Supporting Information

Geggier and Vologodskii 10.1073/pnas.1004809107

SI Text

Sequences of DNA Fragments Used in the Current Study. The sequence names correspond to those mentioned in the text. For each group of fragments, the shown sequence corresponds to the fragment with the highest *j* factor.

SG1 199 bp. AGCTTAGGACTACGAACGCTAGCTTAGCTAC-
CAGCGAGTACACTGCAGCAGCAGCTAGCTAGCGC-
GATGCCCAGCTGAGATCGACGATCGATGGCGATTAT-
CAGCTAGCAGCTAGCGATCGACGCGCGATGCGCAGCT-
GAGCTAGCTGATCAGCTTCAGCTGACGTCAGCTGA-
GAGCTGACCACCGTAGAGTCGATCGA

SG2 200 bp. AGCTTCGCGCGCGTATATATAACACACCAC-
CAACCAGAGAGACTGCAGGAAGAGAGAGAGA-
GAGCGCGCGACACCATATATACCAACCAGAGAGAGA-
GAGAGAGAGCGCGCGCGCGAGCTATATATATATACA-
GAGAGAGAGAGAGAGACGCGCGCGTGCTGCTGCTGC-
TGCTGCTGCTGCTGCGCAGAGAGCGCGAGCA

AC 200 bp, included into elemental set 1. AGCTTCACACACACA-
CACACACACACACACACTGCAGACACACGCACACA-
CACGCACACACACCCACACACACGCACACAACTCACA-
GACACCACGCACACACACATACACACACACTCACACA-
CACACACACCCACACACACTAACACTCACACACACAA-
CACACACCCACACACACACACACACAAACACACA-
CACACCA

AG 199 bp, included into elemental set 1. AGCTTAGAAGGGAAA-
GAAGAGGAGAAGGAGCTGCAGAGGAAAGGGGAGAA-
GAGAAAGAGAGGGAAAAGGGAGGAAAGAAGAGA-
GAAAGGAAAGGAAGGAAGGGAGGGAGGGGAAA-
GAAAGAAGAGGAGAGAGAGAGGGGGAGAGGGAAA-
GAAAGGGAGGAGAGGAAGAGGAGAAGAAGGAAAGA-
GAAGAAGAGGAGGAA

AT 199 bp, included into elemental set 1. AAGCTTATATAATATAA-
TATAATATAACACAACACACTGCAGATATAATATAATA-
TAATATAACATAACATAATATAATATAATATAATATAAGA-
TAAGATAATATAATATAATATAATATAATATAATATAACA-
GAACAGAATATAATATAATATAATATAAGAGAAGAGAA-
TATAATATAATATAATATAAGAGAAG

G 199 bp, included into elemental set 1. AGCTTGGGCTG-
GGCTTGGGCTGGGCTGGGCGGGCTGCAGCTTGGGCT-
GGCTGGGCTTGGGCTGGGCATGGGCTGGGCTTGGGC-
TGGGCATGGGCTGGGCTTGGGCTGGGCTGGG-
CATGGGCTGGGCTGGGCTTGGGCTGGGCTGGGCTGG-
GCATGGGCTGGGCTGGGCTTGGGCTGGGCTGGG-
CATGGGCTTGGGCTGA

GC 198 bp, included into elemental set 1. AGCTTGGCCG-
CGCCGCGGCCGCCGGCGGCTGCAGCCGCGGCGCGG-
CAGCGCCGCGGCCGGTCGCCGGCGCCGG-
CACGGCGCCGGGCCGTGGGCCGCGGCGC-
CAGGGCGCCGGCGCGTGCGGCGGCCGCG-
CAGCCGGCCGCGCGCTGGCCGCCCGG-
CAGCGCCGCCGCGGCTCGCGGCGCCGCC-
GACCGCGCCGCGGTA

ATC 200 bp, included into elemental set 1. AGCTTATCATCATCAT-
CATCGATCATCATCAATCCTGCAGATCATCATCATCATC-
GATCATCATCATCATCAATCATCATCATCATCGATCAT-
CATCATCATCAATCATCATCATCATCGATCATCATCAT-
CATCAATCATCATCATCATCGATCATCATCATCATCAAT-
CATCATCATCATCGATCATCATCATCATCA

