Electronic Supplementary Information

Sequence-dependent twist-bend coupling in DNA minicircles

Minjung Kim^a, Sehui Bae^a, Inrok Oh^b, Jejoong Yoo^c, and Jun Soo Kim^{a*}

^aDepartment of Chemistry and Nanoscience, Ewha Womans University, Seoul 03760, Republic of Korea

^bLG Chem Ltd, LG Science Park, Seoul 07796, Republic of Korea

^cDepartment of Physics, Sungkyunkwan University, Suwon 16419, Republic of Korea

*Corresponding author: jkim@ewha.ac.kr

Contents

Table S1. A summary of Gaussian fitting parameters for twist-angle distribution functions of Fig. 3

Table S2. Average twist angle of circular and linear DNAs

Table S3. List of PDB files of nucleosomal DNAs from the Protein Data Bank

Table S4. A summary of circular DNAs with constituent dinucleotide steps and the number of base pairs

Figure S1. Determination of kink formation based on the hydrogen bond distance between complementary bases in DNA minicircles

Figure S2. Ellipticity of DNA minicircles as a function of time

Figure S3. Time evolution of twist angles of all nucleotide steps in a linear DNA molecule

Figure S4. Time evolution of DNA poloidal orientation as a function of time

Figure S5. Time correlation functions of DNA poloidal orientations

Dinucleotide Steps	μ^{\dagger}	σ^{\dagger}	<i>R</i> ² †	μ^{\ddagger}	σ^{\ddagger}	<i>R</i> ² ‡
AT/AT	30.86	3.44	0.9994	29.86	3.37	0.9999
TA/TA	35.42	5.38	0.9944	29.90	4.46	0.9947
GC/GC	32.73	4.55	0.9996	31.77	4.31	0.9989
CG/CG	37.20	4.72	0.9989	34.11	4.83	0.9997
AC/GT	31.61	4.07	0.9997	30.55	3.85	0.9997
CA/TG	35.19	4.81	0.9993	31.06	4.73	0.9984
AG/CT	30.19	4.14	0.9955	30.38	4.49	0.9956
GA/TC	34.55	4.84	0.9987	30.69	4.94	0.9978
AA/TT	33.75	4.20	0.9991	33.32	4.60	0.9968
GG/CC	30.23	4.27	0.9975	28.85	3.99	0.9976

Table S1. A summary of Gaussian fitting parameters for twist-angledistribution functions of Fig. 3

The parameters μ , σ are the mean, standard deviation of the Gaussian distribution function, respectively. R^2 represents the coefficient of determination in fitting. [†]Dinucleotide step with the direction of the minor groove facing inward (cos θ <

0).

[‡]Dinucleotide step with the direction of the minor groove facing outward ($\cos \theta > 0$).

Sequence	$<\Omega^{XY}>^{\dagger}$	$<\Omega^{YX}>^{\ddagger}$	Sequence	$<\Omega_0^{XY}>^\dagger$	$<\Omega_0^{YX}>^{\ddagger}$
(AT)45/(AT)45	30.22 (3.51)	32.84 (5.62)	$(AT)_{12}/(AT)_{12}$	30.32 (3.64)	32.99 (4.99)
$(GC)_{42}/(GC)_{42}$	32.28 (4.40)	35.39 (5.07)	$(GC)_{12}/(GC)_{12}$	31.64 (4.23)	35.30 (4.37)
(AC) ₄₅ /(GT) ₄₅	30.49 (4.09)	32.29 (5.18)	$(AC)_{12}/(GT)_{12}$	30.87 (3.93)	33.05 (4.47)
(AG) ₄₅ /(CT) ₄₅	30.57 (4.43)	32.59 (5.10)	$(AG)_{12}/(CT)_{12}$	30.80 (4.47)	32.54 (4.70)
$(AA)_{42}/(TT)_{42}$	33.62 (4.37)		$(AA)_{12}/(TT)_{12}$	33.04 (4.26)	
(GG) ₄₈ /(CC) ₄₈	29.67 (4.32)		$(CC)_{12}/(GG)_{12}$	29.35 (3.92)	

Table S2. Average twist angle of circular and linear DNAs

 $<\Omega>$ is the average twist angle in each DNA minicircle, whereas $<\Omega_0>$ is the average twist angle in a linear DNA molecule. The values in parenthesis are the standard deviations.

