, which typically increases by 2–3% in the temperature interval considered. The equilibrium twist angle decreases by 1–2° for about half of the sequences (e.g., TA, GC, and GA) and remains approximately constant for the other ones, except in the case of CG, where it increases, yielding an average decrease of ~0.5°. The relation between the living temperature of organisms and supercoiling is discussed in the next section.

These results show that the entropic effect on basepair step elasticity is detectable, but more limited than that on basepair elasticity. Note that this conclusion may also be affected by slower equilibration times, which would require a greater computational effort to resolve. Our analysis provides, to our knowledge, the first parameterization of a temperature-dependent rigid-basepair model of DNA where the entropic contributions are sequence-dependent. Altogether, the model includes 156 nonzero entropic parameters in addition to the 270 parameters required for describing the elasticity at a single temperature. It is also the first model, to our knowledge, that systematically includes confidence intervals for these parameters, which can be used to estimate error bars for the quantities computed at the coarse-grain level, as we show in the next paragraph. The model parameters are given in Tables S5 and S7. Note that for the AA dinucleotide, treated in two different sequence contexts, we arbitrarily chose the values computed for the CAAA data set, where the mean values are closer to those reported in crystallographic structures (21), and where we had more data points.

DISCUSSION

Temperature dependence of soft vibration modes

Our results demonstrate that the basepair internal stiffness contains an important entropic contribution. This contribution is strongest for the stretch and opening degrees of

freedom (Fig. 3 *B*). Although A-T and G-C basepairs differ by their enthalpic stiffness constants, they show similar ratios, k_s/k_h (see also below). In an experimental setting, the corresponding temperature sensitivity of soft vibration modes might be accessible via Raman spectroscopy or neutron scattering.

Few studies have dealt with the temperature dependence of these spectra. Among those that have, the only one (to our knowledge) that focuses on the frequency shifts in the double helix is that of Grimm and Rupprecht (38). The neutron-scattering spectra in that study exhibit a mode at ~1.1 THz (4.5 meV) at 193 K, which softens to ~0.85 THz at 300 K, i.e., a relative softening of ~20–25%, or, in our terms, a relative entropic weight at room temperature of ~0.45, consistent with our results. These numbers must be treated with some caution, however, since the data only allow a qualitative estimation of this softening. Another problem is the indexing of the associated mode with respect to the base and basepair degrees of freedom. From the distributions in our simulations, we estimate that the internal basepair vibrations have frequencies between 0.6 THz (opening) and 2.4 THz (stretch); the latter value has been reported in Raman spectra (39). The step frequencies are generally lower, in the 0.3–0.6 THz range. This indexation thus requires further experimental support, but if the observed mode corresponds indeed to an intrabasepair eigenmode, it is in qualitative agreement with the entropic contribution observed in our simulations.

The path to the melting transition

Our harmonic model is the first-order approximation of the DNA double-helical free energy. For obvious reasons, these fluctuations have to be bounded for the model to be valid: the eigenvalues of the stiffness matrix have to be positive. Within our description of the temperature-dependent stiffness, $k(T) = k_h - Tk_s$, the double helix is thus predicted to become unstable at $T_s = k_h/k_s$, known in the thermodynamics literature as the spinodal temperature (40). Depending on the sequence, this temperature is estimated to be between 550 and 660 K (Table S4). As a comparison, we note that the spinodal temperature of liquid water at atmospheric pressure, $T_s \simeq 600$ K (41), is comparable to the value we estimate for the double helix.

These large values signify that our biologically relevant range of temperature (273–350 K) remains far below the limit of (meta)stability of the double helix, thus justifying the approximation of constant k_h and k_s values: the melting transition occurs before these values change significantly. Note that the double helix might still exist as a metastable state for temperatures $T_m<T<T_s$.

The eigenvector associated with the decomposition of the basepair at T_s reflects the directions for which the fluctuations are most sensitive to temperature. Fig. 3 *B* shows that the basepair is destabilized in the stretch direction, with a simultaneous partial opening. Qualitatively, these directions may also indicate the direction of the kinetic pathway toward melting.

With our MD simulations, sampling only double-helical states, we rely in the following on standard theories of DNA melting (11,13) to provide information on the relative stability of single-stranded DNA and the thermodynamics of bubble formation.

DNA in vivo: torque control of the local opening probability

How do our results apply to DNA in living cells, in particular for extremophiles living at temperatures close to or above 85°C, the typical melting temperature of unconstrained DNA (42)? In these organisms, many proteins bind to double-helical DNA (43), and the opening of a significant fraction of the molecule would probably be lethal. It has been suggested that the double helix could be stabilized by over-twisting (14). Maintaining DNA under superhelical stress by specific enzymes could thus be one of the strategies used by extremophiles to keep their DNA closed even beyond 85°C. In vivo, the double helix is therefore not free, as in our simulations, but in a constrained torsional state. Consequently, one may wonder if this constraint can affect the elastic and structural properties of the DNA double helix and therefore modify local physicochemical features like the binding constants of transcription enzymatic complexes.

To account for this effect, we use a recently proposed method that allows the efficient computation of DNA melting properties under superhelical stress imposed by an applied torque (19). Augmenting the Benham model to incorporate the temperature and basepair step dependencies of twisting energetics studied in this article (see Models and Methods), we evaluate the suitable torque needed to stabilize the double helix at high temperatures and combine this estimation with our previous results to estimate the elastic properties of the biologically relevant constrained helix.

As a function of the temperature, we compute the torque needed to maintain an open fraction of 1% for the genomic sequence of the extremophilic bacteria *Thermus thermophilus*. This fraction corresponds to the typical opening probability of *Escherichia coli* under standard physiological temperature and superhelical density $\Delta\sigma \simeq -0.06$ (19). This propensity to open is mainly located at gene promoters and transcription start sites (19), enhancing gene expression by polymerase enzymes (44). The constraining torque increases with temperature (see Fig. S6), and as noticed in previous studies by Benham (14), the model predicts that the stability of the double helix can be maintained only up to a critical temperature, which depends on the sequence (data not shown). In Fig. 4, we plot the different contributions of the average basepair step stability. At high temperatures, thermal destabilization of pairing and stacking (*gray dashed*

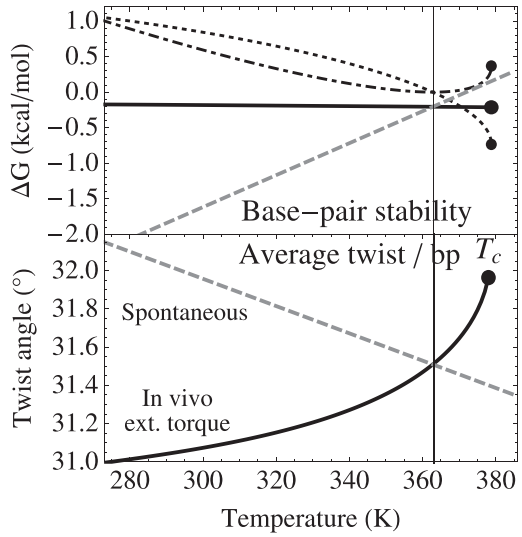

FIGURE 4 Stabilization of the DNA double helix in vivo by superhelical stress for temperatures between 273 K and 380 K, where extremophile organisms exist. Although unconstrained DNA (*gray dashed line*) is unstable at high temperatures, the contribution of the average twist in the free energy (*dotted line*) maintains the stability of the double helix (*black solid line*) for an additional 15 K, with an almost constant opening penalty of ~0.2 kcal/mol. Above a critical temperature $T_c \approx 378$ K (*dots*), the contribution from the twist stiffness of the double helix (*dash-dotted line*) prevents further stabilization. The superhelical stress opposes the spontaneous decrease of the average twist, and results in an ~1°/bp increase in the considered interval.

line) is prevented by the contribution of the average twist (*dotted line*) in the free energy of the constrained helix, maintaining an almost constant basepair-step opening penalty of 0.2 kcal/mol (*black solid line*). Above a certain temperature, the stabilization is limited by the strong twisting stiffness of double-stranded steps (*dash-dotted line*), and the double helix starts to melt. For *T. Thermophilus*, we find a critical temperature of 106°C, which is 15°C higher than the melting temperature in the absence of superhelical constraints. Interestingly, this value is close to the highest living temperatures of extremophilic organisms, suggesting that the stability of the double helix may indeed be the limiting condition for their existence.

At this limit, the stabilizing torque is surprisingly low ($\sim 2\, k_B T$). Although the resulting twist excess is important in 1% of open basepairs (~33°/bp, close to the average twist of B-DNA; see Fig. S7), the action of the torque results only in a weak local twist excess for double-stranded regions, ranging from 0.8°/bp for stiff basepair steps to 2.5°/bp for soft ones. These values should be compared to the typical twist standard deviation at room temperature (~6°/bp). For comparison, adding 1° to the spontaneous twist per basepair increases the superhelical density, $\Delta\sigma$, by ~0.03. At the rigid-basepair level of description, the properties of the constrained helix can therefore be described in the regime of linear response, where the equilibrium conformation is displaced but the elasticity remains identical to that of the relaxed helix. The stiffness parameters extracted from our simulations, with their limited temperature dependence, are therefore relevant to biological DNA. Using these parameters, we compute the mean conformation at the critical temperature by relaxing the sequence-neutral helix under the suitable torque. The resulting twist modification is shown in Fig. 4: the external torque goes against the small spontaneous twist decrease and results in an average increase of ~1° in the interval under consideration, between 0°C and the critical temperature.