AGC 199 bp, included into elemental sets 1 and 2. AGCTTAGCAG-
CAGCAGCAGCTAGCAGCAGCAGCCTGCAGAGCAG-
CAGCAGCAGCTAGCAGCAGCAGCAGCAAGCAGCAG-
CAGCAGCGAGCAGCAGCAGCAGCTAGCAGCAGCAG-
CAGCAAGCAGCAGCAGCAGCGAGCAGCAGCAGCAGC-
TAGCAGCAGCAGCAGCAAGCAGCAGCAGCAGCGAG-
CAGCAGCAGCAGCA

CAA 200 bp, included into elemental sets 1 and 2. AGCTTACAA-
CAACAACAACCTGCAGAACCAACAACAAGCACAACAA-
CAACAACAACAAGAACAACAACAACAACAACCAACAA-
CAACAACAACAACCAACAACAACAAGCAACAACAA-
CAACAACAACCAACAACAACAACCAACAACAACAAC-
CAACAACAACAACAGCAACAACAACAACAACAAGAA-
CAACAAGAAA

ACAT 201 bp, included into elemental set 2. AGCTTACACATATA-
TACACATACATATACACACACATATACTGCAGACATACA-
CATATATACACATACATACACACATACATATATATACACA-
CACATATACATACATATACATACATACATATATACACACA-
TACATACACATATATATACACATACACATACATACACA-
CATATACATATACATACACATATACACATA

ACCAGG 201 bp, included into elemental set 2. AGCTTACCAG-
GAGGACCACCAGGACCACCACCAGGAGGAGCTGCA-
GACCAGGACCACCAGGAGGACCAGGAGGAGGACCAC-
CACCAGGACCACCAGGACCAGGAGGAGGACCACCAG-
GACCACCAGGAGGAGGACCACCAGGACCAGGAGGAC-
CACCACCAGGAGGAGGACCACCAGGACCAGGAGGAC-
CAGGACCACCAA

ACGAGC 199 bp, included into elemental set 2. AGCTTAGCAC-
GACGAGCAGCACGAGCAGCAGCACGACGCTGCAGAG-
CAGCACGACGACGAGCACGAGCAGCACGACGAGCAC-
GAGCAGCAGCACGAGCACGACGAGCAGCACGACGAG-
CAGCACGACGAGCACGACGACGAGCAGCAGCACGAG-
CAGCACGAGCACGACGACGAGCACGAGCAGCACGAG-
CAGCACGAA

AGAT 200 bp, included into elemental set 2. AGCTTAGAGATATA-
TAGAGATAGATATAGAGAGAGATATCTGCAGAGATA-
GAGATATATAGAGATAGATAGAGAGATAGATATATATA-
GAGAGAGATATAGATAGATATAGATAGATAGATATATA-
GAGAGATAGATAGAGATATATATAGAGATAGAGATAGA-
TAGAGAGATATAGATATAGATAGAGATATAGAGATA

CAACTT 198 bp, included into elemental set 2. AGCTT-
CAACTTCTTCAACAACAACTTCTTCTTCAACTCTGCAG-
CAACTTCTTCAACAACTTCAACTTCTTCAACAA-
CAACTTCTTCTTCAACAACTTCAACAACTTCTT-
CAACTTCAACAACTTCTTCAACTTCAACTTCTTCAA-
CAACAACTTCTTCAACTTCTTCAACAACTTCAACTT-
CAACAACTTCTTCAACA

CAGT 200 bp, included into elemental set 2. AGCTTCAGTCAGTCAGTCTGACAGTCAGTCAGTCAGTCAGTCAGTCTGCAGCAGTCAGTCAGTCAGTCAGTCAGTCAGTCTGACAGTCAGTCAGTCAGTCAGTCAGTCAGTCTGACAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCTGACAGTCAGTCAGTCAGTCAGTCTGACAGTCAGTCAGTCAGTA

CATCTA 200 bp, included into elemental set 2. AGCTTCATCTACTACATCATCATCTACTACATCTACATCCTGCAGCATCATCTACTACATCTACTACTACATCATCATCTACATCTACTACATCATCTACTACATCATCTACTACATCATCTACATCTACATCTACTACATCATCTACTACATCATCATCTACTACATCTACTACATCTACATCATCTACTACATCATCA

HPL1 198 bp, included into elemental set 2. AGCTTCGATTGCGCATTCGATTGGAGTCTCCGCTGCCATTGCATTCTGCAGCGATTCGGCATTCGATTCGCGCATTCGATTCGCATTCGATTCGGCATTCGATTCGGCATTCGATTCGCATTCGATTCATTCATTCGATTCGGCATTCGATTCGGCATTCGATTGCGATTGCATTCGGCATTCGATTCGGCGATTCAA

