

A Predictive Model for the Sequence-Dependent Fluorogenic Response of Forced-Intercalation Peptide Nucleic Acid (FIT-PNA)

Itamar Peled^a and Eylon Yavin^{a,*}

^aThe School of Pharmacy, Faculty of Medicine, The Hebrew University of Jerusalem, Hadassah Ein-Kerem, Jerusalem 91120, Israel.

*E-mail: eylony@ekmd.huji.ac.il; Fax: +972-2-6757574; [Tel: +972-2-6758692](tel:+972-2-6758692).

Tables of contents:

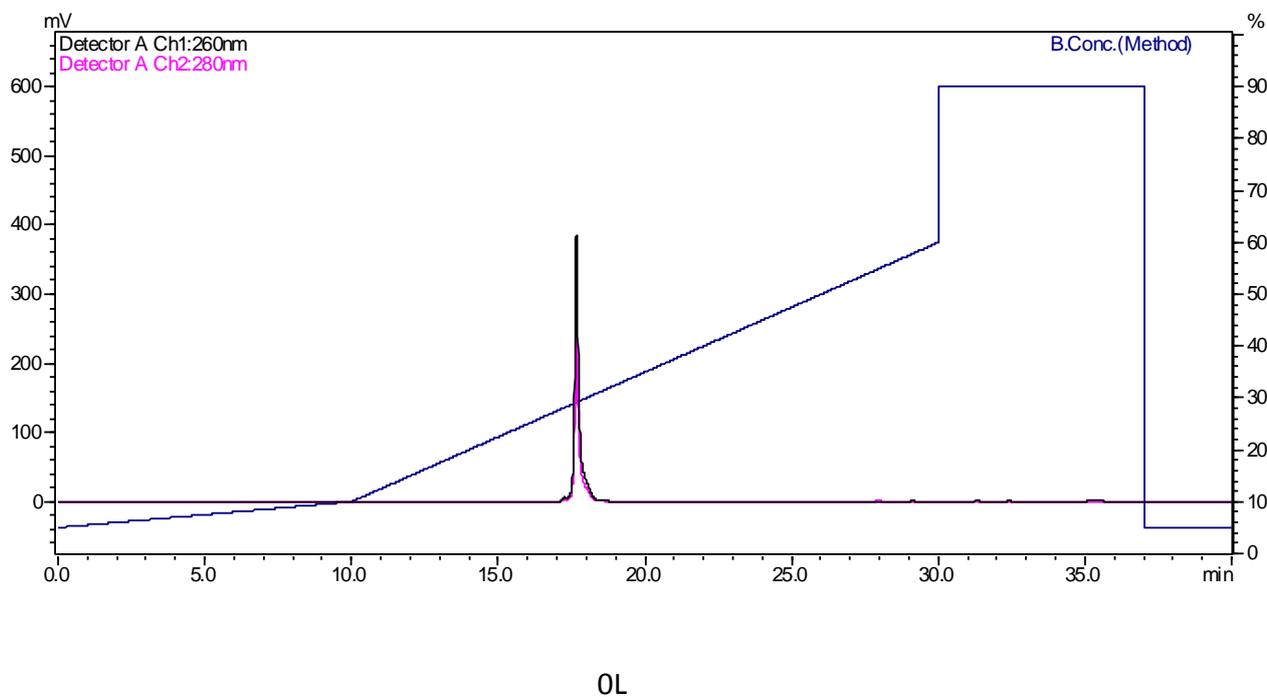
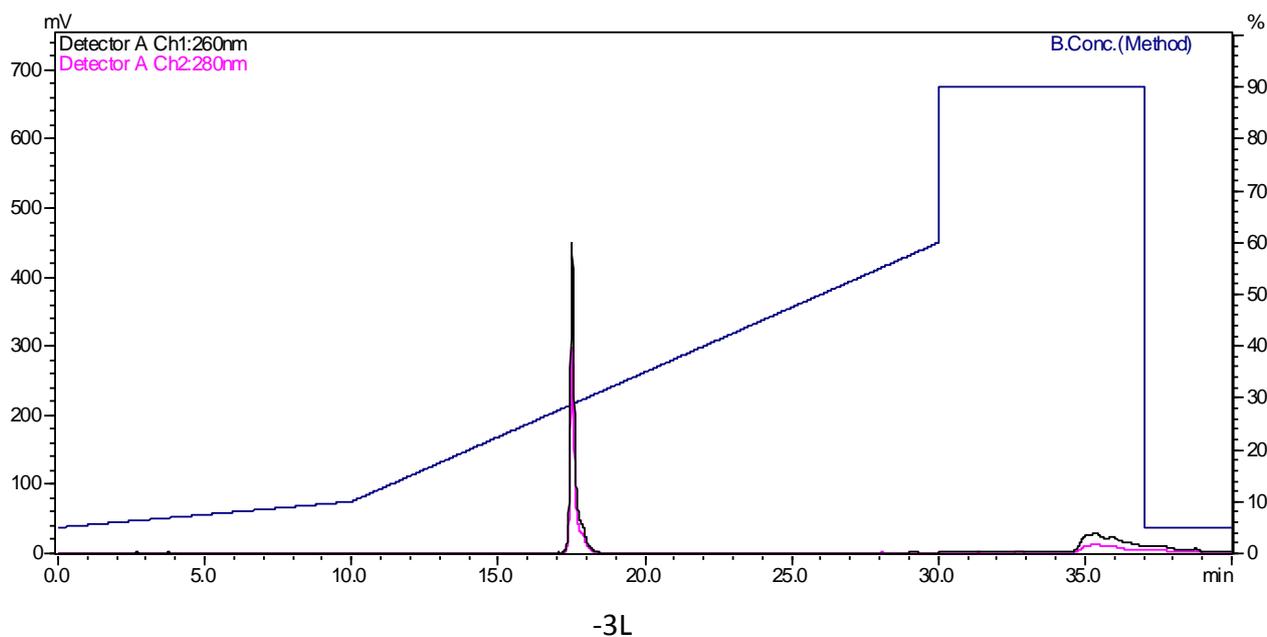
HPLC chromatograms of pure FIT-PNAs	Figure S1	S3-S5
DNA sequences	Table S1	S6
DNA sequences for generating figures 2 and 3	Table S2	S7
Calculation of ΔH NNI		S8
Data used to generate Figure 6	Table S3	S9-S11

