SUPPLEMENTARY INFORMATION

Compensatory Mechanisms in Temperature Dependence of DNA Double Helical Structure: Bending and Elongation

Hana Dohnalová[1], Tomáš Geršata[1,2], Jiří Šponer[2], Martin Zacharias[3], Jan Lipfert[4*], Filip Lankaš[1*]

1 Department of Informatics and Chemistry, University of Chemistry and Technology Prague, Technická 5, 166 28 Prague, Czech Republic

2 Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic

3 Physics-Department T38, Technical University of Munich, James-Franck-Strasse 1, 85748 Garching, Germany

4 Department of Physics and Center for Nanoscience, LMU Munich, Amalienstr. 54, 80799 Munich, Germany

*To whom correspondence should be addressed: filip.lankas@vscht.cz, Jan.Lipfert@lmu.de

S1. Supplementary Theory

A. Helix and helicity

Assume a circular helix whose helical axis coincides with the z-axis of a fixed Cartesian coordinate system. The points **s** constituting the helix can be parameterized by the relation

$$\mathbf{s} = \left(r\cos\omega, r\sin\omega, \frac{l}{2\pi}\omega\right) \tag{A1}$$

where r is the helix radius, l is the pitch and ω is the parameter. The unit tangent vector **t** is

$$\mathbf{t} = d\mathbf{s}/|d\mathbf{s}| \tag{A2}$$

where

$$d\mathbf{s} = \left(-r\sin\omega, r\cos\omega, \frac{l}{2\pi}\right) d\omega \tag{A3}$$

$$|d\mathbf{s}| = \sqrt{r^2 + (l/2\pi)^2}\, d\omega \tag{A4}$$

Substituting eq. A3 and A4 into eq. A2, we find

$$\mathbf{t} = \frac{1}{\sqrt{r^2 + (l/2\pi)^2}}\left(-r\sin\omega, r\cos\omega, \frac{l}{2\pi}\right) \tag{A5}$$

The angle β_c between the tangent vector **t** and the helix axis vector, that is, the coordinate vector **z** = (0,0,1), is given by

$$\cos\beta_c = \mathbf{t}\cdot\mathbf{z} \tag{A6}$$

Inserting the expression for **t** (eq. A5) into the right-hand side, we obtain

$$\cos\beta_c = \frac{l}{2\pi\sqrt{r^2 + (l/2\pi)^2}} \tag{A7}$$

On the other hand, the helix contour length L_c between its points A corresponding to $\omega = 0$ and B corresponding to $\omega = \Omega$ is computed as

$$L_c = \int_A^B |d\mathbf{s}| \tag{A8}$$

Inserting the parameterization of the length element $|d\mathbf{s}|$ from eq. A4, we have

$$L_c = \int_0^{\Omega} \sqrt{r^2 + (l/2\pi)^2}\, d\omega \tag{A9}$$

which yields

$$L_c = \sqrt{r^2 + (l/2\pi)^2}\,\Omega \tag{A10}$$

The length L_h of the same helical fragment but measured along the helical axis, i.e. the spring length, is given by

$$L_h = \int_A^B \mathbf{z} \cdot d\mathbf{s} \tag{A11}$$

Inserting the expression for $d\mathbf{s}$ (eq. A3) into the right-hand side, we find

$$L_h = \int_0^{\Omega} \frac{l}{2\pi}\, d\omega \tag{A12}$$

which yields

$$L_h = \frac{l}{2\pi}\Omega \tag{A13}$$

Thus, it follows from eq. A10 and A13 that

$$\frac{L_h}{L_c} = \frac{l}{2\pi\sqrt{r^2 + (l/2\pi)^2}} \tag{A14}$$

and, comparing this with eq. A7, we have the final result

$$\cos\beta_c = \frac{L_h}{L_c} \tag{A15}$$

In the computations described in the main text, the contour length L_c is approximated by L, the sum of basepair center distances, and β_c is approximated by β defined by

$$\cos\beta = \frac{L_h}{L} \tag{A16}$$

B. Ensemble average of a function

Consider a quantity *y* as a function of *n* variables $x_1, \dots, x_n$. We introduce the vector

$$\mathbf{x} = (x_1, \dots, x_n) \tag{B1}$$

and write

$$y = f(\mathbf{x}). \tag{B2}$$

Expanding the function *f* in a Taylor series and keeping only terms up to the second order, we obtain

$$y = f(\hat{\mathbf{x}}) + \sum_{i=1}^{6N} \frac{\partial f}{\partial x_i}(\hat{\mathbf{x}})(x_i - \hat{x}_i) + \sum_{i,j=1}^{6N} \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{\mathbf{x}})(x_i - \hat{x}_i)\big(x_j - \hat{x}_j\big) \tag{B3}$$

where we introduced a reference vector

$$\hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_n) \tag{B4}$$

for which we choose the ensemble average of **x** at the reference temperature T_0,

$$\hat{\mathbf{x}} = \langle \mathbf{x} \rangle_{T_0}. \tag{B5}$$

Taking now the ensemble average of *y* at temperature *T*, we get

$$\langle y \rangle_T = f(\hat{\mathbf{x}}) + \sum_{i=1}^{6N} \frac{\partial f}{\partial x_i}(\hat{\mathbf{x}})(\langle x_i \rangle_T - \hat{x}_i) + \sum_{i,j=1}^{6N} \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{\mathbf{x}})\langle (x_i - \hat{x}_i)\big(x_j - \hat{x}_j\big)\rangle_T \tag{B6}$$

The first two terms on the right-hand side are a Taylor expansion up to the first order of $f(\langle \mathbf{x} \rangle_T)$. Assuming that the coordinate means change only slightly with temperature, that is, $\langle \mathbf{x} \rangle_T$ is close to $\hat{\mathbf{x}} = \langle \mathbf{x} \rangle_{T_0}$, we can write, to a high precision,

$$f(\hat{\mathbf{x}}) + \sum_{i=1}^{6N} \frac{\partial f}{\partial x_i}(\hat{\mathbf{x}})(\langle x_i \rangle_T - \hat{x}_i) = f(\langle \mathbf{x} \rangle_T) \tag{B7}$$

The quadratic term can be simplified as follows. Writing

$$x_i - \hat{x}_i = x_i - \langle x_i \rangle_T + \langle x_i \rangle_T - \hat{x}_i \tag{B8}$$

we find

$$\langle (x_i - \hat{x}_i)\big(x_j - \hat{x}_j\big)\rangle_T = C_{ij} + (\langle x_i\rangle_T - \hat{x}_i)\big(\langle x_j\rangle_T - \hat{x}_j\big) \tag{B9}$$

where

$$C_{ij} = \langle (x_i - \langle x_i\rangle_T)\big(x_j - \langle x_j\rangle_T\big)\rangle_T \tag{B10}$$

are the elements of the coordinate covariance matrix *C* at temperature *T*, which is associated with the stiffness matrix *K* by the relation

$$C_{ij} = k_B T[K^{-1}]_{ij} \tag{B11}$$

where the superscript -1 denotes the matrix inverse and k_B is the Boltzmann constant. We further neglect the term $(\langle x_i\rangle_T - \hat{x}_i)\big(\langle x_j\rangle_T - \hat{x}_j\big)$ which is quadratic in the temperature differences of the coordinate means and is supposed to be very small. Taken together, we have the approximation

$$\langle y\rangle_T = f(\langle \mathbf{x}\rangle_T) + T\left[k_B \sum_{i,j=1}^{6N} \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{\mathbf{x}})[K^{-1}]_{ij}\right] \tag{B12}$$