LPL1 200 bp, included into elemental set 2. AGCTTTAGTAGCCTAGTAGCCTAGAGTCTCCGCTGCCATTGCCCTACCTGCAGTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGCCTAGTAGACTAA

HPL2 198 bp. AGCTTACGACGAACGACGACGAACGACGAACGAACGACGAACGCTGCAGACGACGAACGACGAACGACGACGAACGAACGACGACGAACGACGAACGACGACGAACGACGACGAACGACGAGACGAACGAACGACGACGAACGACGAACGACGACGAACGACGAACGACGAACGACGACGAACGAACGACGAACGACA

LPL2 201 bp. AGCTTGCATAGGCATTAGCCATGCATAGGCATATGGCATTAGGCACTGCAGGGCCATAGGCATGCATAGGCATAGGCCATGGCATAGGCATTAGGCATGCATAGGCATAGGCATGGCATAGGCATTAGGCATGCATAGGCATAGGCATGGCATAGGCATTAGGCATGCATAGGCATGGCATAGGCCATGGCATAGGCATTA

SI Materials and Methods. ***Accuracy of Shimada-Yamakawa (SY) approximation for the j factors of the worm-like chain (WLC).*** It was found earlier, by comparing the SY solution with Monte Carlo calculations, that the SY approximation gives very accurate results for j factors of WLC fragments corresponding to DNA molecules of 200–600 bp in length (1, 2). We investigated the approximation accuracy for WLC fragments with lengths between a and $1.6a$ using the Monte Carlo algorithm based on a string of conditional probabilities (3). The algorithm allows performing the calculations with very low statistical error even for very short WLC fragments. Specifically, we compared the values of j factors obtained from the SY equation and calculated by MC simulation for the discrete worm-like chain (DWLC). We found that, in the limit of very short segments, when DWLC becomes very close to WLC, MC results approach the values obtained from SY equation (Fig. S6). Although this limit is not reached for the segment length of 0.34 nm (the distance between adjacent base pairs), we decided to ignore this small discrepancy. We found that the difference between the j factors obtained from SY equation and the computed values disappears if the bending rigidity constant, g, in the DWLC is specified as $kT(a-l)/l$ rather than kTa/l (Fig. S6). Thus, if we fit the experimental data by the equation for the DWLC, the value of a would be increased by l, so the correction would be within the error bar of a determination in any case.

Torsional component of j factor. The second component of j factor, j_{tw}, was calculated as

$$j_{tw} = \frac{1}{\sqrt{2\pi\langle(\Delta Lk)^2\rangle}} \sum_{Lk} \exp\left(-\frac{(Lk - N/\gamma)^2}{2\langle(\Delta Lk)^2\rangle}\right), \qquad \textbf{[S1]}$$

where $\langle(\Delta Lk)^2\rangle$ is the variance of the linking number distribution for the circular form of the fragments, Lk is the linking number of the complementary strands, γ is the average helical repeat, and N is the number of base pairs in the fragment. For DNA fragments with length close to 200 bp only one or two terms in the sum of Eq. **S1** make a notable contribution into j_{tw}. The corresponding values of Lk are specified by condition $|Lk - N/\gamma| < 1$. $\langle(\Delta Lk)^2\rangle$ was estimated as sum of twist and writhe variances, $\langle(\Delta Tw)^2\rangle$ and $\langle(Wr)^2\rangle$ (see ref. 4). The contribution from torsional fluctuations was calculated as

$$\langle(\Delta Tw)^2\rangle = \frac{kTL}{4\pi^2 C}, \qquad \textbf{[S2]}$$

where C is the fragment torsional rigidity, in erg · cm, and L are measured in centimeters. This contribution for fragments of 200 bp constituted about 94% of $\langle(\Delta Lk)^2\rangle$. The contribution from bending fluctuations, $\langle(Wr)^2\rangle$, is specified by L and a and was calculated with interpolating equations suggested in ref. 5.