Note that the computed values correspond to the base content of a specific organism and neglect the influence of other physicochemical parameters (ionic strength), as well as alternate structural transitions that may contribute to releasing the superhelical stress, in particular the B-Z transition (45). These contributions may therefore shift the computed critical temperature in an organism-dependent way (for instance, we estimate that $T_c = 90$°C for *E. coli*). However, since they are (at first order) purely enthalpic, they should not modify the basic mechanism.

Altogether, our results show (14) that under constraints, double-stranded DNA can exist above the typical DNA melting temperature. Our simulations make it possible to compute its elastic and structural properties, which remain close to those of unconstrained DNA at room temperature. These observations explain how similar molecular mechanisms of DNA-protein associations can exist for organisms living in the whole temperature range (43). Locally, the melting profiles may also be driven out of equilibrium by in vivo active processes, in particular transcription (46), which may induce a transient superhelical addition (or removal) in the same range of values computed here (47).

Large-scale flexibility of unconstrained DNA

As a final point, we address the (apparent) large-scale bending flexibility of DNA. At the wormlike-chain level, the details of the molecule are averaged out and the flexibility is described by the bending persistence length, $l_p(T)$, i.e., the bending correlation length of the polymer, which is related to the bending stiffness, $k_b(T)$, by the relation

$$l_p(T) = b\frac{k_b(T)}{k_B T}, \tag{11}$$

where b is the length of a rigid element of the polymer. Within the rigid-basepair model, these parameters can be derived from the nanoscale structure and elasticity via a systematic coarse-graining procedure (20). We note that the three step parameters most influential in terms of large-scale bending are also the ones for which the effect of temperature is strongest. Not surprisingly, our estimate for the bending stiffness, $k_b(T)$, of the DNA double helix exhibits a weak, but noticeable, temperature dependence (Fig. 5, *gray line*; shaded area indicates the estimated error bar).

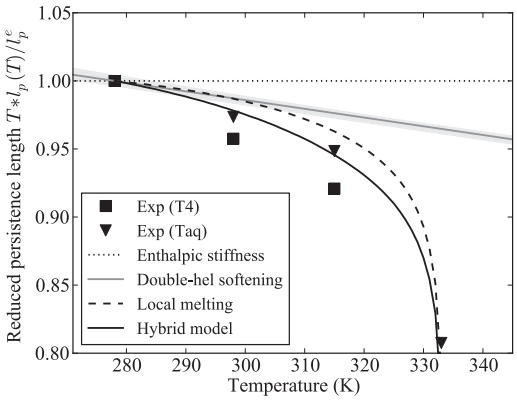

FIGURE 5 Persistence length of sequence-averaged DNA, multiplied by temperature and rescaled by its value at 278 K. In this representation, a purely enthalpic stiffness yields a temperature-independent value (*dotted line*). The markers ▼ and ■ are the experimental data points from Geggier et al. (37) obtained with two types of ligase enzymes. Gray solid line: MD-derived values of the persistence length, with the estimated error bars (*shaded area*), i.e., the contribution from the double-helical elasticity; dashed line: contribution predicted from the denaturation bubbles (9); black solid line: the hybrid model, including both contributions.

Recent measurements (37) have indeed ruled out a temperature-independent stiffness, k_b (Fig. 5, *dotted horizontal line*). However, the data included in Fig. 5 show a temperature effect, which is considerably stronger than expected from our results for the DNA double helix. In addition to softening the double helix, the flexibility of DNA may also increase due to the presence of transient denaturation bubbles that generate local kinks (7,8). In a different study, two of the authors of the article presented here quantitatively implemented this idea and reproduced the dramatic decrease of the apparent stiffness in the premelting stage (Fig. 5, *last datapoint*) (9). However, at lower temperatures, the bubble formation is insufficient to reproduce the experimental slope if the stiffness of the double helix is assumed to be independent of T (Fig. 5, *dashed line*).

Here, we present the results from a hybrid model, which accounts for the temperature dependence of the double-helical stiffness, as well as for the effect of bubbles, and which is in good agreement with the experimental data (Fig. 5, *solid line*). At physiological temperatures, the two effects contribute equally strongly. Closer to the melting transition, the unharmonic effects dominate.

CONCLUSION

The nanoscale mechanical structure and elasticity of the DNA double helix can be inferred from the analysis of crystallographic data and MD simulations. To date, there is almost no information on how these properties vary with temperature. In this article, we present results from MD simulations of DNA oligomers at different temperatures. We show that entropy plays a significant role in double-helical elasticity, both inside the basepair and, to a lesser extent, between successive basepairs.

At the internal-basepair level, this entropic contribution is particularly strong in the stretch and opening degrees of freedom. With increasing temperature, it results in a significant broadening of the harmonic fluctuations in these directions of the basepair plane, which can be considered the initial step of the DNA melting pathway.

At the basepair step level, the entropic contribution is weaker but detectable. It is strongest in the roll, tilt, and rise degrees of freedom relevant to the large-scale bending rigidity. We include the resulting temperature dependence of the persistence length of the DNA double helix in a description of kinking in transient denaturation bubbles. Our predictions for the effective large-scale bending stiffness agree with the experimentally measured values for standard physiological conditions up to temperatures close to the melting temperature of DNA.

It is remarkable that living organisms thrive over an even broader temperature range. To describe DNA in vivo, we have assumed that biological organisms control the stability of the double helix by regulating superhelical stress. We have estimated the required values from the Benham model and included the effect in our description of the nanoscale properties. Overall, we find a remarkably small temperature dependence of the structure and flexibility of genomic DNA, compatible with conservation of protein-DNA binding mechanisms (43).

SUPPORTING MATERIAL

Ten tables, nine figures, references (48 and 49) and Supporting Methods are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(13)01022-9.

S.M. thanks K. Zakrzewska for her help with the simulation software. We acknowledge the Pole Scientifique de Modélisation Numérique for computing resources.

This work was supported by the Agence Nationale de la Recherche grants FSCF (ANR-12-BSV5-0009-01) (to R.E.) and Chrome (ANR-12-BSV5-0017-01) (to R.L.).

REFERENCES

1. Davey, C. A., D. F. Sargent, …, T. J. Richmond. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* 319:1097–1113.
2. Dixon, N. E., and A. Kornberg. 1984. Protein HU in the enzymatic replication of the chromosomal origin of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*. 81:424–428.
3. Kim, Y., J. H. Geiger, …, P. B. Sigler. 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature*. 365:512–520.
4. Tansey, M. R., and T. D. Brock. 1972. The upper temperature limit for eukaryotic organisms. *Proc. Natl. Acad. Sci. USA*. 69:2426–2428.
5. Takai, K., K. Nakamura, …, K. Horikoshi. 2008. Cell proliferation at 122°C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation. *Proc. Natl. Acad. Sci. USA*. 105:10949–10954.