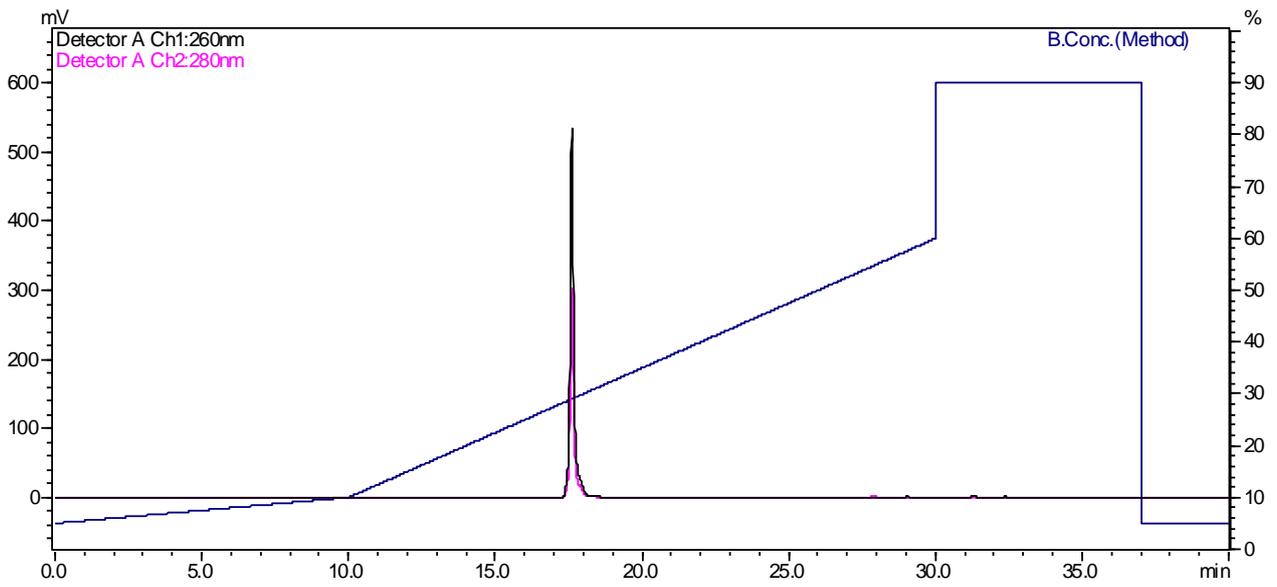
Figure S1. HPLC chromatograms of purified FIT-PNAs. Over 95% purity for all FIT-PNAs

RP-HPLC (Shimadzu LC2010), semi-preparative C18 reverse-phase column (Phenomenex, Jupiter 300 A) at a flow rate of 4 mL/min. Mobile phase: 0.1% TFA in H₂O (A) and acetonitrile (B).