If the stiffness is temperature independent, then the term in the square braces is a constant. However, as shown in ref. (1), the entropic contribution to the deformation free energy cannot be neglected and therefore the stiffness also depends on temperature. We assume a linear dependence of the form

$$[K^{-1}]_{ij} = a_{ij} + (T - T_0)b_{ij} \tag{B13}$$

Substituting this into the previous equation and keeping only terms linear in $(T - T_0)$, we obtain

$$\langle y\rangle_T = f(\langle \mathbf{x}\rangle_T) + a + b(T - T_0) \tag{B14}$$

where the constants *a* and *b* are given by

$$a = k_B T_0 \sum_{i,j=1}^{6N} \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{\mathbf{x}})a_{ij}$$

$$b = k_B \sum_{i,j=1}^{6N} \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{\mathbf{x}})\big(a_{ij} + T_0 b_{ij}\big).$$

(B15)

The constants *a* and *b* can be inferred using eq. B14 by least squares fitting of the data. If

$$a \approx b(T - T_0) \approx 0 \tag{B16}$$

then we simply have the estimate

$$\langle y \rangle_T = f(\langle \mathbf{x} \rangle_T) \tag{B17}$$

that is, the operation of ensemble averaging and the function *f* can be interchanged. This was the approximation used in our previous study (2) and derived there using a slightly different argument. However, in the cases studied in the present work we often have

$$a \neq 0, \ b(T - T_0) \approx 0 \tag{B18}$$

and eq. B14 then implies the estimate

$$\langle y \rangle_T - \langle y \rangle_{T_0} = f(\langle \mathbf{x} \rangle_T) - f\big(\langle \mathbf{x} \rangle_{T_0}\big). \tag{B19}$$

Thus, in this approximation, the difference of the means of a function can be replaced by the difference of the function of the means. This enables one to study the effect of individual ensemble averaged components of **x** on the change of *y*, as done in the main text.

C. Helical rise and the local coordinates

Consider a basepair step consisting of two successive base pairs, 1 and 2, equipped with reference points O_1 and O_2 and right-handed, orthonormal frames ($\mathbf{x}_1$, $\mathbf{y}_1$, $\mathbf{z}_1$) and ($\mathbf{x}_2$, $\mathbf{y}_2$, $\mathbf{z}_2$). According to the 3DNA definition (3), the helical rise *h* is the projection of the basepair center distance $d = |O_2 - O_1|$ onto the local helical axis (in what follows we drop the basepair step index *a* for simplicity – we write *h* instead of h^a, *d* instead of d^a, β instead of β^a and so on). The coordinate systems associated with the two base pairs are representesented by matrices of column component vectors R_1 and R_2, related to each other by a relative rotation represented by the rotation matrix *R*,

$$R_2 = R_1 R \tag{C1}$$

Working in the coordinate system ($\mathbf{x}_1$, $\mathbf{y}_1$, $\mathbf{z}_1$) of base pair 1 implies that

$$R_1 = I, R_2 = R \tag{C2}$$

where *I* is the identity matrix. In the 3DNA coordinate definition (3), *R* takes the form

$$R = R_z\left(\frac{\Omega}{2} - \phi\right) R_y(\Gamma) R_z\left(\frac{\Omega}{2} + \phi\right) \tag{C3}$$

where Ω is the local twist, Γ the local bending angle, and ϕ the phase angle determining the bending direction. Here R_z and R_y are the rotation matrices about the coordinate axes *z* and *y* which for any angle γ are given by

$$R_z(\gamma) = \begin{pmatrix} c & -s & 0 \\ s & c & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad R_y(\gamma) = \begin{pmatrix} c & 0 & s \\ 0 & 1 & 0 \\ -s & 0 & c \end{pmatrix} \tag{C4}$$

where $c = \cos\gamma$, $s = \sin\gamma$. The angles Γ and ϕ are related to roll (ρ) and tilt (τ) as

$$\rho = \Gamma\cos\phi, \quad \tau = \Gamma\sin\phi \tag{C5}$$

The translational coordinates shift, slide and rise are the components of the vector $O_2 - O_1$ in the middle frame ($\mathbf{x}_m$, $\mathbf{y}_m$, $\mathbf{z}_m$) defined by the matrix R_m of its column component vectors. In the coordinate system of base pair 1, R_m takes the form (3)

$$R_m = R_z\left(\frac{\Omega}{2} - \phi\right) R_y(\Gamma/2) R_z(\phi) \tag{C6}$$

We will adopt a simplified model of DNA geometry. We first assume that bending takes place exclusively via roll (i.e. tilt is always zero), so that the phase angle ϕ is 0 or π. This is justified by the fact that tilt is usually very small in real DNA and its thermal change is also much smaller on average than that of roll (Figure S2). Thus, the bending angle Γ is just the magnitude of roll, and we assume it to be small, so that we can adopt the approximations $\cos\Gamma \approx 1$, $\sin\Gamma \approx \Gamma$. We further assume that shift and slide are both zero, so that $O_2 - O_1$is parallel to $\mathbf{z}_m$, the middle frame *z*-axis, and *d* coincides with rise. This is justified, since shift and slide in DNA are usually small and the temperature dependence of *d* is dominated by that of rise (Figure S3).

The vector **v** of the local helical axis, that is, the axis of the screw transformation connecting the two basepair coordinate systems, is given by (3)

$$\mathbf{v} = (\mathbf{x}_2 - \mathbf{x}_1) \times (\mathbf{y}_2 - \mathbf{y}_1) \tag{C7}$$

and the helical rise is defined as

$$h = d\cos\beta \tag{C8}$$

where β, the local helicity, is the angle between the vector $O_2 - O_1$ and the local helical axis vector **v**. In the approximations stated above, we have

$$\cos\beta = \mathbf{v}\cdot\mathbf{z}_m/|\mathbf{v}| \tag{C9}$$

Inserting the component vectors into eq. C7 and C9 and performing the computation, we find eq. C8 to take the form

$$h = d\left(1 - \frac{\Gamma^2}{2\Omega^2}\right) \tag{C10}$$

where we also used the approximation $\sin\frac{\Omega}{2} \approx \frac{\Omega}{2}$. Eq. C10 is identical to eq. 11 in the main text.

S2. Supplementary Methods

A 33 base pair (bp) DNA duplex of the sequence

GAGAT-GCTAA-CCCTG-ATCGC-TGATT-CCTTG-GAC

(here divided into 5 bp blocks for better readability) was built in the canonical B-DNA fiber diffraction form (Arnott B-DNA (4)) using the *nab* module of Amber. The Amber OL15 force field (5) was used to parameterize the DNA internatomic interactions. The duplex was then solvated in a truncated octahedron box using the SPC/E water model (6), with a minimal distance of 10 Å between the DNA and the box walls. In the next step, K+ and Cl- ions parameterized according to Dang (7) were added to the system to ensure its charge neutrality and to create an excess concentration of 150 mM KCl, close to the physiological value. In total, 174 K+, 110 Cl- and ca. 40,000 water molecules were included in the simulated system, which contains roughly 123,000 atoms. The ion positions were randomized using the *cpptraj* module of Amber, so that the ions were no closer than 5 Å from the DNA and 3.5 Å from each other.