[†]XY represents AT/AT in the cases of sequence $(AT)_{45}/(AT)_{45}$ and $(AT)_{12}/(AT)_{12}$, GC/GC in the cases of $(GC)_{42}/(GC)_{42}$ and $(GC)_{12}/(GC)_{12}$, AC/GT in the cases of $(AC)_{45}/(GT)_{45}$ and $(AC)_{12}/(GT)_{12}$, and AG/CT in the cases of $(AG)_{45}/(CT)_{45}$ and $(AG)_{12}/(CT)_{12}$.

[‡]YX represents TA/TA in the cases of sequence $(AT)_{45}/(AT)_{45}$ and $(AT)_{12}/(AT)_{12}$, CG/CG in the cases of $(GC)_{42}/(GC)_{42}$ and $(GC)_{12}/(GC)_{12}$, CA/TG in the cases of $(AC)_{45}/(GT)_{45}$ and $(AC)_{12}/(GT)_{12}$, and GA/TC in the cases of $(AG)_{45}/(CT)_{45}$ and $(AG)_{12}/(CT)_{12}$.

Table S3. List of PDB files of nucleosomal DNAs from the Protein DataBank

1AOI	3A6N	3MGS	4KGC	5B32	5XF6	6IY3	6NE3	6SEE	6Y5E
1EQZ	3AFA	3MNN	4KUD	5B33	5XM0	6J99	6NJ9	6SEF	6YOV
1F66	3AN2	3MVD	4LD9	5B40	5XM1	6JM9	6NN6	6SEG	6Z6P
1ID3	3AV1	3062	4R8P	5CP6	5Y0C	6JMA	6NOG	6T79	6ZHX
1KX3	3AV2	3REH	4WU8	5CPI	5Y0D	6JOU	6NQA	6T7A	6ZHY
1KX4	3AYW	3REI	4WU9	5CPJ	5Z23	6JR0	601D	6T7B	7A08
1KX5	3AZE	3REJ	4X23	5CPK	5Z30	6JR1	6096	6T7C	7AT8
1M18	3AZF	3REK	4XUJ	5DNM	5Z3L	6JXD	60M3	6T7D	7BXT
1M19	3AZG	3REL	4XZQ	5DNN	5Z3O	6JYL	6PA7	6T90	7BY0
1M1A	3AZH	3TU4	4YM5	5E5A	5Z3U	6K1I	6PWE	6T93	7C0M
1P34	3AZI	3UT9	4YM6	5F99	5Z3V	6K1J	6PWF	6T9L	7CCQ
1P3A	3AZJ	3UTA	4YS3	5GSU	5ZBX	6K1K	6PWV	6TEM	7CCR
1P3B	3AZK	3UTB	4Z5T	5GT0	6BUZ	6K1P	6PWW	6UGM	7D1Z
1P3F	3AZL	3W96	4Z66	5GT3	6C0W	6KE9	6PWX	6UH5	7D20
1P3G	3AZM	3W97	4ZUX	5GTC	6DZT	6KIU	6PX1	6UPH	7DBP
1P3I	3AZN	3W98	5AV5	5GXQ	6E0C	6KIV	6R0C	6UPK	7JO9
1P3K	3B6F	3W99	5AV6	5KGF	6E0P	6KIW	6R1T	6UPL	7JOA
1P3L	3B6G	3WA9	5AV8	5MLU	6ESF	6KIX	6R1U	6USJ	7JZV
1P3M	3C1B	3WAA	5AV9	5NL0	6ESG	6KIZ	6R25	6V2K	7K5X
1P3O	3C1C	3WKJ	5AVB	50MX	6ESH	6KVD	6R8Y	6V92	7K5Y
1P3P	3KUY	3WTP	5AVC	50NG	6ESI	6KW3	6R8Z	6VEN	7K60
1S32	3KWQ	3X1S	5AY8	50NW	6FQ5	6KW5	6R90	6VYP	7K61
1U35	3KXB	3X1T	5B0Y	5T5K	6FQ6	6KXV	6R91	6VZ4	7K63
1ZLA	3LEL	3X1U	5B0Z	5WCU	6FQ8	6L9H	6R92	6WZ5	7K6P
2CV5	3LJA	3X1V	5B1L	5X0X	6GEJ	6LE9	6R94	6WZ9	7K6Q
2F8N	3LZ0	4J8U	5B1M	5X0Y	6GEN	6M4D	6RYR	6X0N	7KTQ
2FJ7	3LZ1	4J8V	5B24	5X7X	6IPU	6M4G	6RYU	6X59	
2NQB	3MGP	4J8W	5B2I	5XF3	6IQ4	6M4H	6S01	6X5A	
2NZD	3MGQ	4J8X	5B2J	5XF4	6IRO	6MUO	6SE0	6XJD	
2PYO	3MGR	4JJN	5B31	5XF5	6IY2	6MUP	6SE6	6Y5D	

