

Supplementary Material

The impact of sequence periodicity on DNA mechanics: investigating the origin of A-tract's curvature

Tania Gardasevic and Agnes Noy*

School of Physics, Engineering and Technology, University of York, York, YO10 5DD, UK

E-mail: agnes.noy@york.ac.uk

Supplementary Figures

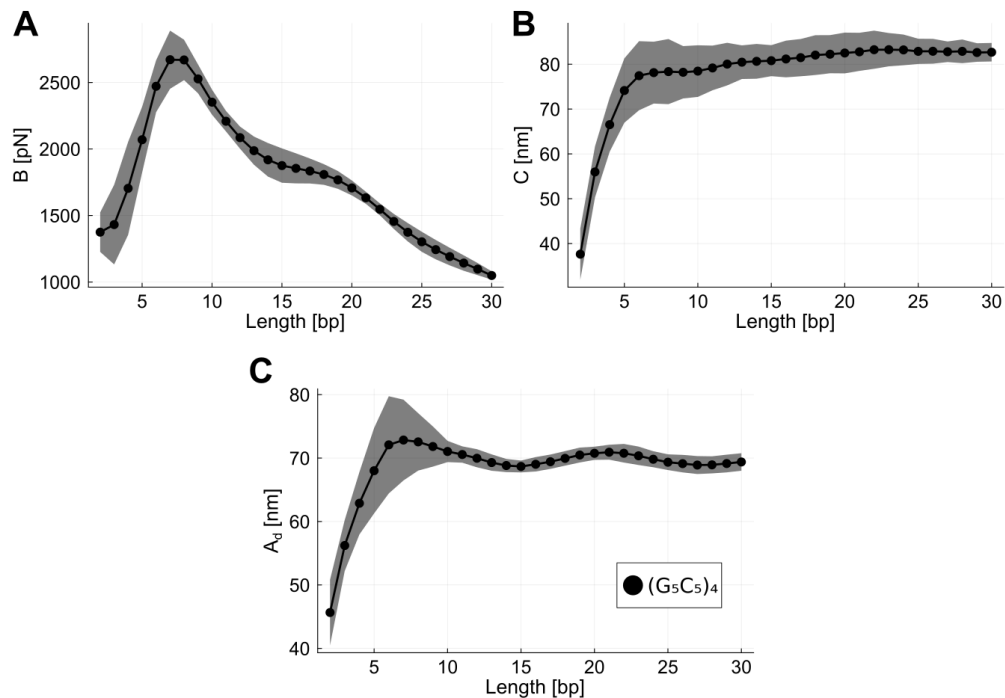


Figure S 1: The evolution of stretch modulus (B), torsional modulus (C) and dynamic persistence length (A_d) as a function of length using $(G_5C_5)_4$ as an example. (A) B defined by the end-to-end distance presents a non-monotonic profile. The strong base-stacking interactions cause stiffening from 2-7 bp in length, as the end-to-end distance is in close alignment with the molecular contour length. The emergence of coordinated motions between bp induces softening from 8-19 bp in length, as the end-to-end distance begins to differentiate from contour length, capturing oscillations in DNA crookedness. Strong softening is observed at 20-30 bp length as a result of long-ranged end effects. The linear fitting is performed using the intermediate lengths, which represent the same regime found in long DNAs and studied by single-molecule experiments. The reader is advised to consult references 26 and 31 from the main text for a more comprehensive explanation. (B-C) The profiles linked to C and A_d are characterized by a gradual increase in values as the lengths become longer, until they hit a plateau within a single DNA-turn.