The equilibration was performed by first minimizing the energy of the system (500 steps of steepest descent and 500 steps of conjugated gradients), restraining the DNA atoms to their initial positions by harmonic restraints using a 25 kcal.mol^{-1}.Å^{-2} force constant. This was followed by heating the system from the initial 100 K to 300 K in a 100 ps NVT molecular dynamics, using the Berendsen thermostat with the default coupling time of 1 ps and 25 kcal.mol^{-1}.Å^{-2} restraints on the DNA. Further, six rounds of energy minimization, each followed by a 50 ps NVT dynamics at 300 K, were performed, gradually releasing the DNA restrains from 5 kcal.mol^{-1}.Å^{-2} through 4, 3, 2, 1 to 0.5 kcal.mol^{-1}.Å^{-2}. Finally, 50 ps of unrestrained NpT dynamics at 300 K and the pressure of 1 atm, using Berendsen thermostat and barostat with coupling times of 5 ps, was performed.

The production was performed in the NpT ensemble at the target temperature (280, 290, 300, 310 or 320 K) and the 1 atm pressure using the Berendsen thermostat and barostat with 5 ps coupling times, periodic boundary conditions, the Particle Mesh Ewald method to compute electrostatic interactions, and SHAKE restraints on hydrogen atoms. The trajectories were prolonged to 1 μs each. In the production phase, mass repartitioning as implemented in Amber was introduced, which allowed us to use a 4 fs time step to integrate the equations of motion. The *pmemd* module of Amber was used for the equilibration and production phases.

S3. Supplementary Figures

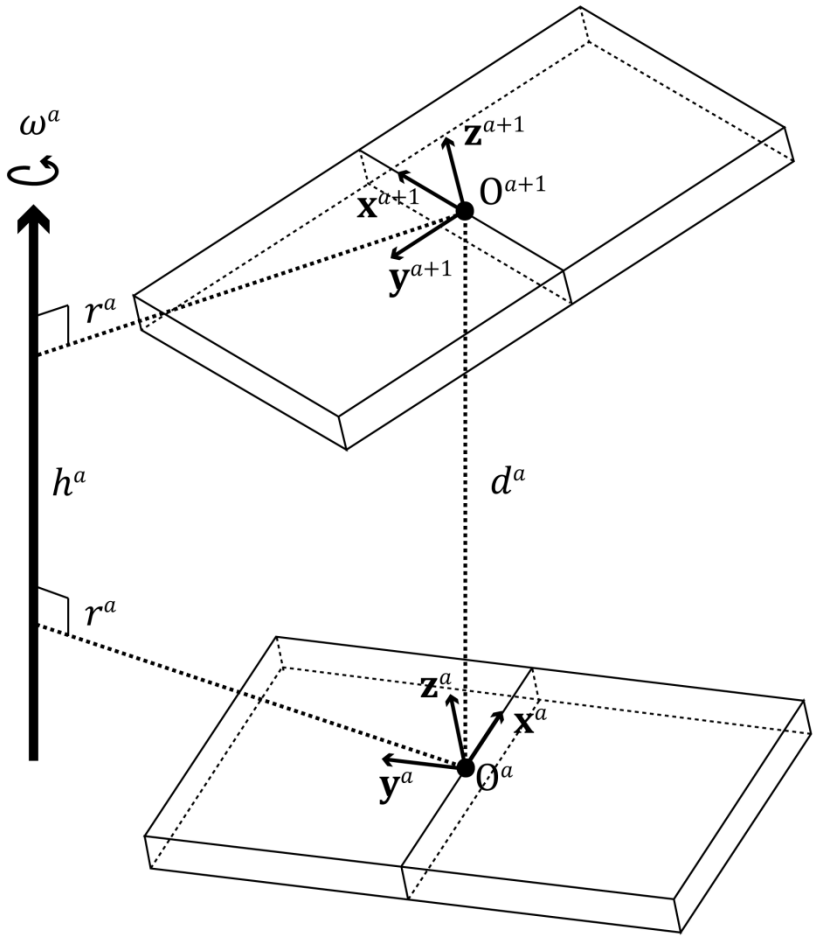

Figure S1. Conformational descriptors used in this work.

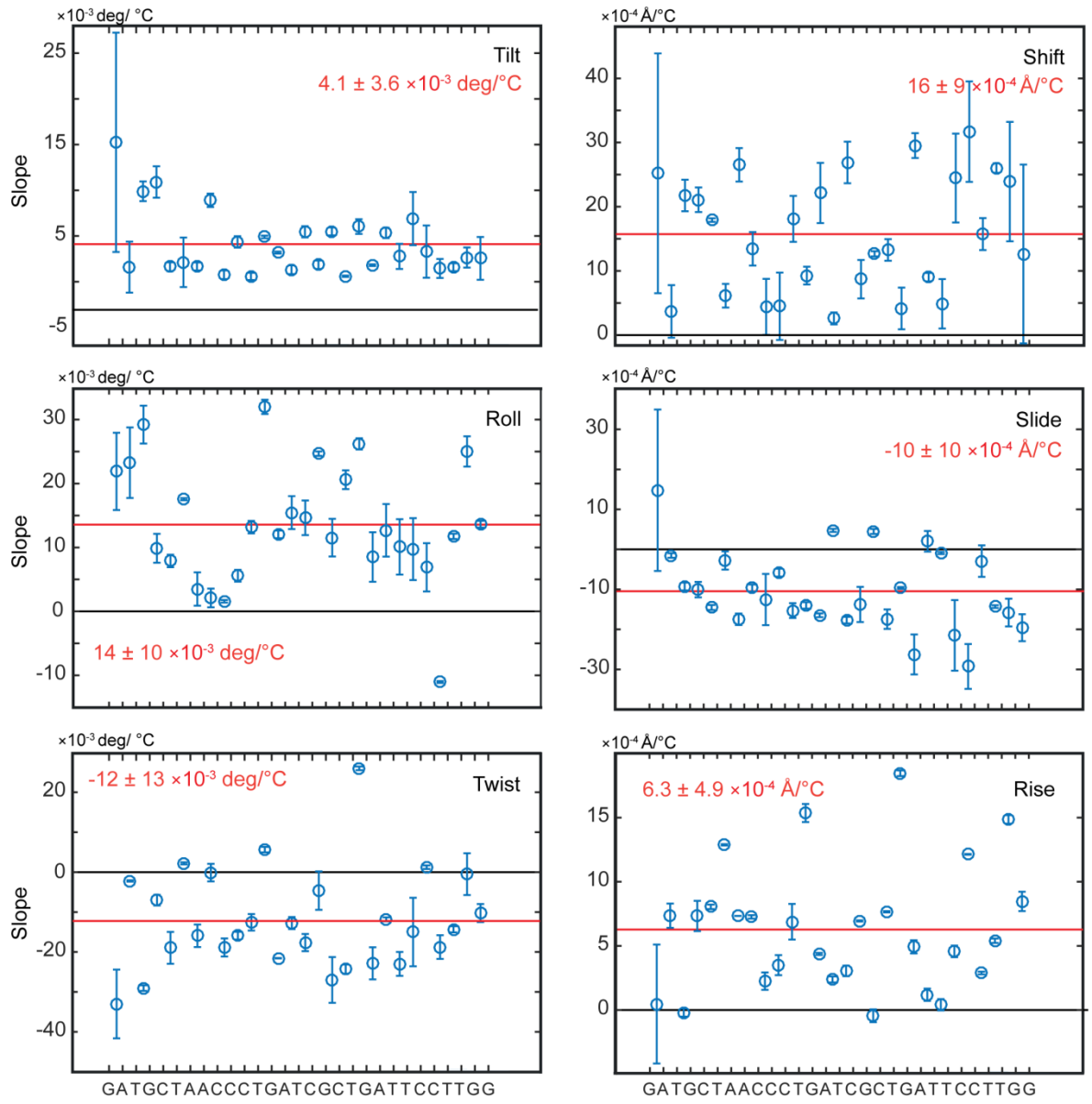

Figure S2. Temperature dependencies of local basepair step coordinates . Values for individual steps between base pairs 3 and 31 (28 steps in total) are shown. Since tilt and shift change sign upon change of the strand selected as the reference strand, the sign of their thermal slopes also depend on this choice. Therefore, absolute values of the tilt and shift thermal slopes are shown. Means over the 28 steps, and standard deviations indicating the spread of the 28 values, are displayed.