6. Rubinstein, M., and R. Colby. 2003. Polymer Physics (Chemistry). Oxford University Press, Cary, NC.
7. Yan, J., and J. F. Marko. 2004. Localized single-stranded bubble mechanism for cyclization of short double helix DNA. *Phys. Rev. Lett.* 93:108108.
8. Wiggins, P. A., R. Phillips, and P. C. Nelson. 2005. Exact theory of kinkable elastic polymers. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 71:021909.
9. Theodorakopoulos, N., and M. Peyrard. 2012. Base pair openings and temperature dependence of DNA flexibility. *Phys. Rev. Lett.* 108:078104.
10. Zimm, B. H., and J. Bragg. 1959. Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* 31:526–535.
11. Poland, D., and H. A. Scheraga. 1966. Phase transitions in one dimension and the helix-coil transition in polyamino acids. *J. Chem. Phys.* 45:1456–1463.
12. Dauxois, T., M. Peyrard, and A. R. Bishop. 1993. Entropy-driven DNA denaturation. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* 47:R44–R47.
13. Jost, D., and R. Everaers. 2009. Genome wide application of DNA melting analysis. *J. Phys. Condens. Matter.* 21:034108.
14. Benham, C. J. 1996. Theoretical analysis of the helix-coil transition in positively superhelical DNA at high temperatures. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* 53:2984–2987.
15. Pérez, A., I. Marchán, …, M. Orozco. 2007. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.* 92:3817–3829.
16. Beveridge, D. L., G. Barreiro, …, M. A. Young. 2004. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys. J.* 87:3799–3813.
17. Lavery, R., K. Zakrzewska, …, J. Sponer. 2010. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.* 38:299–313.
18. Dickerson, R. E. 1989. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.* 17:1797–1803.
19. Jost, D., A. Zubair, and R. Everaers. 2011. Bubble statistics and positioning in superhelically stressed DNA. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 84:031912.
20. Becker, N. B., and R. Everaers. 2007. From rigid base pairs to semiflexible polymers: coarse-graining DNA. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 76:021923.
21. Olson, W. K., A. A. Gorin, …, V. B. Zhurkin. 1998. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA.* 95:11163–11168.
22. Lankas, F., J. Sponer, …, T. E. Cheatham, 3rd. 2003. DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* 85:2872–2883.
23. Becker, N. B., L. Wolff, and R. Everaers. 2006. Indirect readout: detection of optimized subsequences and calculation of relative binding affinities using different DNA elastic potentials. *Nucleic Acids Res.* 34:5638–5649.
24. Pearlman, D., D. Case, …, P. Kollman. 1995. Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 91:1–41.
25. Horn, H. W., W. C. Swope, …, T. Head-Gordon. 2004. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* 120:9665–9678.
26. Berendsen, H., J. Postma, …, J. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
27. Lavery, R., M. Moakher, …, K. Zakrzewska. 2009. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* 37:5917–5929.
28. Olson, W. K., M. Bansal, …, H. M. Berman. 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* 313:229–237.
29. Flyvbjerg, H., and H. Petersen. 1989. Error estimates on averages of correlated data. *J. Chem. Phys.* 91:461–466.
30. Press, W., S. Teukolsky, …, B. Flannery. 2007. Numerical Recipes: The Art of Scientific Computing, 3rd ed. Cambridge University Press, New York.
31. Fye, R. M., and C. J. Benham. 1999. Exact method for numerically analyzing a model of local denaturation in superhelically stressed DNA. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* 59:3408–3426.
32. Jost, D. 2013. Twist-DNA: computing base-pair and bubble opening probabilities in genomic superhelical DNA. *Bioinformatics.* In press.
33. Kratky, O., and G. Porod. 1949. Röntgenuntersuchung gelöster fadenmoleküle. *Recl. Trav. Chim. Pays-B.* 68:1106–1122.
34. Peyrard, M., and A. R. Bishop. 1989. Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.* 62:2755–2758.
35. Theodorakopoulos, N. 2010. Melting of genomic DNA: predictive modeling by nonlinear lattice dynamics. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 82:021905.
36. Theodorakopoulos, N. 2011. Bubbles, clusters and denaturation in genomic DNA: modeling, parametrization, efficient computation. *J. Nonlin. Math. Phys.* 18:429–447.
37. Geggier, S., A. Kotlyar, and A. Vologodskii. 2011. Temperature dependence of DNA persistence length. *Nucleic Acids Res.* 39:1419–1426.
38. Grimm, H., and A. Rupprecht. 1997. Low frequency dynamics of DNA. *Phys. B Cond. Mat.* 234–236:183–187.
39. Urabe, H., and Y. Tominaga. 1981. Low frequency raman spectra of DNA. *J. Phys. Soc. Jpn.* 50:3543–3544.
40. Chandler, D. 1987. Introduction to Modern Statistical Mechanics. Oxford University Press, Cary, NC.
41. Eberhart, J., and V. Pinks, II. 1985. The thermodynamic limit of superheat of water. *J. Colloid Interface Sci.* 107:574–575.
42. Frank-Kamenetskii, F. 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.
43. Grayling, R. A., K. Sandman, and J. N. Reeve. 1996. DNA stability and DNA binding proteins. *Adv. Protein Chem.* 48:437–467.
44. Schneider, R., A. Travers, and G. Muskhelishvili. 2000. The expression of the *Escherichia coli fis* gene is strongly dependent on the superhelical density of DNA. *Mol. Microbiol.* 38:167–175.
45. Zhabinskaya, D., and C. J. Benham. 2012. Theoretical analysis of competing conformational transitions in superhelical DNA. *PLOS Comput. Biol.* 8:e1002484.
46. Liu, L. F., and J. C. Wang. 1987. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA.* 84:7024–7027.
47. Tsao, Y.-P., H.-Y. Wu, and L. F. Liu. 1989. Transcription-driven supercoiling of DNA: direct biochemical evidence from in vitro studies. *Cell.* 56:111–118.
48. Jones, E., T. Oliphant, and P. Peterson. 2001. Scipy: Open source scientific tools for python. http://www.scipy.org/.
49. Hunter, J. 2007. Matplotlib: a 2d graphics environment. *Comput. Sci. Eng.* 9:90–95.

Temperature dependence of the DNA double helix at the nanoscale: structure, elasticity and fluctuations

Sam Meyer, Daniel Jost, Nikos Theodorakopoulos, Michel Peyrard, Richard Lavery and Ralf Everaers

Supporting Material

I. Dataset

Trinuc	Occ
A**A**A	2
A**A**C	2
C**A**A	3
G**A**T	2
T**A**G	3
T**A**T	2
A**C**A	2
C**C**C	9
G**C**G	9
T**C**T	2

Tetranuc	Occ
A**AA**C	2
C**AA**A	3
A**AC**A	2
T**AG**A	3
G**AT**A	2
A**CA**A	2
G**CG**C	5
A**GA**T	2
C**GC**G	4
G**GG**G	9
A**TA**G	2

Table S1: Number of occurrences of trinucleotide (intra parameters) and tetranucleotide (step parameter) sequences. We analyzed these sequences separately, rather than the central mono- or dinucleotides where the Gaussian approximation is not valid (see text and Fig. S3).

II. Methods

Analysis of the trajectory DNA conformations were analyzed by the program Curves+ (1), which computes a full set of helical, backbone and groove parameters. Curves+ uses the commonly agreed "Tsukuba" reference frame for the description of each base (2), and the Cambridge convention for the names and signs of all helical parameters (3). We focused on the internal base-pair and the base-pair step deformations. Starting from the time series of these parameters as provided by Curves+ (and the associated trajectory analysis program Canal), subsequent analysis (Boltzmann inversion, covariance and matrix inversions, error estimates, Monte Carlo generation of "artificial data", linear regressions, plots) was developed in Python, with the use of the Numpy/Scipy libraries (4) and the MatPlotLib library for plotting (5).

We checked for the absence of any melting or end-fraying of the oligomers, using the intra parameters obtained at the different positions and temperatures (Fig. S2), especially the highest (325K and 350K). Only two cases showed aberrant values, which were attributed to partial melting and eliminated. We also checked that the distributions are monomodal and stable at all temperatures (see Fig. S1), and we compute the *covariance matrix*. By inverting the latter, we get the *stiffness matrix* (Eq. 2), which will be used to estimate the enthalpic and entropic contributions according to Eq. 6.

A comparison between the distributions obtained with different occurrences of given sequence fragments shows that the statistical power is sufficient to resolve the relatively small temperature-induced variations, except in some cases at the lowest temperatures where kinetics are slower. To quantify this effect, we systematically estimate statistical confidence intervals for the computed quantities (see next section). Because the runs at the lower temperatures have larger error bars, their relative weight in the subsequent modeling is proportionnally reduced: to compensate this effect, we conducted more simulations for this range of temperatures.

Linear fitting of the simulation data To estimate the errors on the computed stiffness, covariance and equilibrium values, we used the *block averaging method* (7). This is a generalization of the simple idea that the statistical errors can be estimated by comparing the values obtained by splitting a trajectory into shorter parts ("blocks"), and computing the quantity of interest separately for each block.