We have downloaded all PDB files obtained from the Protein Data Bank, using a search keyword of nucleosome. A subset of these structures was selected based on the following criteria: (i) mono-nucleosomal DNA, (ii) the absence of additional proteins other than histone proteins, and (iii) the compatibility with the analysis software Curves+. For nucleosomes with DNA of more than 147 base pairs, the base pairs in the linker region were removed.

$<\!\Omega_{out}^{YX}\!>^{\sharp}$	31.38	30.25	29.60	28.61	30.46	33.95	33.61	32.33	31.65
	(0.01)	(0.01)	(0.05)	(0.01)	(0.06)	(0.08)	(0.20)	(0.14)	(0.14)
$< \Omega_{in}^{YX} > \ddagger$	36.81	35.63	34.26	33.58	34.02	36.91	35.80	35.04	34.24
	(0.01)	(0.02)	(0.06)	(0.02)	(0.05)	(0.01)	(0.05)	(0.05)	(0.08)
$<\Omega_{out}^{XY}>^{\dagger}$	30.14	29.77	29.45	28.77	29.63	31.85	31.44	30.77	30.25
	(0.01)	(0.01)	(0.03)	(0.01)	(0.03)	(0.03)	(0.05)	(0.05)	(0.05)
$<\Omega_{in}^{XY}>^{\dagger}$	30.99	30.70	30.41	29.96	30.08	32.73	32.05	31.72	31.30
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.03)	(0.03)	(0.04)
$* <_{XX} U>$	34.00	32.84	31.80	31.01	32.14	35.39	34.69	33.66	32.93
	(0.01)	(0.01)	(0.01)	(0.01)	(0.05)	(0.04)	(0.08)	(0.08)	(0.11)
$\langle \Omega^{XY} \rangle^{\dagger}$	30.55	30.22	29.90	29.34	29.85	32.28	31.74	31.24	30.77
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)
<0>	32.27	31.53	30.86	30.18	30.98	33.83	33.22	32.44	31.84
	(0.00)	(0.00)	(0.01)	(0.01)	(0.03)	(0.02)	(0.05)	(0.06)	(0.07)
ellipticity	1.14 (0.00)	1.15 (0.00)	1.21 (0.01)	1.17 (0.00)	1.50 (0.01)	1.20 (0.02)	1.59 (0.10)	1.29 (0.04)	1.22 (0.01)
kink	0	0	1	0	1	1	1	7	7
α	0.034	0.011	-0.011	-0.032	-0.052	0.024	0.001	-0.022	-0.044
L_{k_0}	7.74 (0.01)	7.91 (0.01)	8.09 (0.01)	8.26 (0.01)	8.44 (0.01)	7.81 (0.00)	8.00 (0.00)	8.18 (0.00)	8.37 (0.00)
N_{bp}	88	06	92	94	96	84	86	88	06
Seq	AT/AT	AT/AT	AT/AT	AT/AT	AT/AT	GC/GC	GC/GC	GC/GC	GC/GC