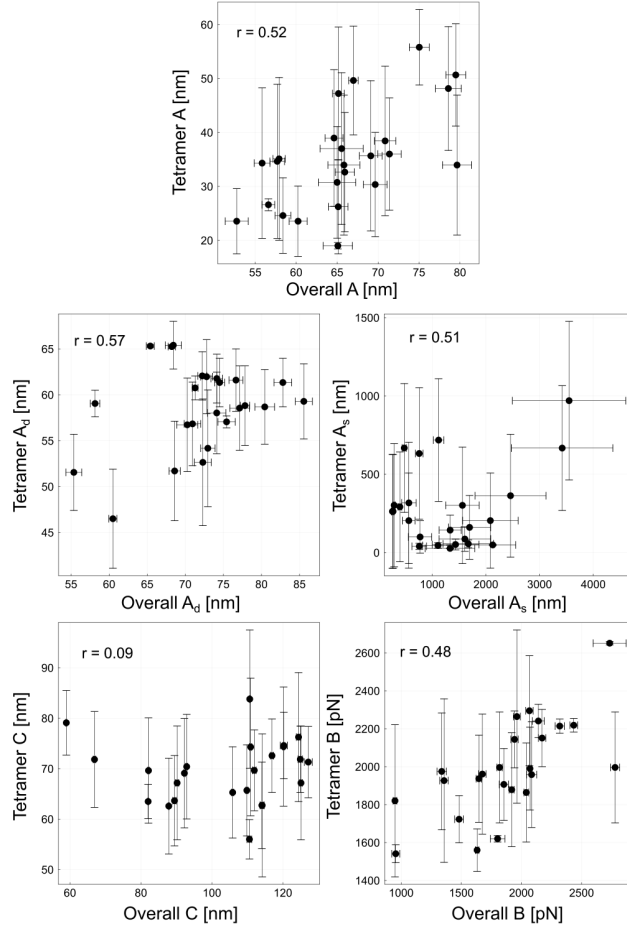


Figure S 2: Correlation between bulk elastic constants for each individual fragment and the averaged values (with standard deviation as error bars) obtained from all constituent tetramers using the ABC simulation database. Error bars for overall bending (A , A_d and A_s) and stretching (B) elastic constants are the 70% confidence intervals of the linear fitting, while for the overall twisting elastic constant (C) is the standard deviation obtained from considering all sub-fragments longer than 10 bp, as it is calculated by SerraNA and described in the Method section. R are the correlation coefficients between the two quantities.

