Science
AAAS

Supplementary Materials for

Hachimoji DNA and RNA. A Genetic System with Eight Building Blocks

Shuichi Hoshika[1,2†], Nicole A. Leal[1,2†], Myong-Jung Kim[1,2], Myong-Sang Kim[1], Nilesh B. Karalkar[1,2], Hyo-Joong Kim[1], Alison M. Bates,[3] Norman E. Watkins Jr.[4], Holly A. SantaLucia[4], Adam J. Meyer,[5] Saurja DasGupta[6], Joseph A. Piccirilli,[6] Andrew D. Ellington,[5] John SantaLucia Jr.[4], Millie M. Georgiadis,[3] and Steven A. Benner[1,2] *

Correspondence to: sbenner@ffame.org

This PDF file includes:

Materials and Methods

The IUPAC systematic names for the added nucleotides in hachimoji DNA are:

d**S**: 3-methyl-6-amino-5-(1'-β-D-2'-deoxyribofuranosyl)-pyrimidin-2-one

d**B**: 6-amino-9-(1'-β-D-2'-deoxyribofuranosyl)-4-hydroxy-5-(hydroxymethyl)-oxolan-2-yl]-1H-purin-2-one

d**Z**: 6-amino-3-(1'-β-D-2'-deoxyribofuranosyl)-5-nitro-1H-pyridin-2-one

d**P**: 2-amino-8-(1'-β-D-2'-deoxyribofuranosyl)-imidazo-[1,2a]-1,3,5-triazin-[8H]-4-one

The IUPAC systematic names for the added nucleotides in hachimoji RNA are:

S: 2-amino-1-(1'-β-D-ribofuranosyl)-4(1H)-pyrimidinone

B: 6-amino-9-(1'-β-D-ribofuranosyl)-4-hydroxy-5-(hydroxymethyl)-oxolan-2-yl]-1H-purin-2-one

Z: 6-amino-3-(1'-β-D-ribofuranosyl)-5-nitro-1H-pyridin-2-one

P: is 2-amino-8-(1'-β-D-ribofuranosyl)-imidazo-[1,2a]-1,3,5-triazin-[8H]-4-one

Synthesis and purification of AEGIS oligonucleotides

Standard phosphoramidites (Bz-dA, Ac-dC, dmf-dG, and dT) and d**B** (isoG) phosphoramidite and CPG having standard residues were from Glen Research (Sterling, VA). The other non-standard hachimoji phosphoramidites (d**Z**, d**P** and d**S**) were synthesized as reported previously (*26*, *27*). All oligonucleotides containing non-standard hachimoji components were synthesized on an ABI 394 DNA Synthesizer following standard phosphoramidite chemistry. The CPGs having oligonucleotides were treated with 2.0 mL of 1 M DBU in anhydrous acetonitrile at room temperature for 24 hours to deprotect the NPE group on the d**Z** nucleobase (*26*). Then the CPGs were filtered, dried, and treated with concentrated ammonium hydroxide at 55 ºC for 16 hours. After removal of ammonium hydroxide, the AEGIS oligonucleotides were purified on ion-exchange HPLC, and then desalted using Sep-Pac® Plus C18 cartridges (Waters). These are collected in **Table S1**.

Thermodynamics Methods

Thermodynamics were determined for duplexes containing **S**:**B** and **Z**:**P** base pairs by measuring UV absorbance versus temperature profiles over six different oligonucleotide concentrations in buffer containing 1M NaCl, 10 mM Na_2HPO_4, and 0.5 mM disodium ethylenediaminetetraacetic acid (Na_2EDTA), pH 7.00 (referred to as "1M NaCl buffer"). These data were then processed using *Meltwin* v.3.5 to obtain a full thermodynamic parameter set through two different methods: T_m^{-1} vs. Ln(Ct) method and the Marquardt non-linear curve fit method (**Table S4, Table S7, and Table S10**). A duplex denaturation data set was considered to be two-state if the two methods yield $\Delta H°$ values that are within 15% of each other *(28)*. By this criterion, all duplexes except for 1 exhibited two-state melting behavior. The one duplex (from the tandem **Z**:**P** and **S**:**B** set in **Table S10**) that was not two-state was 5'-C**SZ**TA**PB**G-3'. It showed no discernable transition at all in the melting curve.