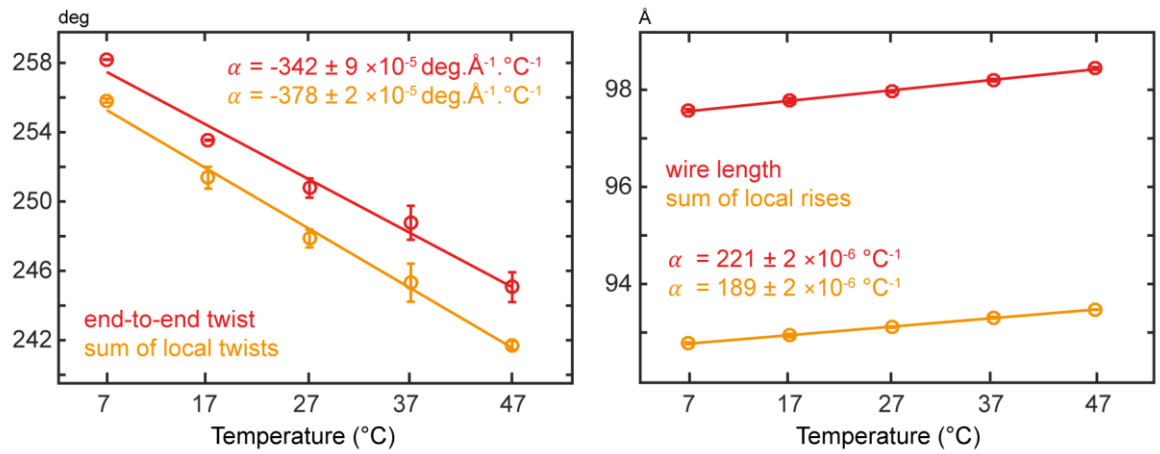

Figure S3. Effect of approximating the end-to-end twist by accumulated local twists (left) and the wire length by accumulated local rises (right). Ensemble averages for each temperature, approximated by averages over the MD trajectory, are shown.

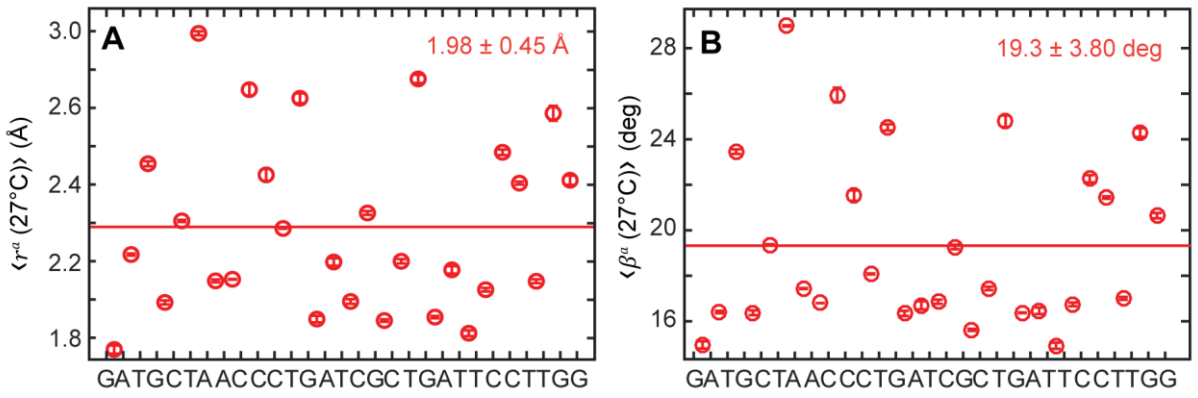

Figure S4. Local radii of the basepair centerline helix (left) and local helicities (right) at 27 °C. Ensemble averages approximated by the means over the MD trajectory are shown. The standard deviations indicate the spread of values of the individual steps.

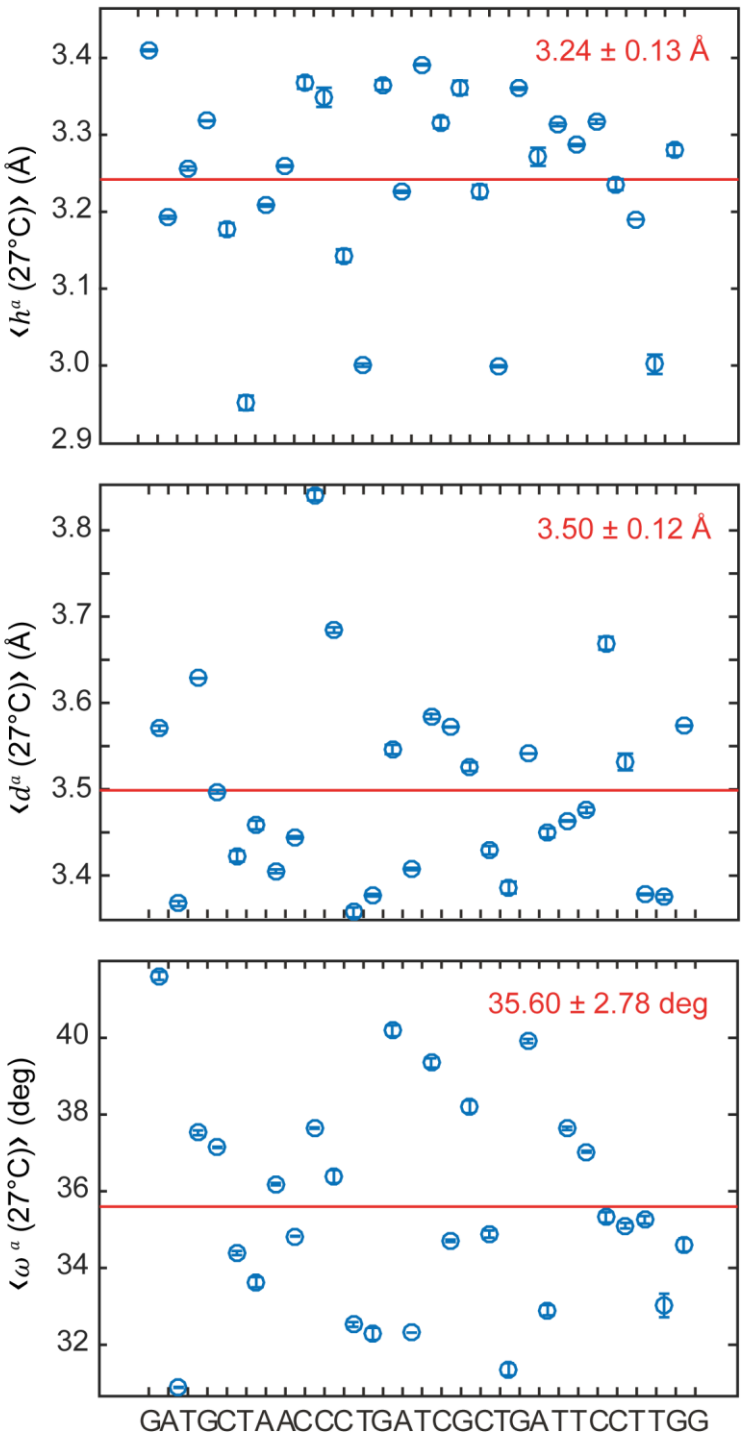

Figure S5. Local helical rises, distances between neighbouring basepair centres, and local helical twists at 27 °C. The standard deviations indicate the spread of values of the individual steps.