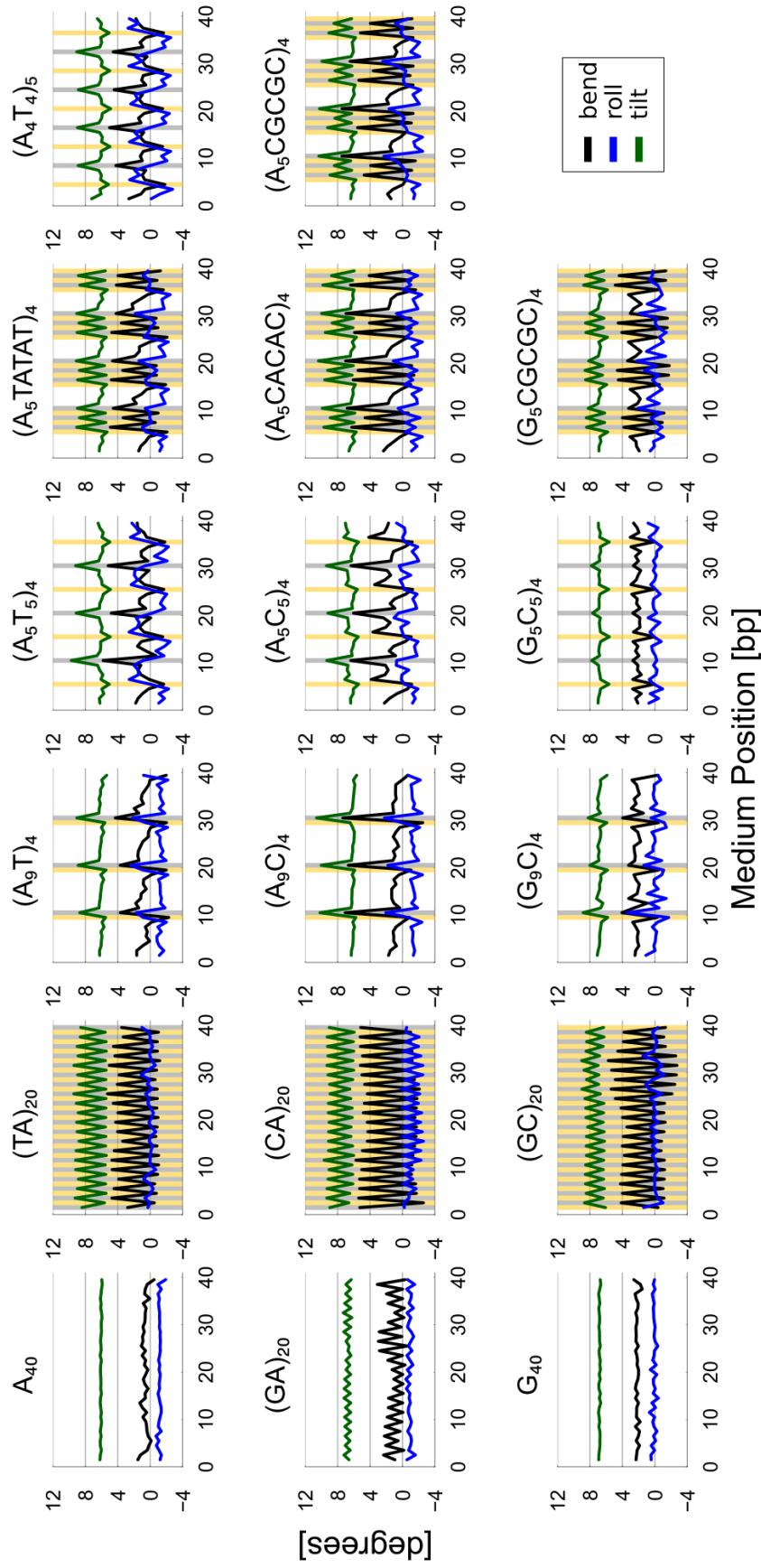


Figure S 3: Bending profiles along DNA molecules at the 2 bp level, together with its two components: roll (towards groove) and tilt (towards backbone). Sequence patterns $((R)_{40}, (YR)_{20}, (R_9Y)_4, (R_5Y_5)_4$ and $(R_5YRYRYR)_4$ are organised by columns and base groups $(A/T, A/C$ and $G/C)$ are organised by rows with the exception of $(A_5CGGGC)_4$ and the out-of-phase periodic sequence $(A_4T_4)_5$. YR and RY bp steps are shaded in grey and yellow, respectively. Standard deviations have been excluded to aid visualisation of key trends.