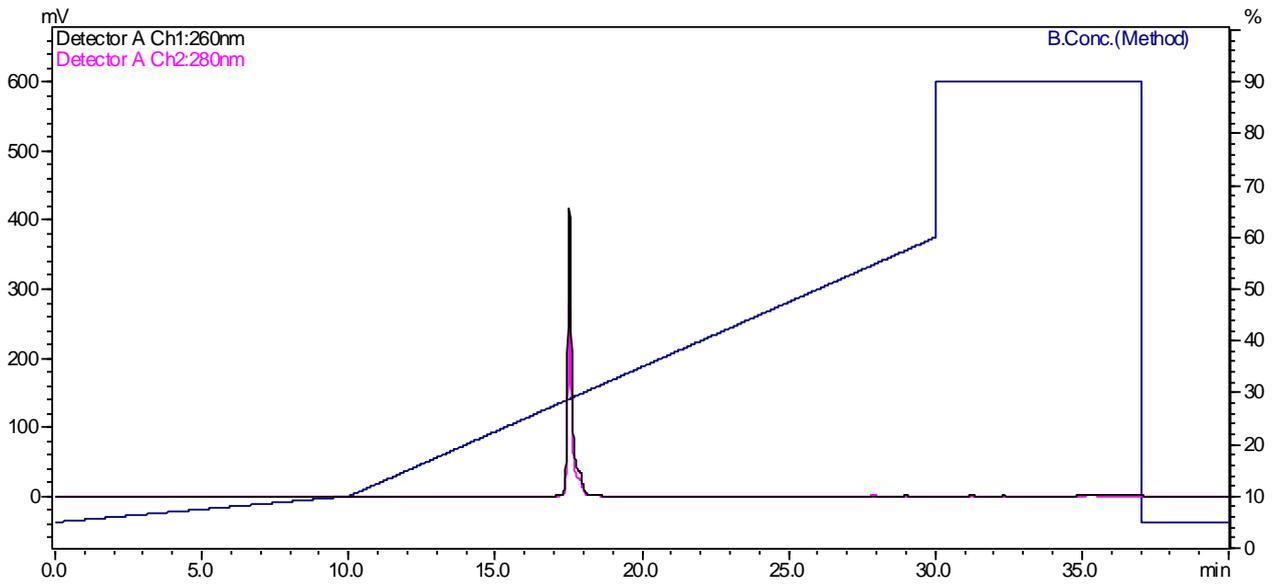
Gradient: 0-10 min: 90% A, 10% B. 10-30 min: linear gradient from 90% A-10%B to 40% A-60% B.

30-37 min: linear gradient from 40% A-60%B to 10% A-90% B. 37-40 min: 5% A-95% B.