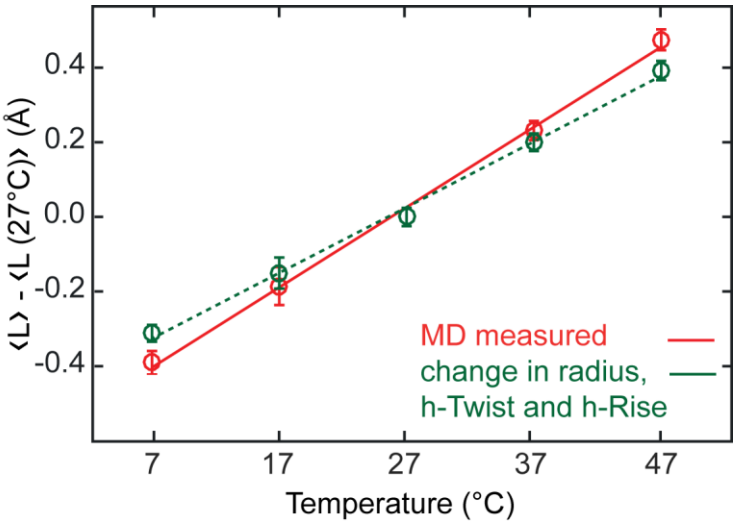

Figure S6. Effect of the approximation used in eq. 7 of the main text. The ensemble average of the difference in wire length L (left-hand side of the equation, red) is estimated by the difference of the function of the ensemble averages of centerline radius, helical rise and helical twist (right-hand side of the equation, green). The data in the Figure indicate good quality of the approximation. Ensemble averages are estimated by averages over the MD trajectory.

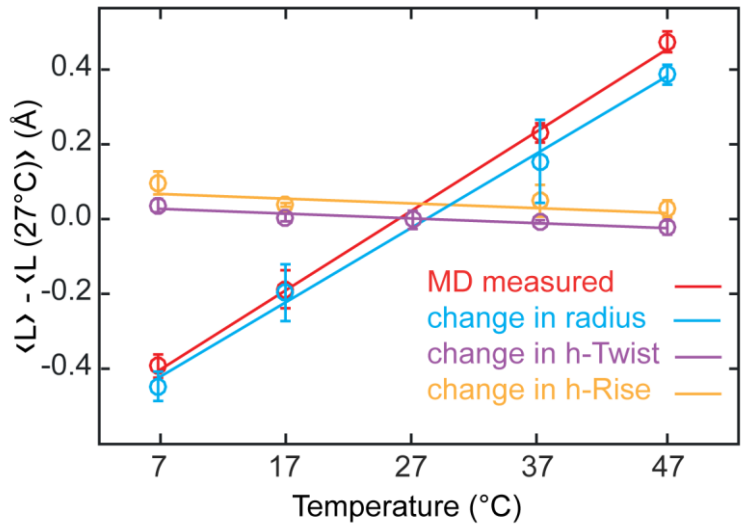

Figure S7. Effect of helical coordinates on the thermal expansion of the wire length L. In eq. 7, each type of coordinates was changed individually, while the others were kept at their ensemble averaged values for $T = T_0$ (27 °C). The change actually observed in the MD data (red) is well approximated by changing the helical radii $\langle r^a \rangle_T$ only (blue), while the effect of the helical twists $\langle \omega^a \rangle_T$ (magenta) and helical rises $\langle h^a \rangle_T$ (yellow) is very weak.

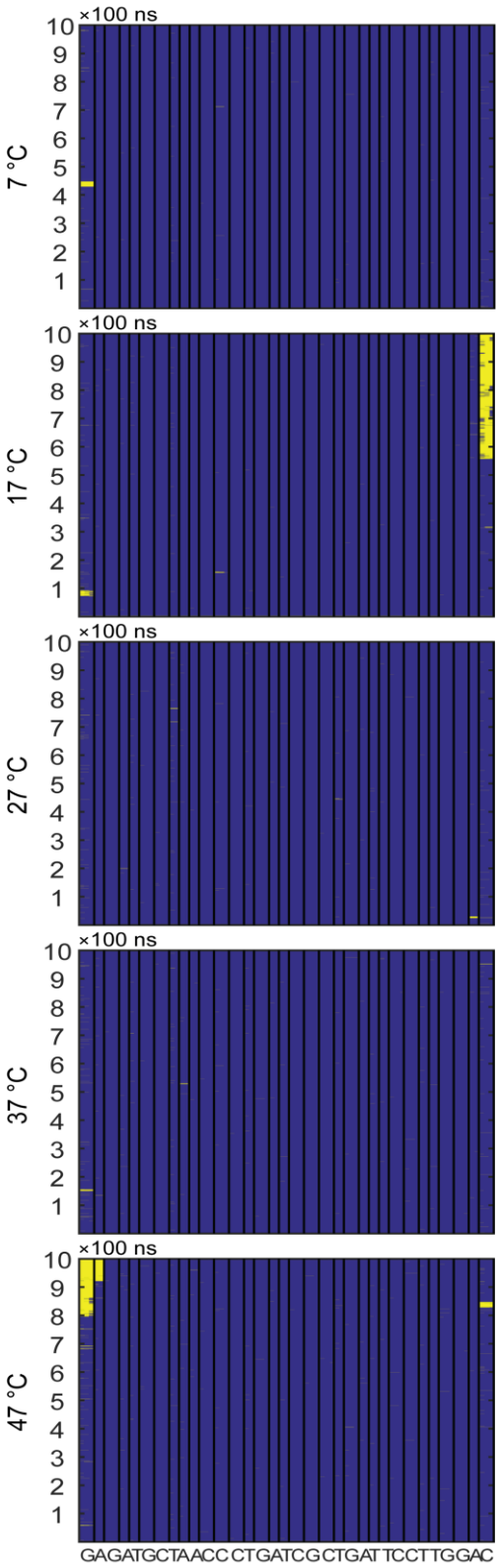

Figure S8. Stability of the hydrogen bonds connecting the two bases in a Watson-Crick pair. Heavy atom distances exceeding 4 Å are in yellow. Black vertical lines separate individual pairs.