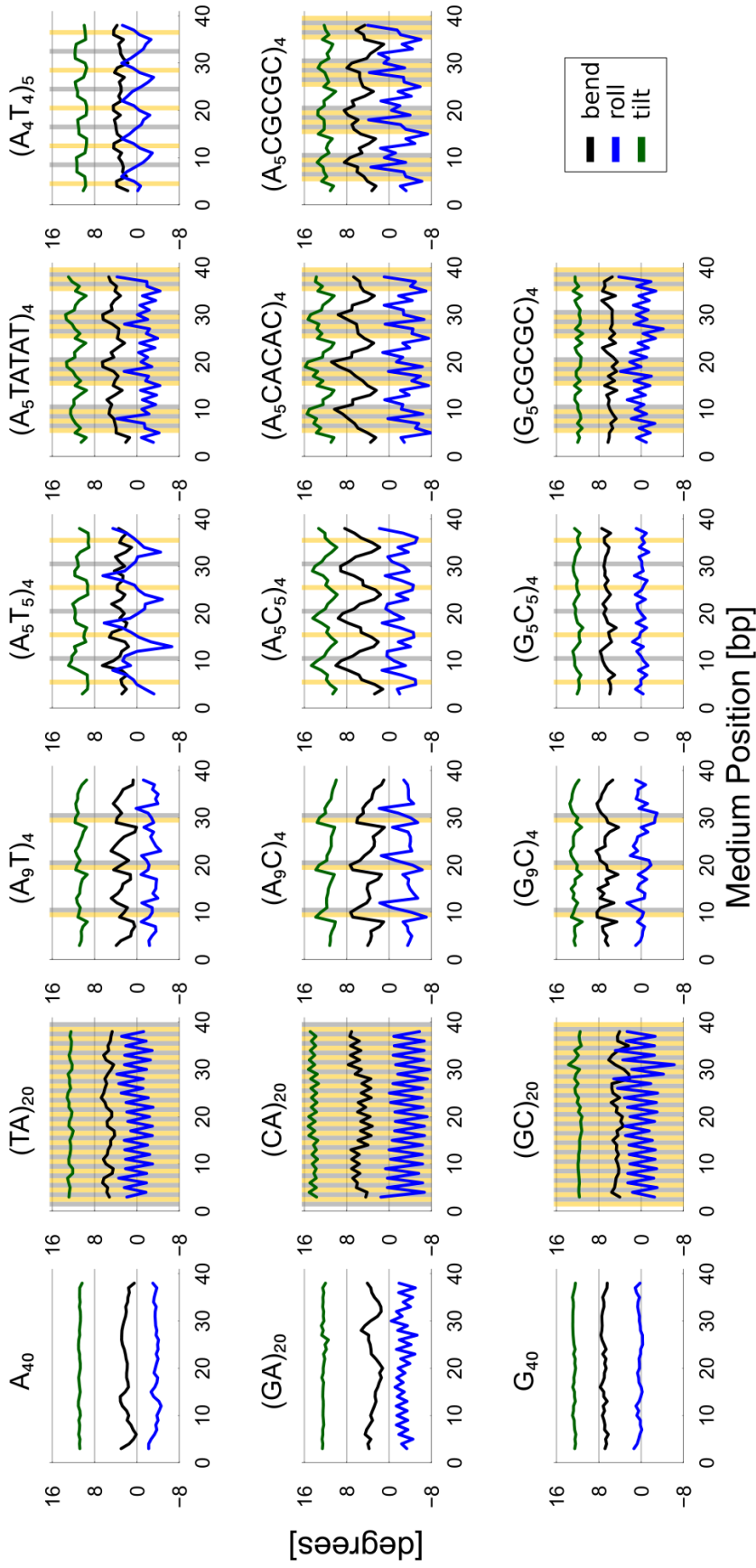


Figure S 4: Bending profiles along DNA molecules considering sub-fragments of 5 bp in length, together with its two components: towards groove and towards backbone, where values are assigned based on the mid-position of each sub-fragment. Sequence patterns ((R)₄₀, (YR)₂₀, (R₉Y)₄, (R₅Y₅)₄ and (R₅YRYRYR)₄) are organised by columns and base groups (A/T, A/C and G/C) are organised by rows with the exception of (A₅CGCGC)₄ and the out-of-phase periodic sequence (A₄T₄)₅. YR and RY bp steps are shaded in grey and yellow, respectively. Standard deviations have been excluded to aid visualisation of key trends.

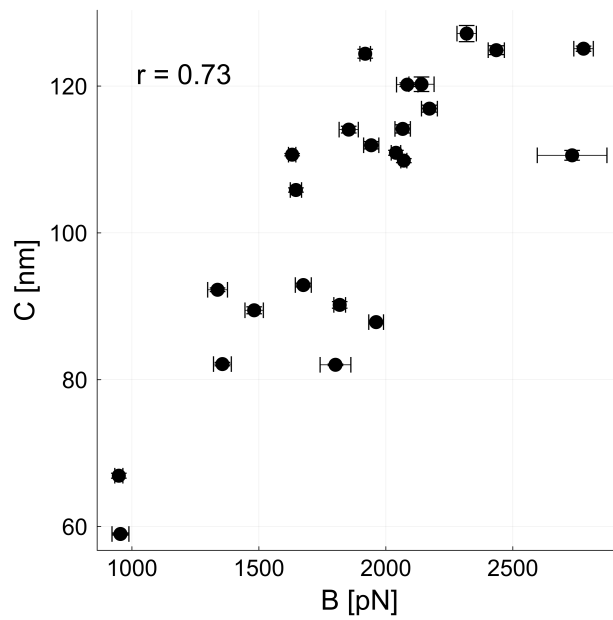


Figure S 5: Correlation between bulk twist (C) and stretch (B) elastic constants for each individual fragment. Error bars for stretching are the 70% confidence intervals of the linear fitting, while twisting are the standard deviation obtained from considering all sub-fragments longer than 10 bp, as it is calculated by SerraNA and described in the Method section. R is the correlation coefficient between the two quantities.

Supplementary Tables

Table S 1: Persistence length A , its static A_s and dynamic A_d components, together with twist C and stretch modulus B for all simulated sequences.