1. Levene SD, Crothers DM (1986) Ring closure probabilities for DNA fragments by Monte Carlo simulation. *J Mol Biol* 189:61–72.
2. Hagerman PJ, Ramadevi VA (1990) Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure. I. Computational analysis. *J Mol Biol* 212:351–362.
3. Podtelezhnikov AA, Mao C, Seeman NC, Vologodskii AV (2000) Multimerization-cyclization of DNA fragments as a method of conformational analysis. *Biophys J* 79:2692–2704.
4. Vologodskii AV, Frank-Kamenetskii MD (1981) Premelting of superhelical DNA: an expression for superhelical energy. *FEBS Lett* 131:178–180.
5. Klenin KV, et al. (1989) Variance of writhe for wormlike DNA rings with excluded volume. *J Biomol Struct Dyn* 6:707–714.
6. Marini JC, Levene SD, Crothers DM, Englund PT (1982) Bent helical structure in kinetoplast DNA. *Proc Natl Acad Sci USA* 79:7664–7668.
7. Koo HS, Wu HM, Crothers DM (1986) DNA bending at adenine-thymine tracts. *Nature* 320:501–506.
8. Hagerman PJ (1986) Sequence-directed curvature of DNA. *Nature* 321:449–450.
9. Hagerman PJ (1990) Sequence-directed curvature of DNA. *Annu Rev Biochem* 59:755–781.
10. Du Q, Vologodskaia M, Kuhn H, Frank-Kamenetskii M, Vologodskii A (2005) Gapped DNA and cyclization of short DNA fragments. *Biophys J* 88:4137–4145.
11. Shimada J, Yamakawa H (1984) Ring-closure probabilities for twisted wormlike chains. Application to DNA. *Macromolecules* 17:689–698.
12. Licinio P, Guerra JC (2007) Irreducible representation for nucleotide sequence physical properties and self-consistency of nearest-neighbor dimer sets. *Biophys J* 92:2000–2006.
13. Balasubramanian S, Xu F, Olson WK (2009) DNA sequence-directed organization of chromatin: Structure-based computational analysis of nucleosome-binding sequences. *Biophys J* 96:2245–2260.

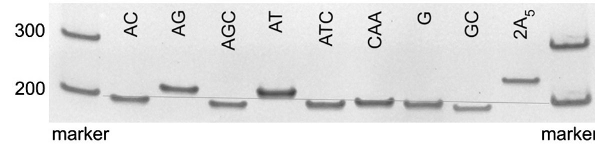

Fig. S1. The test of possible intrinsic curvature for the fragments of elemental set 1 by measuring their mobility in polyacrylamide gel (6–8). The major sequence motifs of the fragments are shown at the top of the gel. Mobility of the fragments in the 4–12% gradient polyacrylamide gel should be compared with the mobility of the marker and 200-bp fragment with two AAAAA ($2A_5$) tracts in the middle of the fragment, separated by 5 bp. The total intrinsic bend in the latter fragment is close to 30° (7, 9, 10). For all fragments of the set, the mobility variations are much smaller than the mobility reduction for the fragment containing A tracts, showing that the fragments do not have essential intrinsic curvature.

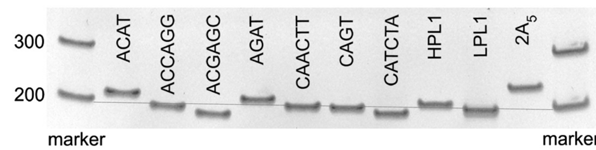

Fig. S2. The test of possible intrinsic curvature for the fragments of elemental set 2. For all fragments of the set the mobility variations are much smaller than the mobility reduction for the fragment containing A tracts, showing that the fragments of this set do not have essential intrinsic curvature.