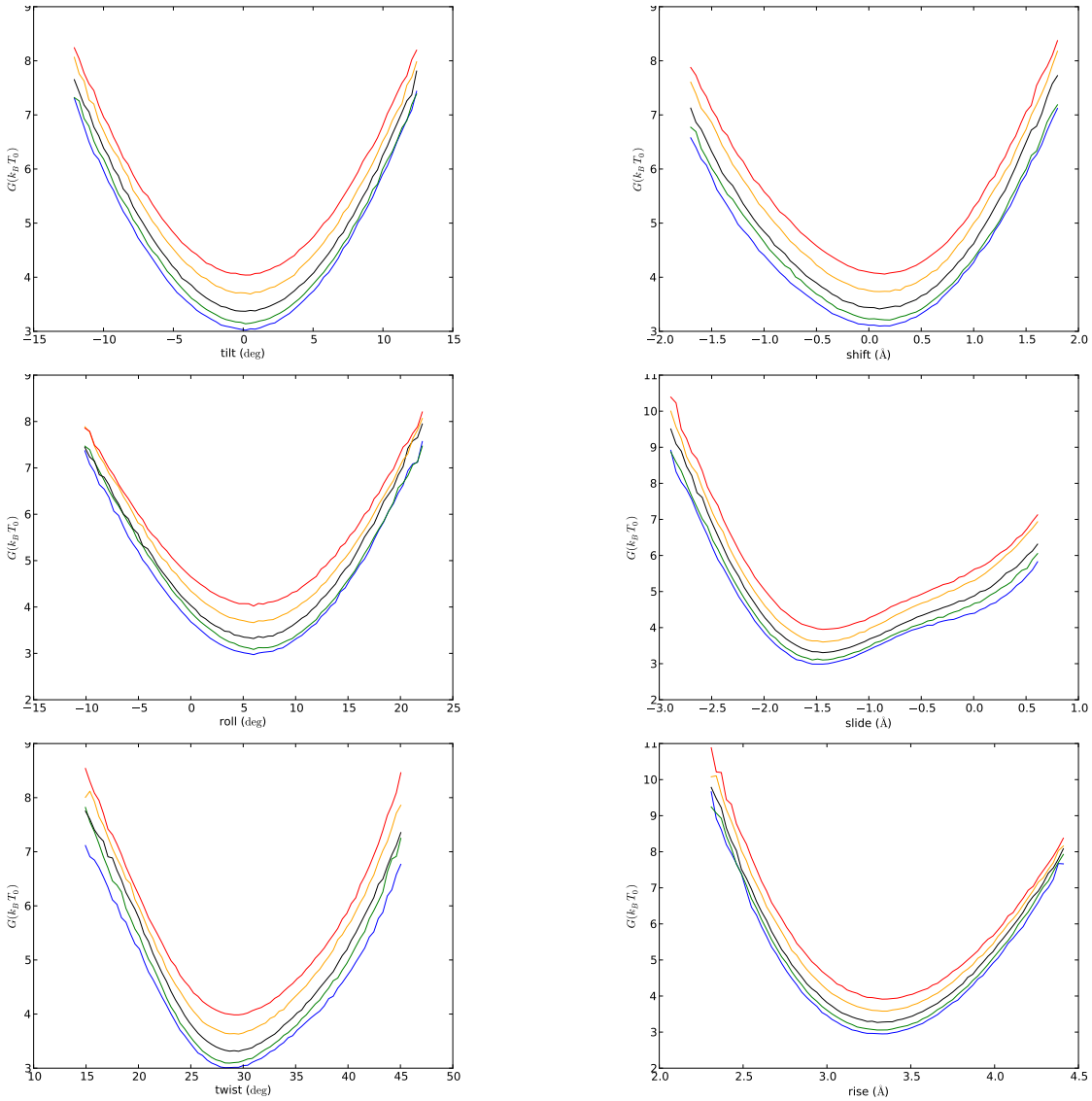

Figure S1: Free energy functions, obtained by Boltzmann inversion of the MD distributions projected along the 6 step degrees of freedom (left column: angular, right column: translational), for the GG sequence. Simulated temperatures: 273K (blue), 283K (green), 300K (black), 325K (orange), 350K (red). The distributions have the same shape at all temperatures (even 350 K), indicating the absence of melting in the analyzed base-pairs. In particular, they are all monomodal: we consider the harmonic approximation in the subsequent analysis. Bimodal distributions were observed in some very specific cases, already noticed in (6).

We separate the enthalpic and entropic contributions by finding the best linear fit for the temperature dependence of (i) each of the stiffness matrix elements (Eq. 6), and (ii) each degree of freedom of the equilibrium values (Eq. 7). The weighted least square fitting procedure from (8) gives a reduced weight to data points with poor precision, and provides not only the best fit parameters (the enthalpic and entropic stiffness contributions), but also error estimates for these parameters and correlations between them. The values and errors of the stiffness matrix at any temperature can be calculated from these numbers, as depicted in the plots by a shaded area rather than a single line: see for instance Fig. 1B.

We introduce a quantitative criterion, to discriminate between the cases where the effect of temperature on the data can be reliably estimated, from those where it is too small to be quantified. The "f-test" (8) compares the accuracy of a model with an enthalpic and entropic contribution (two parameters) and that of a purely enthalpic model, and determines if the gain in precision justifies the more detailed model. Our criterion is a threshold on a number obtained from a combination of this test with the relative uncertainty on the temperature t_s and Pearson's correlation coefficient between the data points. The two latter contributions were added to eliminate some specific cases where the f-test was positive, while the results of the fit lacked precision.

Validation of the analysis method The covariance and stiffness matrices are related by an inversion operation (Eq. 2), which mixes the errors of the different stiffness matrix elements. We made the simplistic hypothesis that the latter are

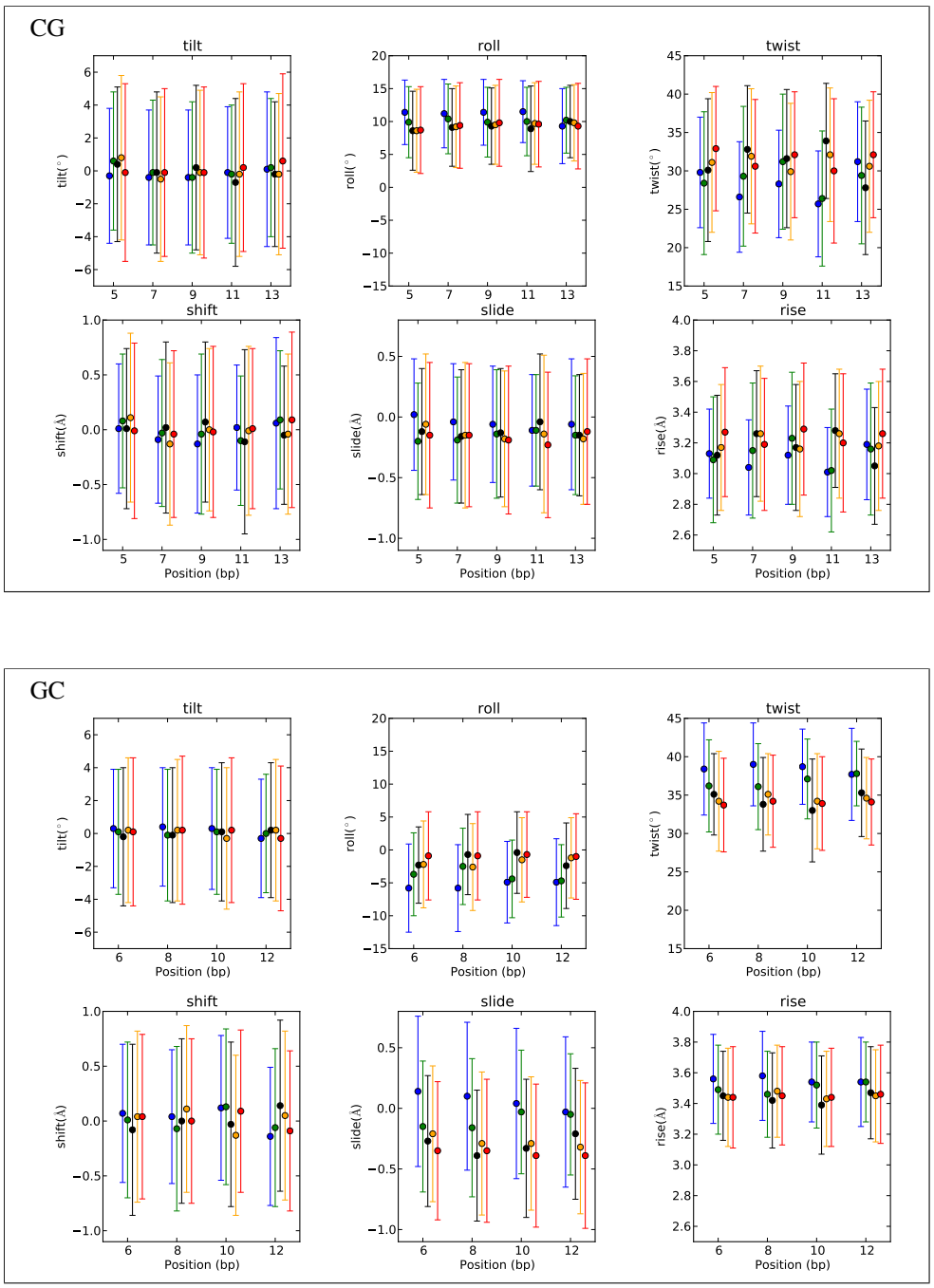

Figure S2: Mean values (dots) and standard deviations (bars) of the base-pair step parameters computed along the oligomer CGCG, at the simulated temperatures: from 273 K (blue) to 350 K (red) (same color code as Fig. S1). The outer 4 bps on either side of the oligomer have been excluded of the analysis to eliminate possible end effects. **(Upper panel)** Odd bps (CG steps); **(lower panel)** even bps (GC steps). The overall regularity of the values shows the absence of long-acting fraying or melting, even at $T = 350K$. Comparison of the values obtained at the different positions for the same dinucleotide indicates that for the lowest temperatures, the remaining statistical noise is sometimes important: see for instance the mean values of twist, for the GC steps (lower panel).

independent. From the fitted stiffness model, the computation of a model-derived covariance matrix (including the error estimates) then involves (i) generating an ensemble of stiffness matrices representative of the estimated uncertainties by Monte Carlo ; (ii) inverting each of these matrices and (iii) computing the average and standard deviation of the resulting ensemble of covariance matrices. The hypothesis of independence of the stiffness errors implies that the computed covariance errors are an upper bound. A comparison with the covariance error bars computed directly on the data shows that they are indeed larger, but are of same order of magnitude, and can therefore be considered acceptable given the possible equilibration problems (see for instance Fig. 1, lower panel).

To validate the entire procedure, we apply it to "artificial" trajectories, which were generated so as to mimic the properties of the real data, but where we know by construction the "true" parameters. The application of the procedure indeed allows the "input" properties to be recovered within the estimated error bars.