Sequence	A (nm)	A_d (nm)	A_s (nm)	C (nm)	B (pN)
A_{40}	75.0 ± 1.2	76.7 ± 0.9	3549.7 ± 1055.1	120.2 ± 1.0	2140.5 ± 49.5
$(TA)_{20}$	58.4 ± 1.0	60.5 ± 0.5	1672.7 ± 469.3	110.7 ± 0.2	1631.1 ± 14.5
$(A_5TATAT)_4$	69.1 ± 1.4	72.3 ± 1.1	1563.5 ± 311.5	124.4 ± 0.6	1918.7 ± 20.8
$(A_4TATA)_5$	70.9 ± 1.3	73.0 ± 0.9	2459.1 ± 660.7	120.2 ± 0.3	2084.3 ± 42.2
$(A_3TAT)_6A_3T$	65.9 ± 1.2	68.6 ± 0.7	1696.2 ± 392.2	110.9 ± 0.3	2039.8 ± 18.7
$(A_5T_5)_4$	66.984 ± 0.6	77.8 ± 0.6	480.3 ± 38.3	127.1 ± 1.1	2318.3 ± 38.2
$(A_6T_6)_3A_4$	79.487 ± 1.2	85.6 ± 1.1	1119.7 ± 96.5	124.9 ± 0.6	2434.9 ± 31.8
$(A_4T_4)_5$	79.6 ± 1.8	82.8 ± 1.2	2083.5 ± 517.0	125.1 ± 0.4	2778.6 ± 38.4
$(A_9T)_4$	78.6 ± 1.6	80.4 ± 1.3	3421.2 ± 945.4	116.9 ± 0.5	2171.6 ± 31.3
G_{40}	65.1 ± 1.8	68.4 ± 1.1	1332.5 ± 462.0	59.0 ± 0.1	954.8 ± 33.3
$(GC)_{20}$	71.4 ± 1.4	75.4 ± 1.1	1334.6 ± 206.9	110.6 ± 0.7	2733.5 ± 137.2
$(G_5CGCGC)_4$	69.7 ± 1.5	72.8 ± 0.8	1608.9 ± 481.7	87.8 ± 0.1	1961.7 ± 29.0
$(G_5C_5)_4$	65.1 ± 1.2	68.2 ± 0.5	1435.4 ± 434.2	82.1 ± 0.2	1356.3 ± 35.1
$(G_9C)_4$	60.2 ± 1.1	65.4 ± 0.5	760.3 ± 135.4	66.9 ± 0.3	948.4 ± 16.0
$(CA)_{20}$	52.7 ± 1.4	55.4 ± 1.0	1107.9 ± 285.9	89.5 ± 0.5	1481.4 ± 36.2
$(A_5CACAC)_4$	55.8 ± 1.0	70.3 ± 1.4	271.96 ± 20.895	105.8 ± 0.3	1645.9 ± 22.2
$(A_4CACA)_5$	65.5 ± 2.6	74.1 ± 1.2	566.2 ± 133.7	109.8 ± 0.3	2070.8 ± 11.9
$(A_3CAC)_6A_3C$	65.0 ± 2.3	70.9 ± 1.1	775.3 ± 217.0	114.1 ± 0.5	1853.7 ± 37.8
$(A_5C_5)_4$	57.7 ± 0.9	74.1 ± 0.8	260.0 ± 21.0	92.3 ± 0.3	1337.3 ± 39.0
$(A_6C_6)_3A_4$	57.9 ± 0.8	72.2 ± 0.6	292.0 ± 20.1	92.9 ± 0.3	1674.9 ± 31.4
$(A_4C_4)_5$	65.8 ± 1.9	74.5 ± 0.7	565.1 ± 117.3	90.2 ± 0.4	1818.2 ± 23.1
$(A_9C)_4$	65.2 ± 0.7	71.3 ± 0.4	758.6 ± 70.4	111.9 ± 0.5	1942.4 ± 29.9
$(GA)_{20}$	56.6 ± 0.8	58.2 ± 0.6	2128.4 ± 429.2	82.0 ± 0.1	1801.6 ± 60.5
$(A_5CGCGC)_4$	64.6 ± 1.1	77.1 ± 1.3	398.5 ± 34.9	114.2 ± 0.6	2066.2 ± 30.3
Average	65.9 ± 7.3	71.9 ± 7.0	1318.4 ± 897.3	103.7 ± 18.5	1881.9 ± 454.6