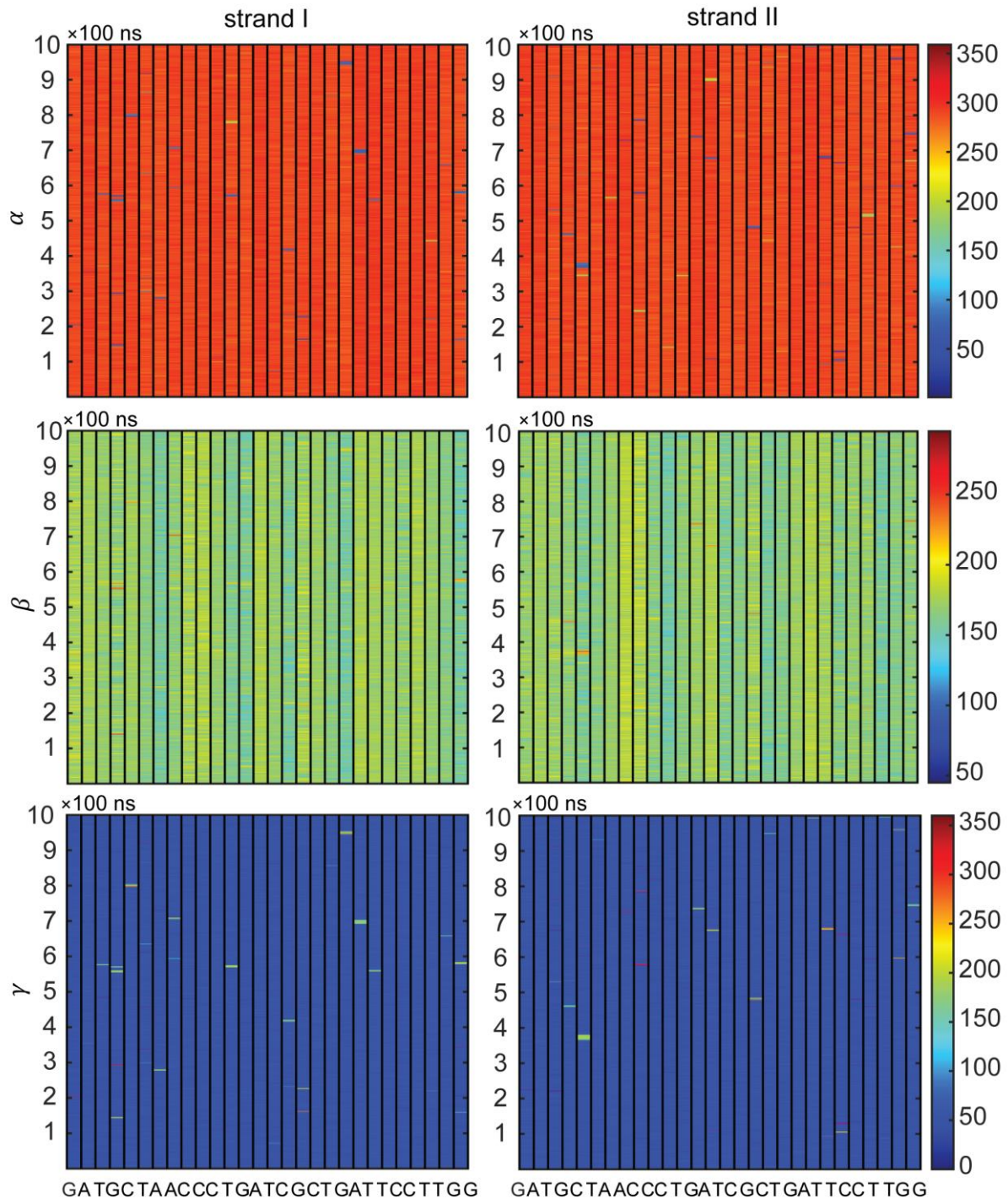

Figure S9. Time series of the backbone torsion angles α, β and γ at 27 ºC. The angles mostly occupy the canonical domain (α/β/γ in *g-/t/g+*) but also show rare and short-lived flips to non-canonical values.

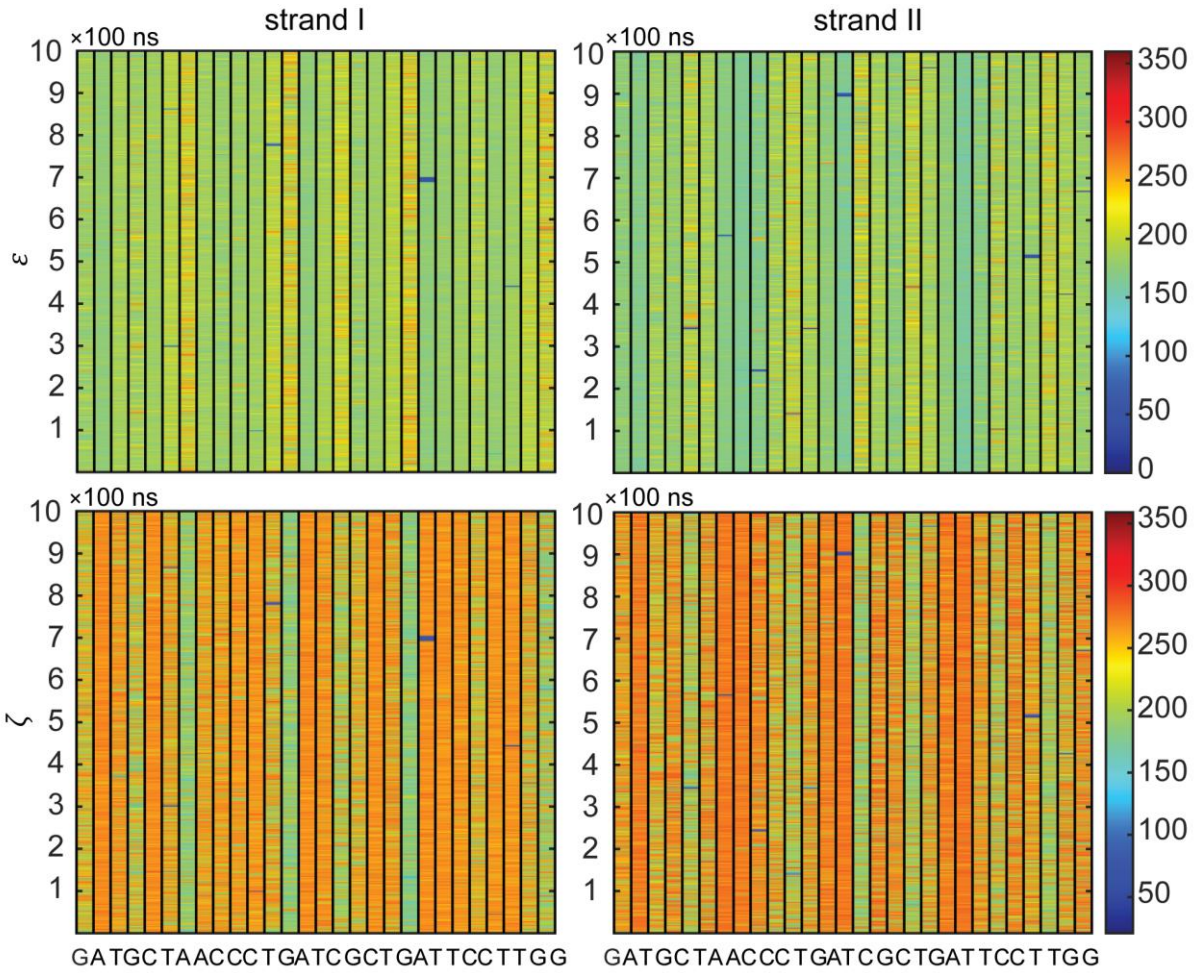

Figure S10. Time series of the backbone torsion angles ε and ζ at 27 ºC. Frequent flips between the BI (ε/ζ in *t*/*g*-) and BII (ε/ζ in *g*-/*t*) are observed. Besides that, a rare non-canonical state of ε/ζ in *g*+/*g*+ accompanied by α in *t* (Figure S9) is observed.