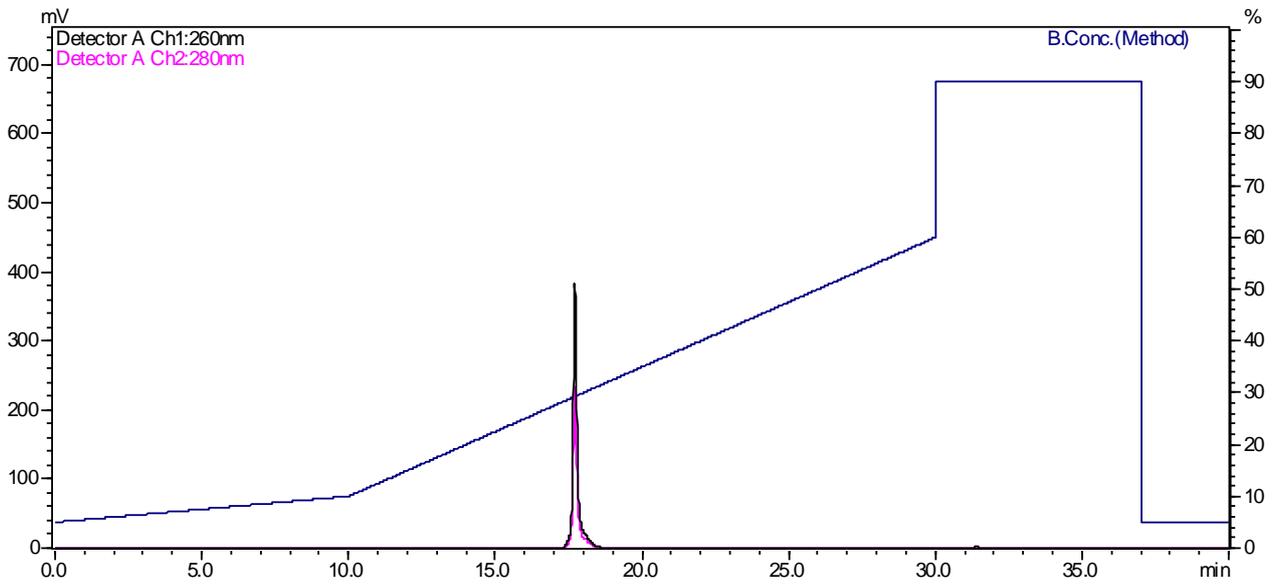




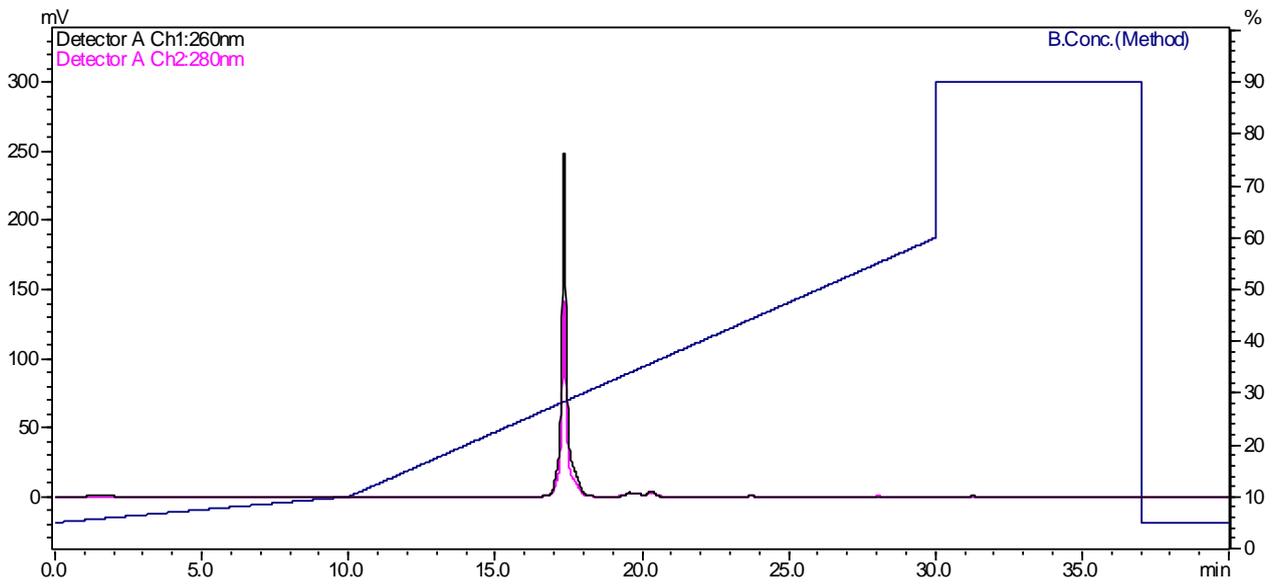
+3L



-3R



OR



+3R

DNA sequences:

A: 5'CCCCTAGATCATCAGCTATGT**A**TGTTGCTTTTGATAATAAAAAT³
B: 5'CCCCTAGATCATCAGCTATGT**G**TGTTGCTTTTGATAATAAAAAT³
C: 5'CCCCTAGATCATCAGCTATGT**T**TGTTGCTTTTGATAATAAAAAT³
D: 5'CCCCTAGATCATCAGCTATGT**C**TGTTGCTTTTGATAATAAAAAT³
E: 5'CCCCTAGATCATCAGCTATGA**A**TGTTGCTTTTGATAATAAAAAT³
F: 5'CCCCTAGATCATCAGCTATGC**A**TGTTGCTTTTGATAATAAAAAT³
G: 5'CCCCTAGATCATCAGCTATGG**A**TGTTGCTTTTGATAATAAAAAT³
H: 5'CCCCTAGATCATCAGCTAT**T**TATGTTGCTTTTGATAATAAAAAT³
I: 5'CCCCTAGATCATCAGCTAT**A**TA**A**TGTTGCTTTTGATAATAAAAAT³
J: 5'CCCCTAGATCATCAGCTAT**C**TA**A**TGTTGCTTTTGATAATAAAAAT³
K: 5'CCCCTAGATCATCAGCTA**A**GT**A**TGTTGCTTTTGATAATAAAAAT³
L: 5'CCCCTAGATCATCAGCT**G**TGT**A**TGTTGCTTTTGATAATAAAAAT³
M: 5'CCCCTAGATCATCAGCA**A**ATGT**A**TGTTGCTTTTGATAATAAAAAT³
N: 5'CCCCTAGATCATCAGCTATGT**A**AGTTGCTTTTGATAATAAAAAT³
O: 5'CCCCTAGATCATCAGCTATGT**A**CGTTGCTTTTGATAATAAAAAT³
P: 5'CCCCTAGATCATCAGCTATGT**A**GGTTGCTTTTGATAATAAAAAT³
Q: 5'CCCCTAGATCATCAGCTATGT**A**T**T**TTGCTTTTGATAATAAAAAT³
R: 5'CCCCTAGATCATCAGCTATGT**A**TATTGCTTTTGATAATAAAAAT³
S: 5'CCCCTAGATCATCAGCTATGT**A**T**C**TTGCTTTTGATAATAAAAAT³
T: 5'CCCCTAGATCATCAGCTATGT**A**TGA**A**TGCTTTTGATAATAAAAAT³
U: 5'CCCCTAGATCATCAGCTATGT**A**TGT**A**AGCTTTTGATAATAAAAAT³
V: 5'CCCCTAGATCATCAGCTATGT**A**TGTT**A**CTTTTGATAATAAAAAT³
W: 5'CCCCTAGATCATCAGCTATGT**A**TGTT**C**TTTTGATAATAAAAAT³