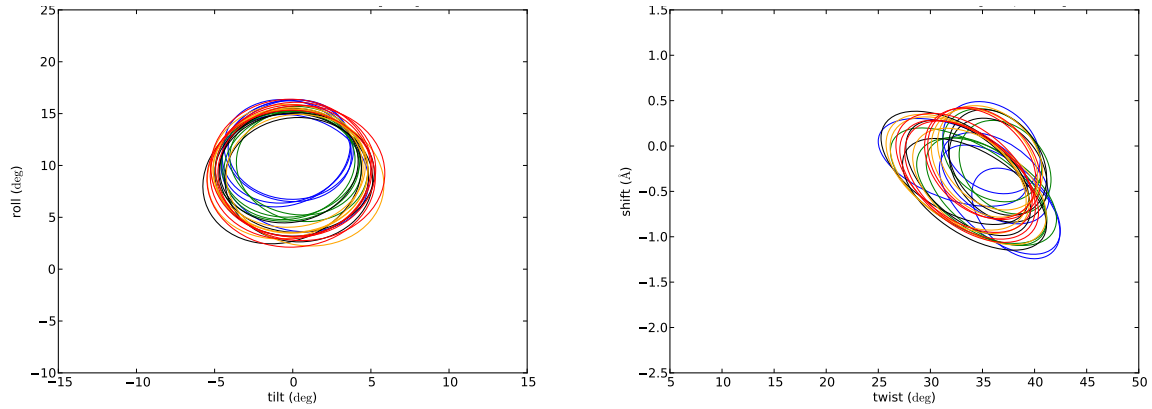

Figure S3: Covariance ellipses computed on the individual positions for the different temperatures (from blue-cold to red-hot, same colors as previous figures). **(Left)** Rise-slide coupling for the CG dinucleotides. Despite a remaining level of statistical noise, one can follow the evolution of the covariance with temperature : the Gaussian approximation is valid. **(Right)** Twist-shift coupling for AA dinucleotides located in the A**AA**C and C**AA**A tetranucleotides. They clearly divide into 2 groups (corresponding to the two different contexts), especially for larger temperatures where the system is better equilibrated, confirming the influence of the neighbor bps on the dinucleotide conformations.

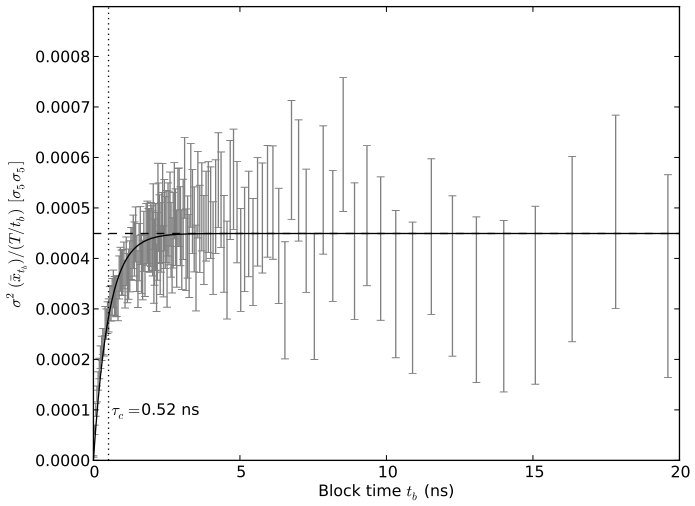

Figure S4: Computation of the statistical errors by the "block averaging" procedure, here for the rise variance of GC at 325K. For each "block size" t_b, we split the trajectory into parts, and the "block error" is then given by standard deviation of the resulting values, divided by the square root of the number of blocks $\sqrt{T/t_b}$. When the block size is small compared to the correlation time of the considered quantity (left part), the block values are not independent, and the computed error is therefore underestimated. For longer sizes, the samples become uncorrelated and the computed error becomes independent of the block size: the value of this plateau provides a reliable error estimate. In each case, we compute the confidence interval (vertical bar) of the computed error, and we fit the curve with an exponential function, yielding the correlation time τ_c. The details of the method are given in (7). In most cases, the correlation time is of the order of 1 ns; in some cases however, the data exhibited very long correlation times, which we attributed to large-scale rearrangements of the molecule: the computed error was then multiplied by a factor 2 for security.

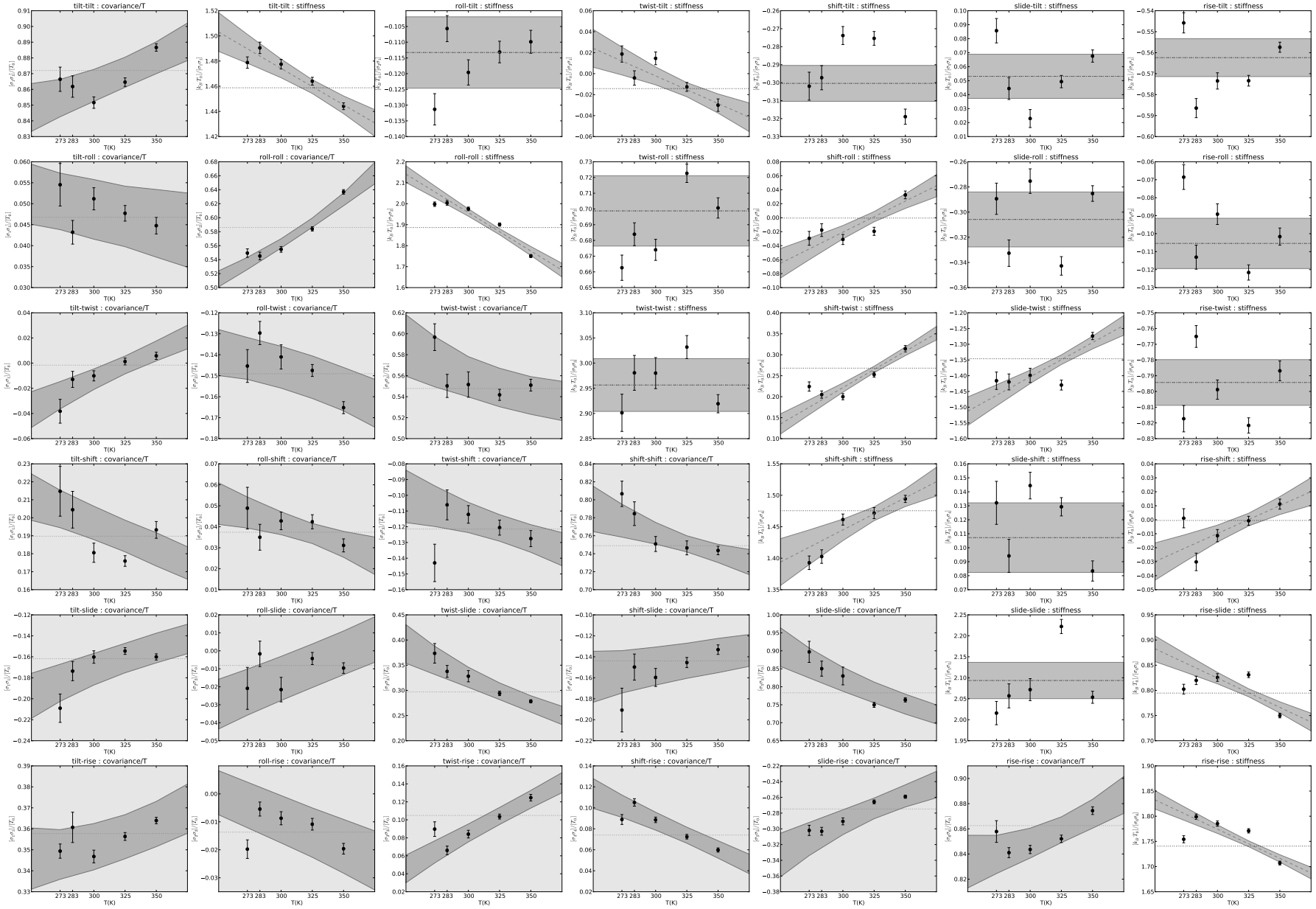

Figure S5: Synthetic illustration of the complete fitting procedure of the GG step parameters. All stiffness matrix elements (white background, upper right half) are fitted independently of each other. The covariance elements (grey background) are obtained by inverting the stiffness matrix. In this operation, the different degrees of freedom get mixed: a given covariance element may therefore exhibit a temperature dependence even though its stiffness counterpart (transposed) does not.