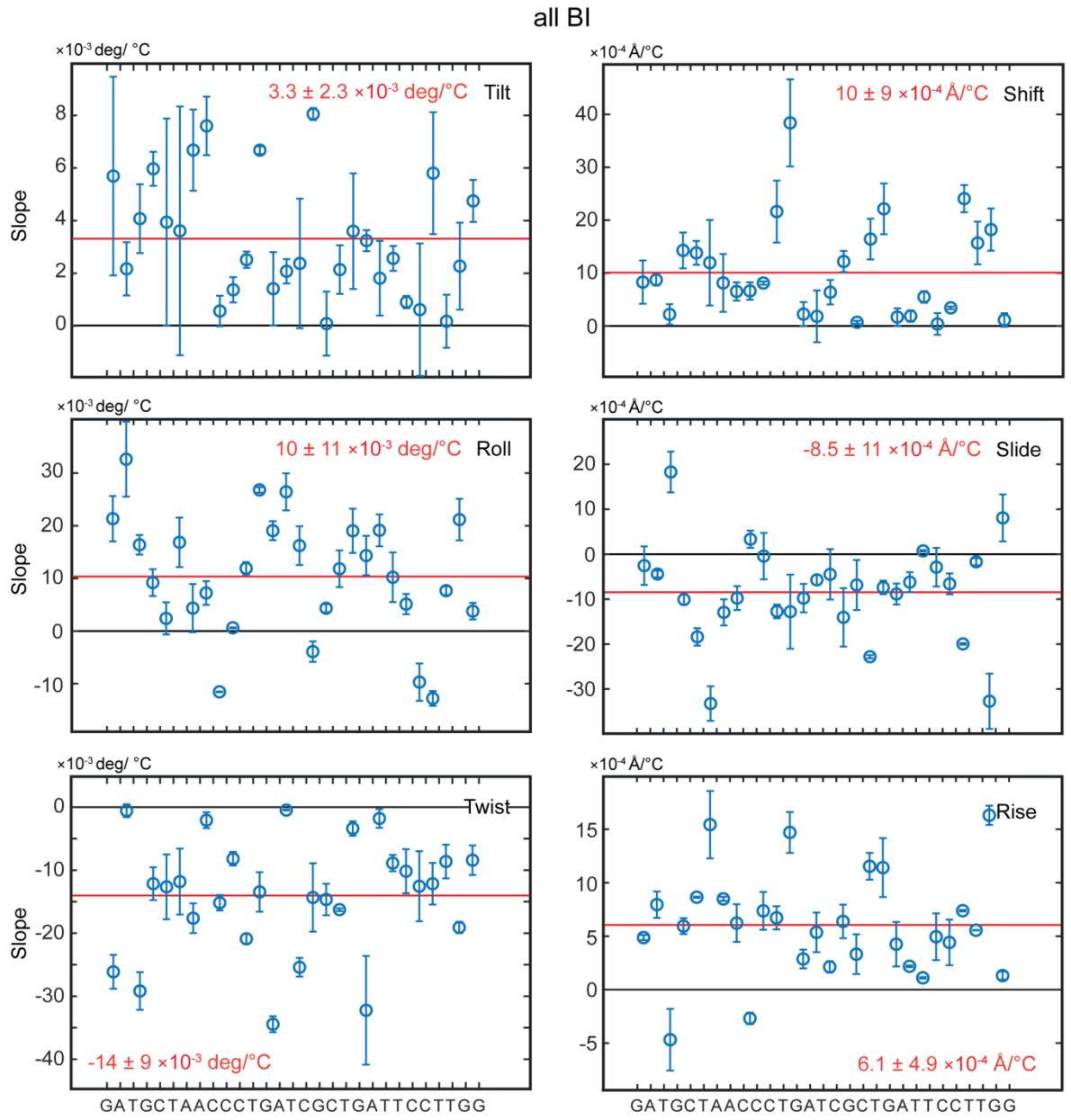

Figure S11. Temperature dependencies of DNA local coordinates filtered for the BI state. For each step, only those MD snapshots were taken into account in which the backbone fragments in the step and in its 5' and 3' neighbouring steps (6 fragments in total) were all in the BI state. The mean values are similar to those obtained without this filtering (Figure S2).

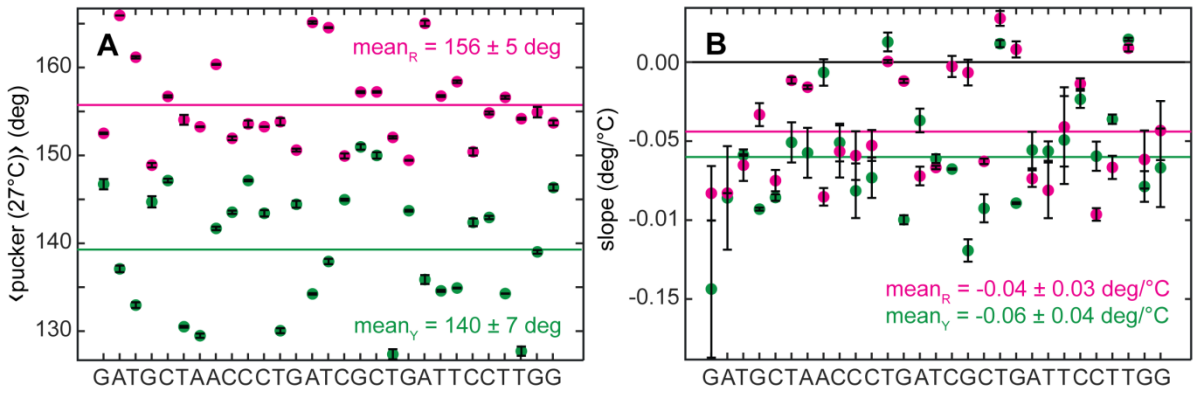

Figure S12. Mean values of the sugar puckers (A) and their temperature dependencies (B). While the means for pyrimidines (Y) are systematically lower than those for purines (R), no such distinction can be observed for the temperature slopes. These are rather noisy, but the sequence averaged pucker clearly decreases with temperature (Figure 6 of the main text).

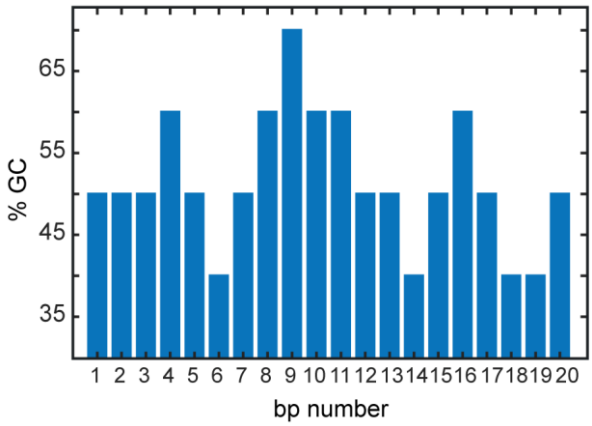

Figure S13. GC content in 10 bp sliding windows covering the portion of the oligomer between basepair 3 and 31 (29 bp in total) examined in the main text. The whole portion has 52 % GC while the GC content of the windows varies between 40 and 70 %.