Table S4. A summary of circular DNAs with constituent dinucleotide stepsand the number of base pairs

Seq	N_{bp}	L_{k_0}	Q	kink	ellipticity	<u></u>	$<\Omega^{XY}>^{\dagger}$	$<\Omega^{YX}>^{\sharp}$	$<\Omega_{in}^{XY}>^{\dagger}$	$<\Omega_{out}^{XY}>^{\dagger}$	$<\Omega^{YX}_{in}>^{\ddagger}$	$<\Omega_{out}^{YX}>^{\pm}$
AC/GT	88	7.81 (0.01)	0.024	0	1.15 (0.00)	32.11 (0.00)	31.05 (0.01)	33.17 (0.01)	31.57 (0.02)	30.54 (0.01)	35.27 (0.02)	31.16 (0.01)
AC/GT	06	7.99 (0.01)	0.001	0	1.16 (0.00)	31.39 (0.00)	30.49 (0.01)	32.29 (0.01)	31.06 (0.01)	29.95 (0.01)	34.42 (0.02)	30.26 (0.02)
AG/CT	88	7.74 (0.01)	0.033	0	1.14 (0.00)	32.20 (0.00)	31.22 (0.01)	33.19 (0.01)	30.85 (0.01)	31.59 (0.01)	34.80 (0.01)	31.59 (0.01)
AG/CT	06	7.92 (0.01)	0.011	0	1.14 (0.00)	31.58 (0.01)	30.57 (0.01)	32.59 (0.01)	30.50 (0.01)	30.63 (0.01)	34.38 (0.02)	30.86 (0.01)

Table S4 - continued from previous page

Seq	N _{bp}	L_{k_0}	σ	kink	ellipticity	<Ω>	$<\Omega_{in}>$	$<\Omega_{out}>$
	84	7.71	0.038	1	1.13	33.62	33.81	33.44
	0-	(0.01)	0.038		(0.00)	(0.00)	(0.01)	(0.01)
	86	7.89	0.014	2	1.18	33.13	33.45	32.84
AA/11	00	(0.01)	0.014	5	(0.01)	(0.02)	(0.01)	(0.03)
AA/TT	00	8.08	0.010	2	1.15	32.36	33.05	31.72
	00	(0.01)	-0.010	3	(0.01)	(0.02)	(0.02)	(0.04)
	00	8.26	0.032	3	1.30	32.04	32.43	31.74
AA/11	90	(0.01)	-0.032		(0.01)	(0.01)	(0.04)	(0.05)
GG/CC	88	7.17	0.115	r	2.59	31.14	31.16	31.12
00/00	00	(0.00)	0.115	2	(0.07)	(0.12)	(0.15)	(0.10)
GG/CC	00	7.34	0.090	2	Not	30.96	31.18	30.76
	90	(0.00)	0.090		calculable	(0.03)	(0.04)	(0.02)
GG/CC	94	7.66	0.044	4	1.30	30.10	30.78	29.52
		(0.00)			(0.01)	(0.01)	(0.02)	(0.04)
GG/CC	06	7.83	0.022	Л	1.19	29.67	30.44	29.04
	90	(0.00)	0.022	4	(0.01)	(0.02)	(0.03)	(0.03)