II. Internal bp elasticity

Entropic contribution to the stiffness and equilibrium values: Results for the different parameters and sequences

Seq	buckle	propel	opening	shear	stretch	stagger
AAA	∞	∞	$\mathbf{2.82 \pm 0.12}$	$\mathbf{5.21 \pm 0.34}$	$\mathbf{2.32 \pm 0.04}$	$\mathbf{3.74 \pm 0.25}$
AAC	∞	$\mathbf{2.60 \pm 0.19}$	$\mathbf{3.05 \pm 0.12}$	$\mathbf{6.50 \pm 0.46}$	$\mathbf{2.39 \pm 0.04}$	$\mathbf{2.99 \pm 0.14}$
CAA	∞	∞	$\mathbf{2.78 \pm 0.12}$	$\mathbf{5.62 \pm 0.31}$	$\mathbf{2.21 \pm 0.04}$	$\mathbf{4.80 \pm 0.48}$
GAT	∞	∞	$\mathbf{2.34 \pm 0.15}$	$\mathbf{4.23 \pm 0.33}$	$\mathbf{1.95 \pm 0.05}$	$\mathbf{2.92 \pm 0.14}$
TAG	∞	∞	$\mathbf{2.83 \pm 0.13}$	$\mathbf{6.76 \pm 0.55}$	$\mathbf{1.95 \pm 0.04}$	$\mathbf{5.50 \pm 0.44}$
TAT	∞	∞	$\mathbf{3.25 \pm 0.20}$	$\mathbf{4.34 \pm 0.21}$	$\mathbf{1.98 \pm 0.04}$	$\mathbf{2.93 \pm 0.13}$
ACA	∞	∞	$\mathbf{2.95 \pm 0.09}$	$\mathbf{3.22 \pm 0.13}$	$\mathbf{2.51 \pm 0.06}$	$\mathbf{4.59 \pm 0.38}$
CCC	∞	$\mathbf{4.44 \pm 0.25}$	$\mathbf{3.41 \pm 0.06}$	$\mathbf{4.15 \pm 0.13}$	$\mathbf{2.51 \pm 0.04}$	$\mathbf{3.94 \pm 0.09}$
GCG	∞	∞	$\mathbf{2.90 \pm 0.06}$	$\mathbf{3.91 \pm 0.14}$	$\mathbf{2.22 \pm 0.03}$	$\mathbf{3.58 \pm 0.16}$
TCT	∞	∞	$\mathbf{3.38 \pm 0.14}$	$\mathbf{2.91 \pm 0.09}$	$\mathbf{2.50 \pm 0.06}$	$\mathbf{3.40 \pm 0.16}$

Table S2: Spinodal temperature computed separately for the diagonal elements of the intra-bp stiffness matrix, in units of 300K: $t_s = k_h/(300K.k_s)$. This value is the inverse of the relative entropic weight shown in Fig. 3: an infinite spinodal temperatures signifies the absence of an entropic contribution (buckle, propeller), and a strong contributions yields a value close to 1 (stretch, opening).

Seq	buckle (°)	propel (°)	opening (°)	shear (Å)	stretch (Å)	stagger (Å)
AAA	8.21	-6.25	-1.52	0.0	-0.05	0.21
	8.79	-17.45	2.91	0.12	0.01	0.00
AAC	0.0	-5.95	0.0	0.0	-0.06	0.29
	5.65	-16.38	2.21	0.14	0.02	0.20
CAA	5.45	0.0	0.0	-0.06	-0.06	0.12
	7.34	-11.02	3.92	0.06	0.02	-0.16
GAT	0.0	0.0	1.54	-0.12	-0.08	0.37
	-3.40	-13.80	3.13	0.09	0.03	0.13
TAG	3.81	-3.66	0.0	-0.03	-0.06	0.27
	6.39	-12.84	5.12	0.05	0.03	0.17
TAT	0.0	0.0	-2.48	0.0	-0.07	0.18
	2.81	-11.46	3.52	-0.07	0.04	0.11
ACA	0.0	0.0	-1.12	0.0	-0.06	0.21
	4.41	-9.97	1.87	-0.03	0.05	0.08
CCC	1.62	1.90	-1.27	0.03	-0.03	0.0
	-1.83	-4.28	0.49	0.01	0.01	-0.25
GCG	5.73	0.0	-1.81	0.0	-0.05	0.18
	2.27	-9.27	0.22	0.03	0.02	0.20
TCT	0.0	-7.20	-0.84	0.0	-0.04	0.07
	2.71	-13.45	1.19	0.04	0.02	0.05

Table S3: Temperature dependence of the equilibrium values, for the intra-bp parameters. For each sequence, the first line indicates the entropic contribution at room temperature ($T_0.q_0'(s)$ in the notation of Eq. 7 in the main text), and the second line gives the value at room temperature ($q_0^0(s)$). From these two numbers, the equilibrium value at any temperature T is given by: $q_0(s,T) = q_0^0(s) - (T/T_0 - 1)(T_0.q_0'(s))$.

Spinodal decomposition of the base-pairs: temperature and direction

	t_s [300K]	buckle	propel	opening	shear	stretch	stagger
AAA	2.034	-0.287	0.337	-0.56	0.12	-0.606	-0.33
AAC	2.179	-0.019	0.149	-0.05	-0.014	-0.856	-0.492
CAA	2.102	0.067	-0.051	0.456	-0.057	0.873	0.139
GAT	1.743	0.109	0.156	-0.427	0.04	0.799	0.376
TAG	1.848	0.129	-0.072	-0.065	-0.032	-0.955	-0.246
TAT	1.802	0.075	0.094	-0.479	-0.072	-0.84	-0.213
ACA	1.753	0.096	0.082	-0.565	-0.36	-0.707	-0.188
CCC	1.877	-0.04	-0.14	0.554	-0.23	0.782	0.083
GCG	2.034	0.002	0.213	-0.521	0.039	-0.792	-0.235
TCT	1.802	-0.024	0.141	-0.569	0.361	-0.711	-0.143

Table S4: Spinodal decomposition of the base-pairs: temperature and direction of instability. The components of the eigenvector are expressed in reduced units of the different parameters, where they have comparable values at T_0. It therefore represents their respective weight in the direction of instability. The element of strongest weight is stretch, then opening.

III. Step elasticity

All following values are also given in separate ASCII files, with the associated error estimates

Equilibrium values

Param	CG	CA	TA	AG	GG	AA	GA	AT	AC	GC
Twist (°)	-7.92	0.	4.74	4.30	0.	0.	7.10	0.	2.27	9.943
Tilt (°)	0.	0.	0.	0.	0.	0.	0.	0.93	0.	0.
Roll (°)	2.86	0.	-9.60	-2.20	1.10	-3.95	-6.18	-1.59	-2.77	-11.96
Shift (Å)	-0.076	-0.192	0.565	-0.436	0.	-0.363	0.483	-0.400	0.	0.
Slide (Å)	0.189	0.	0.	0.339	-0.222	0.367	0.169	-0.199	0.	0.956
Rise (Å)	-0.49	0.	-0.348	-0.128	-0.163	-0.308	-0.161	-0.264	-0.102	0.

Table S5: Temperature dependence of the equilibrium conformation ($q'_0(s).T_0$ in the notation of Eq. 7). From these values and those of Table S6, the equilibrium value at any temperature T is given by: $q_0(s,T) = q^0_0(s) - (T/T_0 - 1)(T_0.q'_0(s))$.

Param	CG	CA	TA	AG	GG	AA	GA	AT	AC	GC
Twist (°)	**30.30**	**27.76**	**31.21**	**32.37**	**29.99**	**33.54**	**36.2**	**29.71**	**32.67**	**35.56**
	36.1	37.3	37.8	31.9	32.9	35.1	36.3	29.3	31.5	33.6
	27.93	25.69	29.74	28.84	29.93	31.04	33.32	29.02	31.46	33.13
Tilt (°)	**-0.02**	**1.80**	**0.11**	**-1.66**	**0.18**	**-2.21**	**-0.73**	**0.30**	**0.02**	**0.03**
	0.	0.5	0.	-1.7	-0.1	-1.4	-1.5	0.	-0.1	0.
	0.	-0.01	0.	0.99	0.84	-1.52	-0.28	0.	-0.36	0.
Roll (°)	**9.74**	**11.80**	**8.70**	**4.31**	**6.02**	**1.15**	**1.35**	**-0.75**	**-1.13**	**-2.86**
	5.4	4.7	3.3	4.5	3.6	0.7	1.9	1.1	0.7	0.3
	8.75	8.36	10.33	2.75	5.28	2.31	2.27	0.21	0.63	1.23
Shift (Å)	**-0.01**	**-0.15**	**0.20**	**-0.47**	**0.05**	**-0.41**	**-0.01**	**0.00**	**0.32**	**0.01**
	0.	0.09	0.	0.09	0.05	-0.03	-0.28	0.	0.13	0.
	0.	0.18	0.	0.05	0.07	-0.14	-0.01	0.	-0.12	0.
Slide (Å)	**-0.13**	**-0.40**	**-0.28**	**-0.54**	**-1.16**	**-0.17**	**-0.66**	**-0.80**	**-0.78**	**-0.21**
	0.41	0.53	0.05	-0.25	-0.22	-0.08	0.09	-0.59	-0.58	-0.38
	-0.76	-0.57	-1.02	-1.33	-1.64	-1.04	-0.92	-1.08	-1.16	-1.09
Rise (Å)	**3.16**	**3.30**	**3.17**	**3.38**	**3.36**	**3.27**	**3.42**	**3.20**	**3.39**	**3.45**
	3.39	3.33	3.42	3.34	3.42	3.27	3.37	3.31	3.36	3.4
	3.26	3.13	3.34	3.38	3.61	3.31	3.37	3.2	3.38	3.37

Table S6: Equilibrium conformations at room temperature, q^0_0 . We give the value estimated from our simulations (bold), from the analysis of DNA-protein crystallographic structures (9) (second line), and from a previous generation of MD simulations (10) (third line).

