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

Itamar Peled† and Eylon Yavin∗∗

†The 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.
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 $\Delta H$ 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.
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:**
  - 3TAGTCGATACBCTACAAAC
  - 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT

- **'L' PNA's:**
  - 3TAGTCGATACBCTACAAAC

**DNA SEQUENCES:***

| A | 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT |
| B | 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT |
| C | 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT |
| D | 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT |
| E | 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT |
| F | 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT |
| G | 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT |
| H | 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT |
| I | 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT |
| J | 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT |
| K | 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT |
| L | 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT |

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:** 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT
- **'0L' PNA:** 3TAGTCGATACBCTACAAAC

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:** 5CCCCTAGATCAGCTATGATACGTTTGTGATAATAAAAT
- **'0R' PNA:** 3TAGTCGATACBCTACAAAC

ACS Omega Supporting Information

S7
Calculation of $\Delta H_{\text{NNI}}$

The predictive value for each duplex was calculated by summing the $\Delta H^\circ$ (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:

\[
\text{NNI factor} = \Delta H_{\text{NNI}}^0 (\text{BisQ triad}) + \Delta H_{\text{NNI}}^0 (\text{mismatch})
\]

\[
\Delta H_{\text{NNI}}^0 (\text{BisQ triad}) = \Delta H^0 (GT) + \Delta H^0 (TA) \quad \Delta H_{\text{NNI}}^0 (\text{mismatch}) = \Delta H^0 (TT)
\]

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 $\text{'-3L':'U'}$ PNA:DNA mismatched duplex shown below has a GT+TA triad NNI and an AG NNI 5’ to 3’ from the mismatched DNA nucleobase.

\[
\begin{align*}
\text{U:} & \quad 5'\text{CCCCCTAGATCATCAGCTATGTATGTAGCTTTTGATAATAAAAT}^3 \\
\text{-3L:} & \quad 3'\text{TAGTCGATACTACAGGAAAC}^5
\end{align*}
\]

$\Delta H^\circ (GT) = 6.5, \Delta H^\circ (TA) = 6.0$ and $\Delta H^\circ (AG) = 7.8$.

Therefore the predicted value for the $\text{'-3L':'U'}$ duplex is $\Delta H_{\text{total}}^0 = 6.5 + 6 + 7.8 = 20.3$ (Table 1 shows $\Delta H$ values).
**Table S3**: Data used to generate the scatter plot (Figure 6) including statistical analysis and NNI calculations:

| exp1 | exp2 | exp3 | NNI(MM) | NNI(BisQ) |
|------|------|------|---------|-----------|
| 0.48 | 0.79 | 0.58 | 0.47 | 0.79 | 0.71 |
| 0.81 | 0.81 | 0.38 | 0.49 | 0.29 | 0.46 |
| 0.46 | 0.77 | 0.87 | 0.88 | 0.63 | 0.94 |
| 0.13 | 0.53 | 0.81 | 0.49 | 0.81 | 0.71 |

**Figure 6**: Statistical analysis and NNI calculation.
### 'R' N-3 results and outlier analysis

|   | M   | L   | K   | H   | I   | J   | E   | B   | Q   | R   | S   | T   | U   | W   | V   |
|---|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 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 | 0.96 |
| exp2 | 0.62 | 0.94 | 1.06 | 0.73 | 0.74 | 0.86 | 0.68 | 0.75 | 1.20 | 0.83 | 0.73 | 0.50 | 0.84 | 0.77 | 0.94 | 0.97 | 0.81 | 1.34 | 1.21 |
| 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 | 0.82 | 0.88 | 0.78 | 1.35 | 1.17 |

### Outlier Analysis

- **N=3 results and outlier analysis**
- **Statistics after outlier (P<0.05, Grubbs test) removal**

### 'R' statistics after outlier removal (P<0.05, Grubbs test) removal

|   | M   | L   | K   | H   | I   | J   | E   | B   | Q   | R   | S   | T   | U   | W   | V   |
|---|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Full Avg | 0.74 | 1.02 | 0.92 | 0.84 | 0.86 | 0.87 | 0.87 | 0.77 | 0.95 | 0.73 | 0.67 | 0.50 | 0.80 | 0.64 | 0.77 | 0.93 | 0.70 | 1.26 | 1.09 |

### SD

|   | 0.13 | 0.13 | 0.18 | 0.19 | 0.24 | 0.06 | 0.07 | 0.06 | 0.16 | 0.14 | 0.06 | 0.04 | 0.10 | 0.12 | 0.16 | 0.17 | 0.10 | 0.08 | 0.09 |

### MIN

|   | 0.55 | 0.84 | 0.47 | 0.56 | 0.49 | 0.77 | 0.54 | 0.65 | 0.72 | 0.57 | 0.61 | 0.46 | 0.62 | 0.50 | 0.60 | 0.74 | 0.53 | 0.10 | 0.95 |

### MAX

|   | 0.92 | 1.27 | 1.06 | 1.25 | 1.30 | 0.95 | 0.76 | 0.85 | 1.20 | 0.98 | 0.77 | 0.61 | 0.97 | 0.82 | 0.98 | 1.33 | 0.81 | 1.35 | 1.21 |

### AVG-MIN

|   | 0.19 | 0.17 | 0.45 | 0.28 | 0.37 | 0.19 | 0.14 | 0.12 | 0.24 | 0.16 | 0.07 | 0.04 | 0.18 | 0.14 | 0.17 | 0.19 | 0.17 | 0.14 | 0.14 |

### MAX-AVG

|   | 0.18 | 0.25 | 0.14 | 0.40 | 0.44 | 0.07 | 0.08 | 0.04 | 0.25 | 0.25 | 0.09 | 0.11 | 0.16 | 0.18 | 0.21 | 0.40 | 0.11 | 0.09 | 0.12 |

### G (max)

|   | 1.36 | 1.99 | 0.81 | 2.24 | 1.86 | 1.18 | 1.19 | 1.31 | 1.55 | 1.84 | 1.41 | 2.29 | 1.67 | 1.60 | 1.35 | 2.43 | 1.08 | 1.14 | 1.34 |

### G (min)

|   | 1.49 | 1.37 | 2.51 | 1.51 | 1.58 | 1.61 | 2.14 | 1.87 | 1.47 | 1.16 | 1.05 | 0.89 | 1.88 | 1.24 | 1.07 | 1.15 | 1.73 | 2.03 | 1.65 |

### G crit 0.05

|   | 2.11 | 2.11 | 2.11 | 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 | 2.11 | 2.11 | 2.11 |

### 'R' statistics after outlier (P<0.05, Grubbs test) removal

|   | 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 | 1.09 |

### 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 | 0.09 |
Table S3 shows the data analysis and NNI factor calculations for generating Figure 6 in manuscript. The two tables present the $I/I_f$ 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 $I/I_f$ value. Next, the average $I/I_f$ 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 $I/I_f$ 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. $I/I_f$ and its standard deviation are shown again for convenience.