 Table S4 - continued from previous page

 N_{bp} is the number of base pairs. $\langle \Omega \rangle$ is the average twist angle in each DNA minicircle, whereas $\langle \Omega_0 \rangle$ is the average twist angle in a linear DNA molecule with 24 base pairs of the same repeating nucleotides. The linking number L_k is the number of helical turns in the DNA minicircles. The fractional linking number L_{k_0} is the number of helical turns predicted based on the average twist angle $\langle \Omega_0 \rangle$ of the corresponding linear molecules. The superhelical density is defined as $\sigma = (L_k - L_{k_0})/L_{k_0}$ with non-zero values indicating deviation of the DNA minicircles from the ideal structure free from torsional stress. A kink is defined in terms of base-pair dissociation that lasts for some duration in the course of the 1-µs simulation. Ellipticity is measured by fitting the 2D-projection of the helical axis to the equation of an ellipse. The values in parenthesis are the error estimated by batch means method.

[†]XY represents AT/AT, GC/GC, AC/GT, and AG/CT in the case of sequence $(AT)_n/(AT)_n$, $(GC)_n/(GC)_n$, $(AC)_n/(GT)_n$, and $(AG)_n/(CT)_n$, respectively.

[‡]YX represents TA/TA, CG/CG, CA/TG, and GA/TC in the case of sequence $(AT)_n/(AT)_n$, $(GC)_n/(GC)_n$, $(AC)_n/(GT)_n$, and $(AG)_n/(CT)_n$, respectively.



Fig. S1. Determination of kink formation based on the hydrogen bond distance between complementary bases in DNA minicircles. Time evolution of hydrogen bond distances, h, of all base pairs in DNA minicircles. The value of h was calculated as a distance between the N1 atom of adenine and the H3 atom of thymine for adenine-thymine (AT) complementary base pairs and between the H1 atom of guanine and the N3 atom of cytosine for guanine-cytosine (GC) complementary base pairs. The color in the scale bar on the right indicates the value of h in unit of nm. The kink formation is determined by the value of h greater than 0.29 nm.













Fig. S2. Ellipticity of DNA minicircles as a function of time. (a) $(AT)_{45}/(AT)_{45}$, (b) $(GC)_{42}/(GC)_{42}$, (c) $(AC)_{44}/(GT)_{44}$, (d) $(AG)_{45}/(CT)_{45}$, (e) $(AA)_{42}/(TT)_{42}$, and (f) $(GG)_{48}/(CC)_{48}$. Ellipticity is defined as the ratio of the major axis to the minor axis of the ellipse. The ellipse is obtained by fitting the 2D-projection of the helical axis to the equation of an ellipse. The numbers on top of each figure are the simulation time at which the DNA conformation was taken and the ellipticity calculated for the conformation.







Fig. S4. Time evolution of DNA poloidal orientation as a function of time. Values of $\theta(t)$ for each base pair were calculated at each time step and subtracted by the initial value $\theta_0 = \theta(t = 0)$, so that all DNA poloidal orientations start from 0 at the beginning. Then the orientations at each time step were averaged over all base pairs of DNA minicircles, resulting in the time evolution of $\langle \theta - \theta_0 \rangle$ for (a) (AT)₄₅/(AT)₄₅, (b) (GC)₄₂/(GC)₄₂, (c) (AC)₄₄/(GT)₄₄, (d) (AG)₄₅/(CT)₄₅, (e) (AA)₄₂/(TT)₄₂, and (f) (GG)₄₈/(CC)₄₈.



Fig. S5. Time correlation functions (TCF) of DNA poloidal orientations $\langle \theta - \theta_0 \rangle$, as presented in Fig. S4, for sequences of (a) $(AT)_{45}/(AT)_{45}$, (b) $(GC)_{42}/(GC)_{42}$, (c) $(AC)_{44}/(GT)_{44}$, (d) $(AG)_{45}/(CT)_{45}$, and (e) $(AA)_{42}/(TT)_{42}$. Solid lines are time correlation functions and dashed lines represent exponential fittings of the data. We obtained relaxation times of $\tau = 36$, 55, 35, 87, and 67 ns from the exponential fittings of (a), (b), (c), (d), and (e).