Table S1: A-W DNA sequences were hybridized to FIT-PNA's to create a 114 mismatched FIT-PNA:DNA duplexes ('A' corresponds to K13-C580Y and 'G' to K13-WT.)

Table S2: DNA sequences used to create data in Figures 2 and 3:

‘R’ PNA’s: ³TAGTCGATACAT**BC**AAACGAAAAC⁵
⁵CCCCTAGATCATCAGCTATGT**AT**GTTGCTTTTGATAATAAAAT³
 ‘L’ PNA’s: ³TAGTCGATAC**BT**ACAACGAAAAC⁵

DNA SEQUENCES:

A: ⁵CCCCTAGATCATCAGCTATGT**AT**GTTGCTTTTGATAATAAAAT³
B: ⁵CCCCTAGATCATCAGCTATGT**G**TGTTGCTTTTGATAATAAAAT³
C: ⁵CCCCTAGATCATCAGCTATGT**AA**TGTTGCTTTTGATAATAAAAT³
D: ⁵CCCCTAGATCATCAGCTAT**ATA**TGTTGCTTTTGATAATAAAAT³
E: ⁵CCCCTAGATCATCAGCTA**AGT**ATGTTGCTTTTGATAATAAAAT³
F: ⁵CCCCTAGATCATCAGCT**G**TGT**AT**GTTGCTTTTGATAATAAAAT³
G: ⁵CCCCTAGATCATCAGCA**AA**TGT**AT**GTTGCTTTTGATAATAAAAT³
H: ⁵CCCCTAGATCATCAGCTATGT**AA**GTTGCTTTTGATAATAAAAT³
I: ⁵CCCCTAGATCATCAGCTATGT**ATA**TGTTGCTTTTGATAATAAAAT³
J: ⁵CCCCTAGATCATCAGCTATGT**ATGA**TGTTGCTTTTGATAATAAAAT³
K: ⁵CCCCTAGATCATCAGCTATGT**ATGTAG**CTTTTGATAATAAAAT³
L: ⁵CCCCTAGATCATCAGCTATGT**ATGTTA**CTTTTGATAATAAAAT³

A-L DNA sequences are hybridized to FIT-PNA’s to create a duplex with a mismatch at a position with a defined distance and direction from BisQ. (‘A’ corresponds to K13-C580Y and ‘B’ to K13-WT.)

K13-C580Y SNP position nucleotide is marked in red, mismatched position nucleotide is highlighted in yellow.

For example: the ‘0L’:‘E’ PNA:DNA mismatched duplex is a ‘-2’ mismatch, since the mismatch position of this duplex is 2 bases from BisQ towards the DNA 5’ direction:

‘E’ DNA: ⁵CCCCTAGATCATCAGCTA**AGT**ATGTTGCTTTTGATAATAAAAT³
 ‘0L’ PNA: ³TCGATA**ACBT**ACAACGAAA⁵

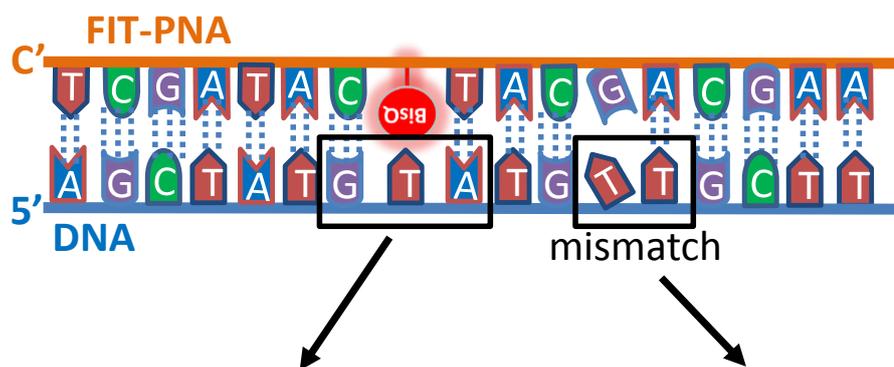
The same DNA sequence will create a ‘-4’ mismatch when hybridized with ‘0R’ FIT-PNA to create the ‘0R’:‘E’ PNA:DNA mismatched duplex, since now the mismatch position of the duplex is 4 bases from BisQ to the DNA 5’ direction:

‘E’ DNA: ⁵CCCCTAGATCATCAGCTA**AGT**ATGTTGCTTTTGATAATAAAAT³
 ‘0R’ PNA: ³TCGATA**ACAT**BCAACGAAA⁵

Calculation of ΔH NNI

The predictive value for each duplex was calculated by summing the ΔH° (NNI) of the two NNI's in the nucleobases triad opposite BisQ and the NNI of the base stack 5' to 3' from the DNA mismatched base. The general equation is shown below:

$$\text{NNI factor} = \Delta H^\circ_{\text{NNI}} (\text{BisQ triad}) + \Delta H^\circ_{\text{NNI}} (\text{mismatch})$$



$$\Delta H^\circ_{\text{NNI}} (\text{BisQ triad}) = \Delta H^\circ (\text{GT}) + \Delta H^\circ (\text{TA})$$

$$\Delta H^\circ_{\text{NNI}} (\text{mismatch}) = \Delta H^\circ (\text{TT})$$

For a mismatch occurring adjacent to BisQ at the 5' end, one of the NNI of the triad opposite BisQ is considered un-stacked, and therefore calculated as zero.

For example: the '-3L':'U' PNA:DNA mismatched duplex shown below has a GT+TA triad NNI and an AG NNI 5' to 3' from the mismatched DNA nucleobase.



ΔH° (GT) = 6.5, ΔH° (TA) = 6.0 and ΔH° (AG) = 7.8.

Therefore the predicted value for the '-3L':'U' duplex is $\Delta H^\circ_{\text{total}} = 6.5+6+7.8=20.3$ (Table 1 shows ΔH values).

Table S3: Data used to generate the scatter plot (**Figure 6**) including statistical analysis and NNI calculations:

'L' N=3 results and outlier analysis																				
	M	L	K	H	I	J	C	D	B	N	P	O	Q	R	S	T	U	W	V	
-3L	exp1	0.48	0.47	0.86	0.62	0.39	0.46	0.81	0.53	0.53	0.26	0.67	0.63	1.09	0.42	0.94	0.57	0.70	0.92	0.82
	exp2	0.70	1.28	1.12	0.56	0.40	0.49	0.91	0.62	0.69	0.67	0.68	0.83	1.05	0.55	0.75	0.99	0.84	1.02	1.04
	exp3	0.78	1.06	0.96	0.60	0.51	0.41	0.97	0.58	0.74	0.64	0.70	0.65	1.04	0.64	0.67	0.88	0.82	1.03	0.99
0L	exp1	0.58	0.72	0.69	0.41	0.50	0.40	0.78	0.54	0.75	0.26	0.63	0.82	1.17	0.47	1.04	0.68	0.55	0.66	0.49
	exp2	0.84	1.22	0.97	0.51	0.47	0.53	0.92	0.62	0.77	0.65	0.84	1.15	1.07	0.60	0.78	0.94	0.80	0.92	0.94
	exp3	0.63	0.98	0.75	0.55	0.50	0.50	1.03	0.65	0.80	0.55	0.57	0.71	1.11	0.67	0.77	0.77	0.67	0.78	0.76
+3L	exp1	0.90	0.90	0.41	0.66	0.57	0.65	1.06	0.73	0.89	0.59	0.78	0.82	1.26	0.78	0.94	0.81	0.75	0.91	1.02
	exp2	0.67	0.69	0.64	0.53	0.50	0.48	0.98	0.63	0.71	0.49	0.60	0.70	1.19	0.73	0.89	0.70	0.65	0.72	0.96
	exp3	0.92	0.95	0.85	0.56	0.53	0.53	1.01	0.68	0.77	0.66	0.73	0.76	1.33	0.76	0.89	0.89	0.84	0.94	1.01
	//If Avg	0.72	0.92	0.81	0.56	0.48	0.49	0.94	0.62	0.74	0.53	0.69	0.79	1.15	0.62	0.87	0.80	0.73	0.88	0.89
	SD	0.15	0.26	0.21	0.07	0.06	0.08	0.10	0.06	0.10	0.16	0.09	0.15	0.10	0.13	0.13	0.14	0.10	0.13	0.18
	MIN	0.48	0.47	0.41	0.41	0.39	0.40	0.78	0.53	0.53	0.26	0.57	0.63	1.04	0.42	0.67	0.57	0.55	0.66	0.49
	MAX	0.92	1.28	1.12	0.66	0.57	0.65	1.06	0.73	0.89	0.67	0.84	1.15	1.33	0.78	1.04	0.99	0.84	1.03	1.04
	AVG-MIN	0.24	0.45	0.39	0.14	0.10	0.16	0.09	0.09	0.21	0.27	0.12	0.16	0.11	0.21	0.20	0.23	0.18	0.22	0.40
	MAX-AVG	0.20	0.36	0.32	0.10	0.09	0.16	0.12	0.11	0.15	0.14	0.15	0.36	0.18	0.16	0.16	0.19	0.11	0.15	0.15
	Gmax	1.33	1.38	1.50	1.43	1.47	1.86	1.25	1.75	1.57	0.83	1.75	2.32	1.83	1.23	1.28	1.39	1.09	1.19	0.84
	Gmin	1.64	1.72	1.87	2.05	1.64	1.15	1.66	1.40	2.16	1.65	1.37	1.03	1.05	1.65	1.58	1.71	1.83	1.70	2.27
	Gcrit005	2.11	2.11	2.11	2.11	2.11	1.94	2.11	2.11	2.11	2.11	2.11	2.11	2.11	2.11	1.94	2.11	2.11	2.11	2.11
'L' statistics after outlier (P<0.05, Grubbs test) removal																				
	//If Avg	0.72	0.92	0.81	0.56	0.48	0.49	0.94	0.62	0.76	0.53	0.69	0.74	1.15	0.62	0.87	0.80	0.73	0.88	0.93
	SD	0.15	0.26	0.21	0.07	0.06	0.08	0.10	0.06	0.06	0.16	0.09	0.08	0.10	0.13	0.13	0.14	0.10	0.13	0.10

'L'																			
	M	L	K	H	I	J	C	D	B	N	P	O	Q	R	S	T	U	W	V
LNNI(BI+O)	12.5	12.5	12.5	12.5	12.5	12.5	15.6	12.1	12.3	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5
Z-NNI(W/M)	6	8.6	5.8	6.5	6.5	6.5	8.6	8.6	8.6	5.8	5.8	5.8	6.5	6.5	6.5	9.1	5.8	11.1	11.1
NNI Fac	18.5	21.1	18.3	19	19	19	24.2	20.7	20.9	18.3	18.3	18.3	19	19	19	21.6	18.3	23.6	23.6
//If Avg	0.72	0.92	0.81	0.56	0.48	0.49	0.94	0.62	0.76	0.53	0.69	0.74	1.15	0.62	0.87	0.80	0.73	0.88	0.93
//IfSD	0.15	0.26	0.21	0.07	0.06	0.08	0.10	0.06	0.06	0.16	0.09	0.08	0.10	0.13	0.13	0.14	0.10	0.13	0.10