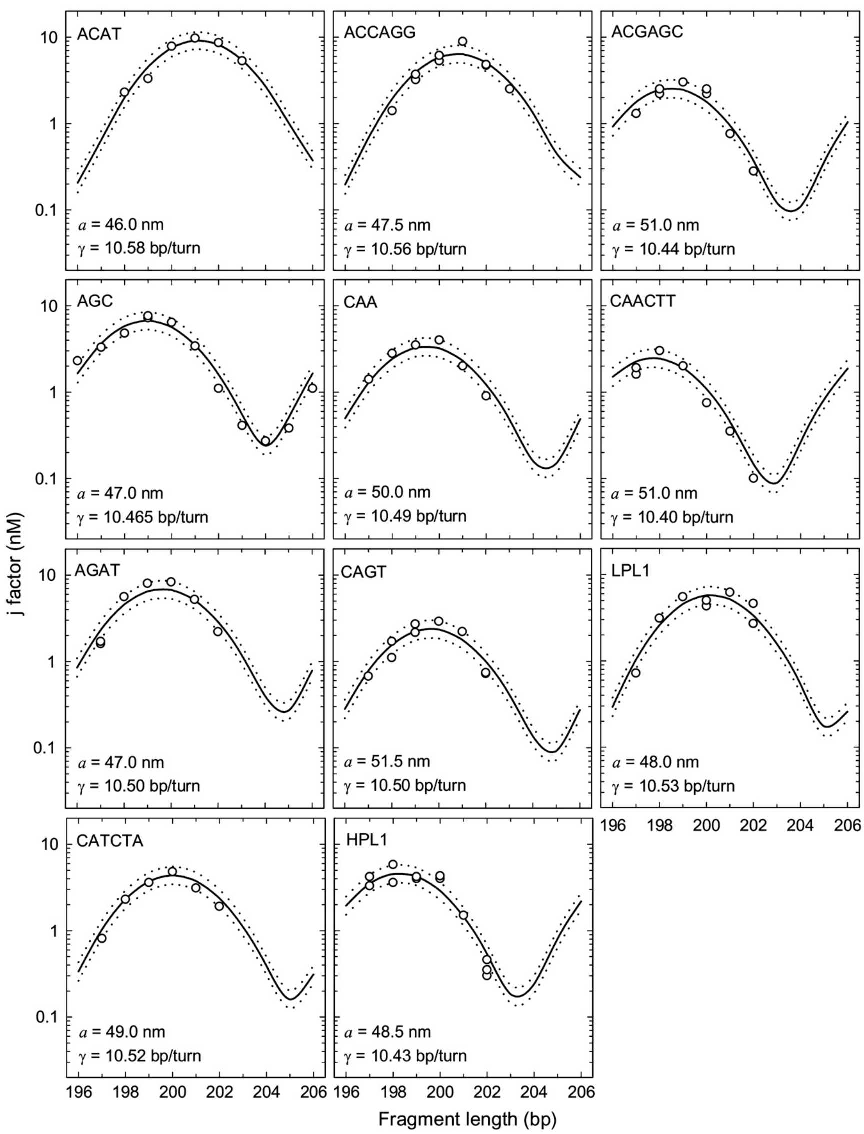

Fig. S3. The measured values of j factors for 11 fragments of elemental set 2 and their theoretical fit. The values of a and γ corresponding to the best fit are shown inside each plot. The theoretical curves corresponding to the best fit are shown by the solid lines; the dotted lines correspond to values of a differed from the best fit by ± 1 nm. The value of C used in the fitting procedure equals 2.6×10^{-19} erg · cm for all fragments. This value of C is the average over a few fragments used in the current study.

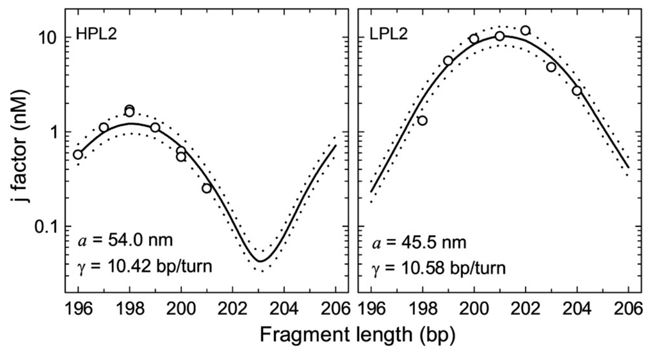

Fig. S4. Test of the obtained set of bending rigidity parameters. The measured values of j factors for the fragments with predicted very high (HPL) and very low (LPL) persistence length were fitted by the theoretical equation to obtain the values of a and γ corresponding to the experimental data. The value of C used in the fitting procedure equals 2.6×10^{-19} erg · cm for both fragments.

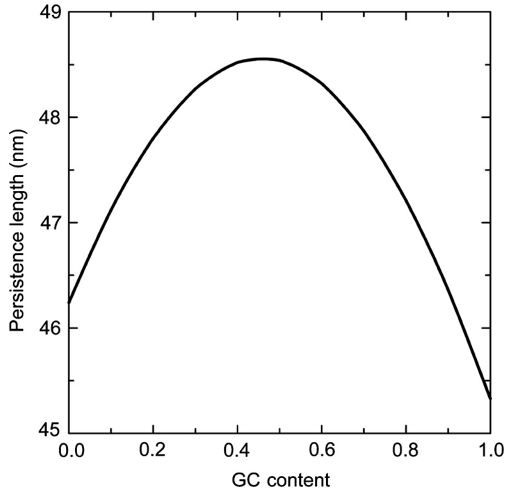

Fig. S5. Calculated average persistence length of DNA as a function of its GC content. The shown values of a are based on the data presented in Table 2 and correspond to averaging over all sequences with a given GC content.