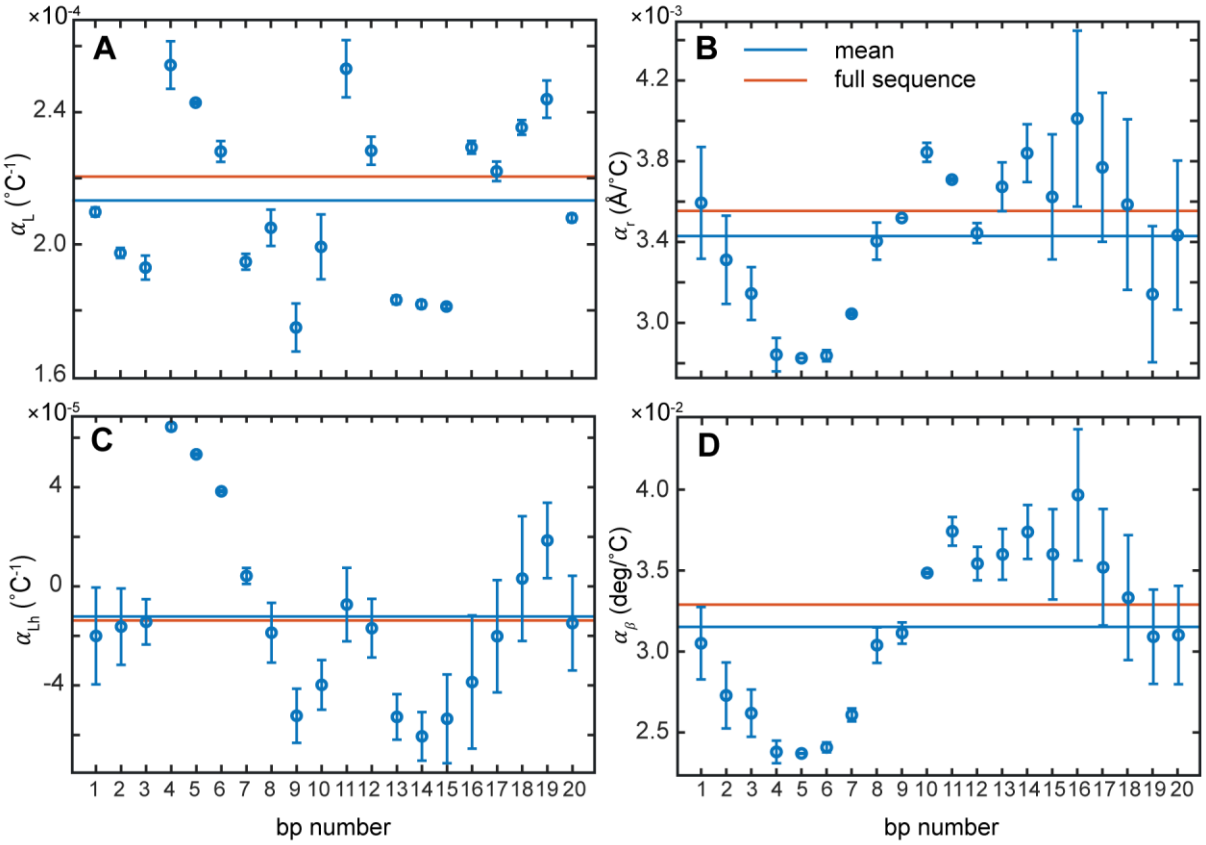

Figure S14. Thermal expansion coefficients for sequences in 10 bp sliding windows. The means over the values are shown together with the values for the whole 29 bp portion examined in the main text.

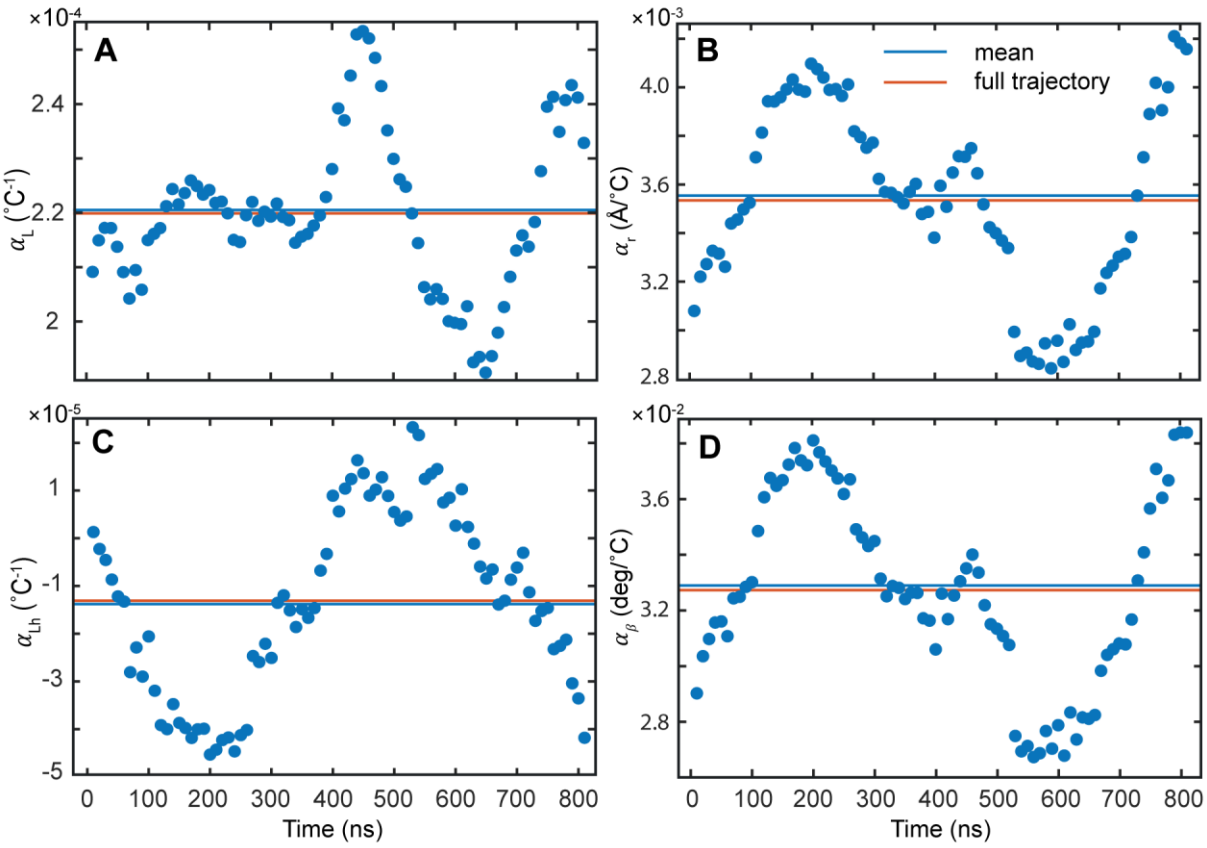

Figure S15. Thermal expansion coefficients computed for a 200 ns sliding window. The mean of the window values and the value for the whole microsecond trajectory are shown.

References

1. Meyer, S., Jost, D., Theodorakopoulos, N., Peyrard, M., Lavery, R. and Everaers, R. (2013) Temperature dependence of the DNA double helix at the nanoscale: Structure, elasticity, and fluctuations. *Biophys. J.*, **105**, 1904-1914.
2. Kriegel, F., Matek, C., Drsata, T., Kulenkampff, K., Tschirpke, S., Zacharias, M., Lankas, F. and Lipfert, J. (2018) The temperature dependence of the helical twist of DNA. *Nucleic Acids Res.*, **46**, 7998-8009.
3. Lu, X.-J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108-5121.
4. Arnott, S. and Hukins, D.W.L. (1972) Optimised parameters for A-DNA and B-DNA. *Biochem. Biophys. Research Commun.*, **47**, 1504-1509.
5. Zgarbova, M., Sponer, J., Otyepka, M., Cheatham III, T.E., Galindo-Murillo, R. and Jurecka, P. (2015) Refinement of the sugar-phosphate backbone torsion beta for Amber force fields improves the description of Z- and B-DNA. *J. Chem. Theory Comput.*, **11**, 5723-5736.
6. Berendsen, H.J.C., Grigera, J.R. and Straatsma, T.P. (1987) The missing term in effective pair potentials. *J. Phys. Chem.*, **91**, 6269-6271.
7. Dang, L.X. (1995) Mechanism and thermodynamics of ion selectivity in aqueous solutions of 18-crown-6 ether: a molecular dynamics study. *J. Am. Chem. Soc.*, **117**, 6954-6960.