Stiffness values

Param	CG	CA	TA	AG	GG	AA	GA	AT	AC	GC
Twist-twist	0.0	0.0	0.027	0.0	0.0	-0.019	0.0	0.0	0.0	0.0
Twist-tilt	0.0	0.0	-0.003	0.0	0.003	0.0	-0.007	0.0	0.0	0.002
Twist-roll	0.005	0.0	0.0	0.0	-0.002	0.0	0.0	0.0	0.0	-0.006
Twist-shift	0.0	-0.041	0.063	0.0	-0.054	0.033	0.0	0.0	0.0	0.0
Twist-slide	0.0	-0.109	-0.193	0.0	-0.078	0.147	-0.245	0.065	0.0	-0.0
Twist-rise	-0.173	-0.154	-0.207	0.0	0.0	0.0	0.066	0.0	0.0	-0.121
Tilt-tilt	0.015	0.0	0.0	0.011	0.005	0.015	0.019	0.007	0.011	0.025
Tilt-roll	0.0	0.0	0.0	0.0	0.0	-0.005	0.0	-0.002	0.0	-0.0
Tilt-shift	-0.023	-0.031	0.0	-0.037	0.0	-0.045	-0.103	0.0	0.0	-0.096
Tilt-slide	0.037	0.0	0.0	0.075	0.0	0.124	0.0	0.0	0.0	-0.041
Tilt-rise	0.0	0.0	0.0	-0.095	0.0	-0.099	-0.226	-0.104	0.0	-0.0
Roll-roll	0.020	0.004	0.0	0.0	0.012	0.007	0.005	0.011	0.015	0.011
Roll-shift	0.0	0.0	0.0	0.0	-0.028	0.0	0.071	0.0	0.0	-0.0
Roll-slide	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.121	0.163
Roll-rise	-0.141	0.0	0.0	0.104	0.0	0.114	0.0	0.0	0.178	0.250
Shift-shift	0.0	0.0	-1.249	0.0	-0.427	0.676	0.0	0.0	0.0	0.0
Shift-slide	0.0	0.273	-0.309	0.431	0.0	0.343	0.0	0.0	-0.623	0.0
Shift-rise	0.0	0.0	0.0	0.675	-0.268	0.847	0.0	0.0	0.0	0.293
Slide-slide	0.866	0.0	0.836	0.0	0.0	0.0	2.002	2.712	1.019	1.225
Slide-rise	0.0	2.177	1.518	0.0	0.0	-0.833	0.0	3.406	1.560	0.0
Rise-rise	5.722	2.613	2.264	3.655	1.537	5.867	5.326	6.833	6.007	3.45

Table S7: Entropic stiffness constants, k_s, in units of (kcal/mol)/(deg^2.[300K]), (kcal/mol)/(Å^2.[300K]), or (kcal/mol)/(deg.Å.[300K]). From the values k_0 at T_0=300K (Table S8) and these numerical values, the stiffness at a given temperature T is given by (Eq. 6): $k(T) = k_0 - (T/T_0 - 1)k_s$.

Param	CG	CA	TA	AG	GG	AA	GA	AT	AC	GC
Twist-twist	0.0184	0.0169	0.0248	0.0305	0.0386	0.0318	0.0433	0.0422	0.0483	0.034
	(0.0227)	(0.021)	(0.0357)	(0.0441)	(0.0482)	(0.0461)	(0.0422)	(0.0463)	(0.0489)	(0.0421)
Twist-tilt*	0.0001	0.0026	0.0001	0.0087	0.0	0.0097	0.0003	-0.001	0.0047	0.0002
Twist-roll	0.008	0.0089	0.0117	0.0094	0.0085	0.0115	0.0107	0.0126	0.0108	0.0062
Twist-shift*	-0.0003	-0.0097	0.0125	0.0922	0.0278	0.0805	0.012	-0.0166	0.0192	0.0029
Twist-slide	-0.0688	-0.0429	-0.0687	-0.0725	-0.174	-0.1107	-0.216	-0.101	-0.175	-0.139
Twist-rise	-0.255	-0.213	-0.212	-0.207	-0.205	-0.218	-0.174	-0.160	-0.189	-0.256
Tilt-tilt	0.0348	0.0284	0.0241	0.04	0.0442	0.0453	0.0416	0.0404	0.0433	0.0459
	(0.0278)	(0.0275)	(0.0245)	(0.0371)	(0.0414)	(0.0389)	(0.0392)	(0.0404)	(0.0411)	(0.0396)
Tilt-roll*	-0.0001	0.001	0.0001	0.0033	-0.0022	-0.0006	-0.0028	0.0003	0.0041	-0.0003
Tilt-shift	-0.107	-0.0451	-0.0097	-0.0367	-0.0547	-0.0269	-0.0623	0.0239	-0.0297	-0.1154
Tilt-slide*	0.0017	-0.003	-0.0156	0.0138	0.0101	0.0073	0.0082	-0.0147	-0.0436	-0.0046
Tilt-rise	0.0051	-0.0096	-0.0104	-0.166	-0.219	-0.236	-0.275	-0.0213	0.0575	0.0034
Roll-roll	0.0224	0.0209	0.0178	0.0215	0.0231	0.0231	0.0211	0.0262	0.0261	0.0248
	(0.0153)	(0.0184)	(0.0136)	(0.0227)	(0.0241)	(0.0235)	(0.0211)	(0.0272)	(0.0267)	(0.0275)
Roll-shift	-0.0019	0.0258	0.0077	0.0248	-0.0025	-0.0036	0.0287	0.0215	0.0338	0.0028
Roll-slide	-0.0034	-0.0124	0.0087	-0.0567	-0.036	-0.0788	-0.0262	-0.0369	0.03	0.0904
Roll-rise	-0.1039	-0.1066	-0.1016	-0.0447	-0.0251	-0.0066	-0.0292	0.0624	0.104	0.1651
Shift-shift	1.480	1.524	0.781	1.595	1.690	2.257	1.700	1.425	1.526	1.420
	(1.346)	(1.600)	(1.529)	(1.657)	(1.984)	(1.975)	(1.430)	(1.193)	(1.341)	(1.761)
Shift-slide*	0.0098	-0.0735	-0.076	0.252	0.1347	0.4312	0.3307	-0.269	-0.3643	-0.0199
Shift-rise*	0.0058	0.0585	0.145	0.3118	-0.022	0.653	0.572	-0.168	0.133	-0.0122
Slide-slide	2.362	1.716	1.457	1.951	2.469	2.373	2.592	4.346	3.559	3.184
	(2.034)	(2.286)	(2.269)	(2.706)	(3.215)	(2.914)	(2.518)	(3.310)	(2.974)	(2.708)
Slide-rise	1.095	1.314	0.860	1.329	1.928	0.960	1.405	2.857	2.539	2.258
Rise-rise	7.358	6.083	6.967	7.906	8.879	9.468	9.098	10.56	10.35	10.54
	(4.390)	(6.290)	(5.055)	(6.388)	(7.335)	(7.621)	(8.330)	(10.50)	(9.88)	(10.28)

Table S8: Stiffness k_0 at room temperature, $T_0 = 300K$, in units of (kcal/mol)/deg^2, (kcal/mol)/Å^2, or (kcal/mol)/(deg.Å) for angular, translational, or mixed deformations, respectively. For the diagonal elements of the stiffness matrix, we indicate in parenthesis the previous MD values from (10), obtained with different force fields, water model, and oligomer sequences. The non-diagonal parameters indicated with a star should be zero for self-complementary sequences (CG, GC, AT, TA) for symmetry reasons.

IV. Torque-control of DNA stability

Param	CG	CA	TA	AG	GG	AA	GA	AT	AC	GC
w_0 (°)	30.3	27.8	31.2	32.4	30.0	33.5	36.2	29.7	32.7	35.6
w_0' (°)	-7.9	0.0	4.7	4.3	0.0	0.0	7.1	0.0	2.3	9.9

Table S9: Natural helical twist w for the 10 different base-pair steps. At any temperature T, $w = w_0 - (T/T_0 - 1)w_0'$ ($T_0 = 300$K) and represents the equilibrium value of "Twist" given in Tables S5 and S6

Param	CG	CA	TA	AG	GG	AA	GA	AT	AC	GC
K_0	0.81	0.78	1.15	1.47	2.33	1.38	2.09	3.16	3.24	1.84
K_0'	-0.44	-0.72	0.77	-0.35	-0.56	-1.44	-2.30	0.73	0.05	-0.59

Table S10: Torsional stiffness K for the 10 different base-pair steps. At any temperature T, $K = K_0 - (T/T_0 - 1)K_0'$ ($T_0 = 300$K) and represents the inverse of the twist-twist element of the matrix $\underline{k}^{-1}$ with $\underline{k}$ the stiffness matrix given in Tables S7 and S8. In units of 10^{-2} kcal/mol/degree^2.