'R' N=3 results and outlier analysis																			
	M	L	K	H	I	J	E	G	F	C	D	B	Q	R	S	T	U	V	
-3R	exp1	0.63	0.84	0.94	0.91	0.77	0.84	0.54	0.76	0.89	0.98	0.62	0.49	0.82	0.75	0.98	0.74	0.56	1.24
	exp2	0.62	0.94	1.06	0.73	0.74	0.86	0.68	1.20	0.83	0.73	0.50	0.84	0.77	0.94	0.97	0.81	1.34	
	exp3	0.68	1.02	1.06	0.95	0.85	0.68	0.65	0.73	0.86	0.76	0.51	0.88	0.82	0.88	0.88	0.78	1.35	
OR	exp1	0.55	0.90	0.89	0.56	0.49	0.92	0.68	0.83	0.62	0.61	0.48	0.72	0.60	0.75	0.84	0.53	1.09	
	exp2	0.74	0.96	0.91	0.76	0.75	0.88	0.70	1.11	0.71	0.67	0.50	0.77	0.56	0.64	0.92	0.76	1.25	
	exp3	0.79	1.10	0.99	0.79	0.75	0.88	0.75	0.92	0.71	0.68	0.46	0.78	0.58	0.64	1.33	0.76	1.31	
+3R	exp1	0.83	1.03	0.47	1.25	1.30	0.95	0.82	0.97	0.69	0.77	0.61	0.97	0.64	0.71	0.83	0.64	1.19	
	exp2	0.92	1.27	0.99	0.83	1.06	0.77	0.76	1.06	0.58	0.62	0.49	0.62	0.53	0.60	0.92	0.72	1.32	
	exp3	0.91	1.09	0.96	0.83	1.03	0.87	0.71	1.00	0.57	0.62	0.46	0.83	0.50	0.73	0.91	0.73	1.25	
	Wif Avg	0.74	1.02	0.92	0.84	0.87	0.68	0.77	0.95	0.73	0.67	0.50	0.80	0.64	0.77	0.93	0.70	1.26	
	SD	0.13	0.13	0.18	0.19	0.24	0.07	0.06	0.16	0.14	0.06	0.04	0.10	0.12	0.16	0.17	0.10	0.08	
	MIN	0.55	0.84	0.47	0.56	0.49	0.77	0.54	0.72	0.57	0.61	0.46	0.62	0.50	0.60	0.74	0.53	1.09	
	MAX	0.92	1.27	1.06	1.25	1.30	0.95	0.85	1.20	0.98	0.77	0.61	0.97	0.82	0.98	1.33	0.81	1.35	
	AVG-MIN	0.19	0.17	0.45	0.28	0.37	0.10	0.14	0.24	0.16	0.07	0.04	0.18	0.14	0.17	0.19	0.17	0.14	
	MAX-AVG	0.18	0.25	0.14	0.40	0.44	0.07	0.08	0.25	0.25	0.09	0.11	0.16	0.18	0.21	0.40	0.11	0.09	
	Gmax	1.36	1.99	0.81	2.14	1.86	1.18	1.19	1.55	1.84	1.41	2.49	1.67	1.60	1.35	2.43	1.08	1.14	
	Gmin	1.49	1.37	2.51	1.51	1.58	1.61	2.14	1.47	1.16	1.05	0.89	1.88	1.24	1.07	1.15	1.73	2.03	
	Gcrit0.05	2.11	2.11	2.11	2.11	2.11	2.11	2.11	2.11	2.11	2.11	2.11	2.11	2.11	2.11	2.11	2.11	2.11	
'R' statistics after outlier (P<0.05, Grubbs test) removal																			
	Wif Avg	0.74	1.02	0.97	0.80	0.86	0.87	0.68	0.77	0.73	0.67	0.48	0.80	0.64	0.77	0.93	0.70	1.26	
	SD	0.13	0.13	0.06	0.12	0.24	0.06	0.04	0.16	0.14	0.06	0.02	0.10	0.12	0.16	0.07	0.10	0.08	

'R'																		
	M	L	K	H	I	J	E	G	F	C	D	B	Q	R	S	T	U	V
L.NNI(BisQ)	14.4	14.4	14.4	14.4	14.4	14.4	14.4	14.4	14.4	5.8	5.8	5.8	17.7	14.2	14.6	14.4	14.4	14.4
2.NNI(MMM)	6	8.6	5.8	6.5	6.5	6.5	6	6	6	8.6	8.6	8.6	6.5	6.5	6.5	9.1	5.8	11.1
NNI Fac	20.4	23	20.2	20.9	20.9	20.9	20.4	20.4	20.4	14.4	14.4	14.4	24.2	20.7	21.1	23.5	20.2	25.5
Wif Avg	0.74	1.02	0.97	0.80	0.86	0.87	0.68	0.77	0.95	0.73	0.67	0.48	0.80	0.64	0.77	0.93	0.70	1.26
Wif SD	0.13	0.13	0.06	0.12	0.24	0.06	0.04	0.06	0.16	0.14	0.06	0.02	0.10	0.12	0.16	0.07	0.10	0.08

Table S3 shows the data analysis and NNI factor calculations for generating Figure 6 in manuscript. The two tables present the I_f values obtained for each PNA:DNA duplex (N=3 repetitions). Letters (in orange) notate the DNA sequences as given in Table S1 and -3\0\+3 (in light blue) notate the FIT-PNAs. Each data point corresponds to an I_f value. Next, the average I_f and standard deviation (SD) of each FIT-PNA frame (-3\0\+3) for N=3 is shown. Grubs test statistics to evaluate outliers in the data is shown next in the table, performed over the nine values of each averaged group. Outliers of P=0.05 (in red) were excluded and I_f and standard deviation was then recalculated. The NNI factor calculation table is shown beneath each data table; for each DNA sequence the BisQ triad and mismatch NNI values are shown followed by the total NNI factor as their sum. I_f and its standard deviation are shown again for convenience.