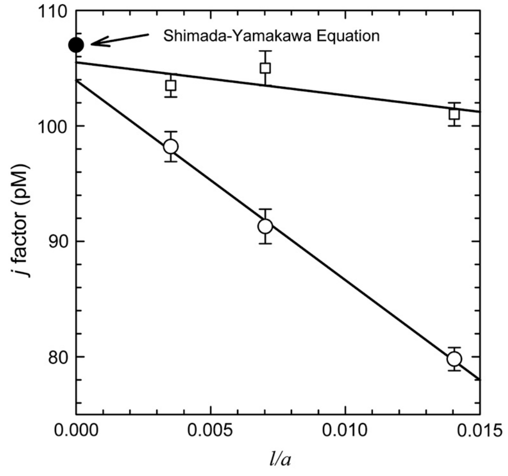

Fig. S6. Comparison between SY approximation for the j factor of the worm-like chain (Eqs. **45** and **74** of ref. 11) and Monte Carlo calculation for the discrete worm-like chain. The value of g in the simulation was equal to kTa/l (○) and $kT(a-l)/l$ (□). If l corresponds to the length of one base pair step and a equals 48.416 nm, then l/a equals 0.007. The calculation was performed for DNA fragment of 51 nm in length.

Table S1. The values of invariants of DNA persistence lengths, a, and helical repeats, γ, obtained from elemental set 1; these values did not sustain the test of their validity

Invariant combination of a^{XY} or γ^{XY}; symbol r corresponds to a or γ, correspondingly	Persistence length, a, nm	Helical repeat, γ, bp/(helix turn)
$r^{AA/TT}$	56.6	10.41
$r^{CC/GG}$	44.4	10.49
$2/(1/r^{AT}$	37.3	10.46
$2/(1/r^{GC}$	52.7	10.38
$2/(1/r^{AC/GT} + 1/r^{CA/TG})$	48.4	10.54
$2/(1/r^{AG/CT} + 1/r^{GA/TC})$	45.3	10.50
$3/(1/r^{AT} + 1/r^{GA/TC} + 1/r^{CA/TG})$	53.1	10.54
$3/(1/r^{AC/GT} + 1/r^{CG} + 1/r^{GA/TC})$	49.2	10.47

Table S2. Persistence lengths and helical repeats corresponding to 10 dinucleotide steps

Dinucleotide step	Persistence length, a, nm	Helical repeat, bp/(helix turn)
AA/TT	50.4	10.27
AC/GT	55.4	10.58
AG/CT	51.0	10.42
AT	40.9	10.59
CA/TG	46.7	10.60
CC/GG	41.7	10.76
CG	56.0	10.35
GA/TC	54.4	10.36
GC	44.6	10.43
TA	44.7	10.65

The values of 10 a^{XY} and γ^{XY} could not be determined unambiguously from the type of data obtained in this study and are shown for convenience of calculation of a and γ for a particular DNA fragment. The shown values are based on the treatment suggested in ref. 12.

Table S3. Comparison of the values of persistence length, a, and DNA helical repeat, γ, predicted from the statistical analysis of DNA–protein crystals (13) with the results of the current study

DNA fragment	a obtained from ref. 13, nm	a, current study, nm	γ obtained from ref. 13, bp/turn	γ, current study, bp/turn
ACAT	48.0	46.0	10.68	10.58
ACCAGG	49.1	47.5	10.76	10.56
ACGAGC	48.6	51.0	10.66	10.44
AGAT	54.1	47.0	10.65	10.50
AGC	50.1	47.0	10.62	10.465
CAA	50.2	50.0	10.56	10.49
CAACTT	54.4	51.0	10.53	10.40
CAGT	56.4	51.5	10.65	10.50
CATCTA	52.4	49.0	10.66	10.52
HPL1	47.5	48.5	10.62	10.43
LPL1	49.7	48.0	10.67	10.53
HPL2	47.5	54.0	10.64	10.42
LPL2	46.1	45.5	10.70	10.58

The persistence lengths and corresponding 10 dinucleotide steps were calculated as $a = R/(d_t^2 + d_r^2)$, where d_t and d_r are the dispersions of tilt and roll for a dinucleotide step of B DNA taken from table S3 of ref. 13, and R is a constant adjusted so that the average persistence length was equal to 48.5 nm. The obtained values and the values of twist angles from ref. 13 were then used to predict a and γ of all fragments from elemental set 2.