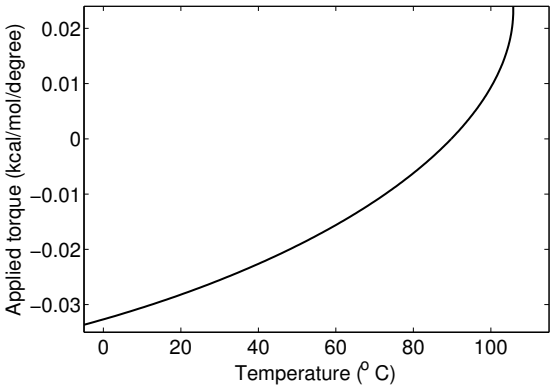

Figure S6: External torque applied on DNA *in vivo* for temperatures between 0°C and 110°C, maintaining a constant open fraction of 1% of the *T. thermophilus* genome (see text). Without external torque, the DNA melting occurs at $T_m \approx 91°$ C. A positive supercoiling maintains the stability of the double helix for an additional 15° C, up to a critical temperature $T_c \approx 106°$ C. The resulting twist excess is shown in Fig. 4 for the double-helical DNA, and on Fig. S7 for the small open fraction, which absorbs most of this excess.

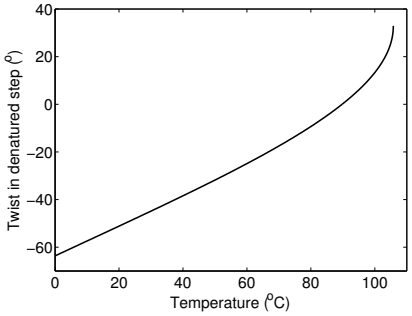

Figure S7: Average twist in the $\sim 1\%$ open DNA, as a result of the external torque (Fig. S6). Because of the much larger twist rigidity of double-helical DNA, the open fraction absorbs much more twist than the closed parts (compare with Fig. 4).

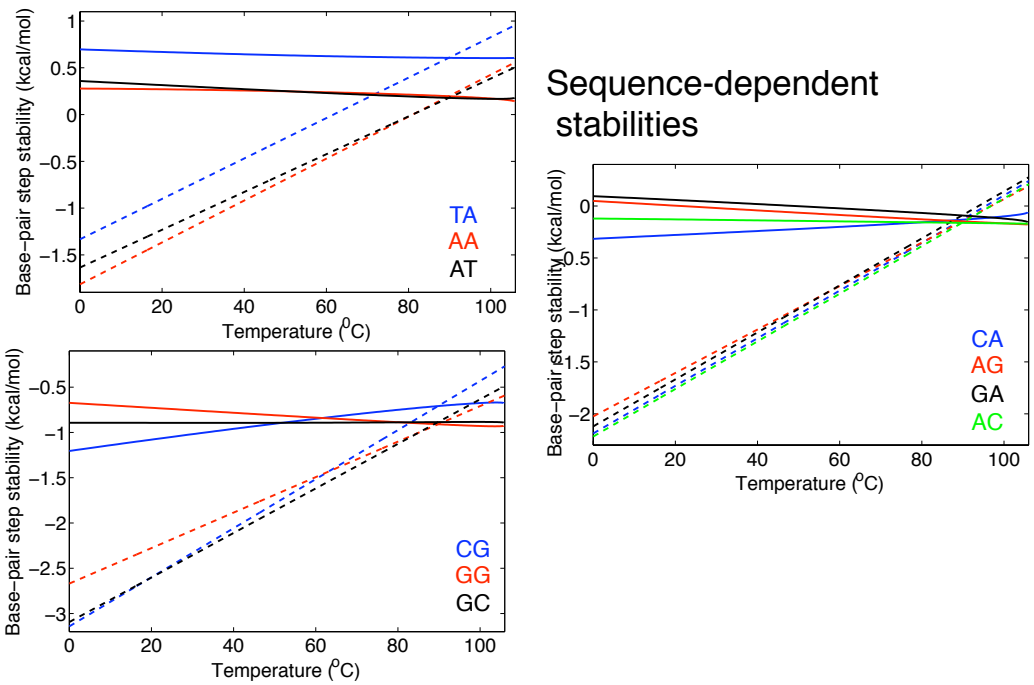

Figure S8: Double-helical stability of the individual step sequences, for unconstrained DNA (dashed) and under the biologically relevant torque (solid lines) as computed in Fig. S6. While the opening penalty becomes approximately constant at the sequence-neutral level, it is not rigorously true for the individual sequences. In particular, the order of stability is temperature-dependent, and is also different from that of unconstrained DNA (see for instance GC and GG). Note that the less stable sequences (upper left) are destabilized by the torque, but the stability of the double-helix is maintained by the cooperative terms between successive base-pairs.

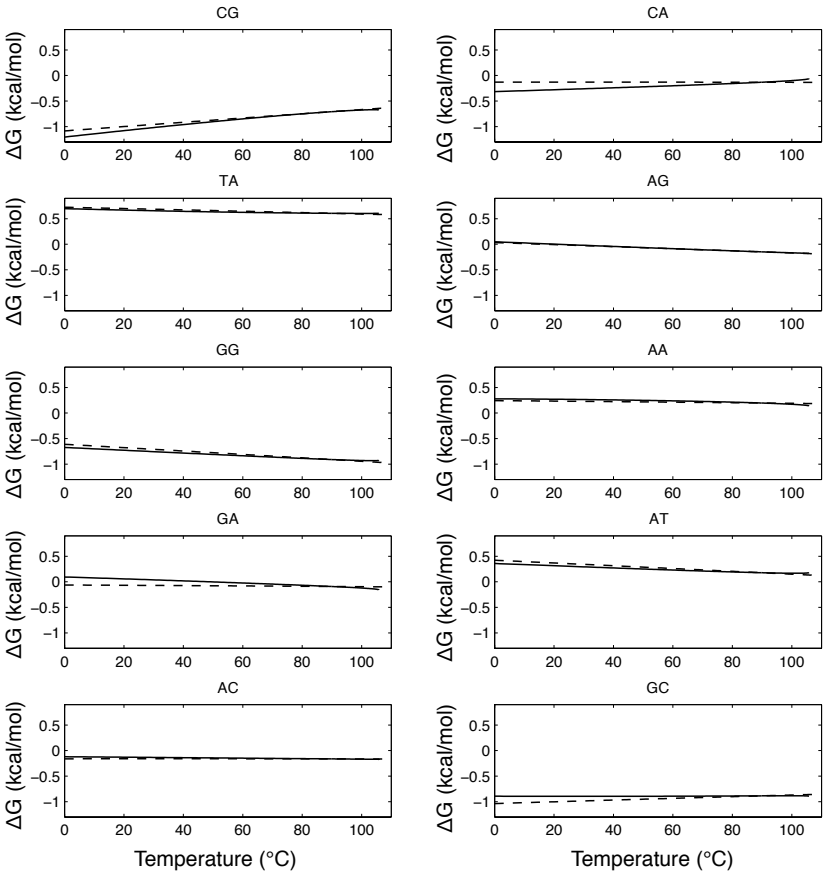

Figure S9: Base-pair step stabilities as predicted by the Benham model (11) (dashed line), and our model which includes the sequence- and temperature-dependent elasticity of the double-helix (solid line). The differences are only marginal, since the dominant contribution comes from the sequence-dependent melting properties.

Supporting references

(1) Lavery, R., M. Moakher, J. Maddocks, D. Petkeviciute, and K. Zakrzewska. 2009. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* 37(17):5917–5929.

(2) Olson, W. K., M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X.-J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C.-S. Tung, E. Westhof, C. Wolberger, and H. M. Berman. 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* 313(1):229 – 237.

(3) Dickerson, R. 1989. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.* 17(5):1797–1803.

(4) Jones, E., T. Oliphant, and P. Peterson. 2001. Scipy: Open source scientific tools for python.

(5) Hunter, J. 2007. Matplotlib: a 2d graphics environment. *Computing in Science & Engineering* pages 90–95.

(6) Lavery, R., K. Zakrzewska, D. Beveridge, T. C. Bishop, D. A. Case, T. r. Cheatham, S. Dixit, B. Jayaram, F. Lankas, C. Laughton, J. H. Maddocks, A. Michon, R. Osman, M. Orozco, A. Perez, T. Singh, N. Spackova, and J. Sponer. 2010. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.* 38(1):299–313.

(7) Flyvbjerg, H. and H. Petersen. 1989. Error estimates on averages of correlated data. *J. Chem. Phys.* 91(1):461–466.

(8) Press, W., S. Teukolsky, W. Vetterling, and B. Flannery. 2007. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press. ISBN 0521880688.

(9) Olson, W. K., A. A. Gorin, X. J. Lu, L. M. Hock, and V. B. Zhurkin. 1998. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA* 95(19):11163–11168.

(10) Lankas, F., J. Sponer, J. Langowski, and T. E. r. Cheatham. 2003. DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* 85(5):2872–2883.

(11) Benham, C. J. 1996. Theoretical analysis of the helix-coil transition in positively superhelical DNA at high temperatures. *Phys. Rev. E* 53(3):2984.