The error-weighted average of these two methods were then used as the "experimental values" in **Table S3, Table S6, and Table S9**, and plotted in **Figure** 2, **S1-S6**. These were used in singular value decomposition (SVD) to determine base pair dimer nearest-neighbor (BPD "NN") parameters shown in **Figure S1**, **S3**, and **S5**. To perform SVD analysis, for each duplex the total $\Delta G°_{37}$ we subtracted the contributions from the Watson-Crick (i.e. A:T and G:C) containing nearest-neighbors, the initiation parameter and the symmetry penalty using the nearest-neighbor

parameters from SantaLucia *(17)*. This resulted in a set of equations in which for each sequence a total ΔG°_{37} increment was equal to a sum of **S**:**B** and **Z**:**P** containing NN (i.e. unknowns). The SVD calculations were carried out as described in *(28)*. For the **Z**:**P** containing NN, a total of 41 equations were used to solve for 11 NN parameters plus a parameter for a terminal **Z**:**P** pair (**Figure S1**). For the **S**:**B** containing NN, a total of 37 equations were used to solve for 11 NN parameters plus a parameter for a terminal **S**:**B** pair (**Figure S3**). For the tandem **Z**:**P** and **S**:**B** containing NN, a total of 15 equations were used to solve for 4 NN parameters (**Figure S5**). Similar calculations were also performed to determine the NN ΔH° parameters. Nearest-neighbor values for ΔS° may be calculated from the equation: $\Delta S^\circ = (\Delta H^\circ - \Delta G^\circ_{37})*1000/310.15$. Note that while this study was performed with 1 M NaCl conditions, the NN presented can be used to make predictions at other salt conditions by using the salt extrapolations described in the literature *(17)*. Note that we also melted the three 16 base-pair duplexes that were submitted for X-ray crystallography. However, 2 of these sequences had melting transitions with T_m's that were too high and could not be fitted. One duplex, (CTTAT**PPSBZZ**ATAAG)$_2$, did give a discernable high temperature baseline, but the thermodynamics are non-two-state due to the competition between random-coil, hairpin, and duplex, which is typical of long self-complementary duplexes (in addition such a long duplex would exhibit slow unfolding kinetics, and the high-temperature baseline is limited). Thus, we included this duplex in the validation plots (**Figure 2**) and **Table S9 and S10**, but we excluded that sequence from the NN determination.

Quality of the **Z**:**P** and **S**:**B** nearest-neighbor (NN) parameters

The quality of the NN parameters can be assessed by three different methods: the level of over-determination of the parameters, the standard error values, and the quality of the predictions using the NN parameters. In this study, a total of 93 measurements (the **SBZP**-5: 5'-C**SZ** TA**P** **B**G duplex was omitted from the SVD determination because it is non-two-state) were used to solve for 28 unknowns (i.e. 3.3-fold over-determined). Experience in the SantaLucia lab for many different modified nucleotides has indicated that this level of over-determination is sufficient for accurate determination of NN parameters *(17)*. The standard error values for the NN parameters are shown in **Figures S1, S3** and **S5** and **Table S2, Table S5, and Table S8**. Most of the NN ΔG°_{37} parameters have errors of about 0.1 kcal/mol. This error level is essentially converged to the limit of what can be expected for the limits of the NN model itself. In principle, the errors could be lowered further by making more measurements, but that would provide only marginal improvement. For example, if the number of measurements in the study was expanded from 94 measurements to 188 (i.e. double the number of measurements), would reduce the standard errors by a factor of SQRT(2) to about 0.07 kcal/mol. In other words, negligible improvement would be obtained by further measurements. Lastly, the NN can be used to predict the duplex thermodynamics (**Figure 2** and **Table S3**, **Table S6**, and **Table S9**). On average, the duplex ΔG°_{37} parameters are predicted within 0.41, 0.31, and 0.36 kcal/mol for duplexes with **Z**:**P**, **S**:**B**, and tandem **Z**:**P** and **S**:**B** pairs, respectively. The quality of predictions of the melting temperatures (T_m) is another metric of the quality of the NN parameters. On average, the duplex T_m's are predicted within 2.6, 2.0, and 2.1 °C for duplexes with **Z**:**P**, **S**:**B**, and tandem **Z**:**P** and **S**:**B** pairs, respectively. We note that such quality of prediction is exceptional for such short 8-mer duplexes, which are highly demanding of the NN model. These error values are consistent with those observed for standard DNA/DNA duplex thermodynamic determinations in the literature from the SantaLucia Lab and for RNA duplexes from the Turner

Lab. For typical oligonucleotides that would be used for PCR and other applications that are in the length range of 18-24 base pairs the T_m prediction errors will likely be significantly smaller than 2.0 °C. Compare this performance to the average error of 1.6 °C for a dataset of ~320 duplexes using the published DNA NN parameters for the usual 4-letter alphabet *(15)*. Lastly, the plots of Experimental vs. Predicted $\Delta G°_{37}$ (**Figures 2**, **S1**, **S3** and **S5**) and T_m shown in **Table S3, Table S6 and Table S9** show the quality of the predictions with slopes near 1, intercepts near 0, and R^2 >0.8 for all plots. Again, such high-quality plots for 8-mer sequences is impressive. For longer sequences (e.g. 18-24 base pairs) with larger ranges in $\Delta G°_{37}$ and T_m, the regression parameters for experiments vs. predictions would likely be even better.

Trends in NN parameters

Table S2 compares the NN trends for NN containing a **Z**:**P** pair compared to the similar NN with C:G pair. For example, the NN A**P**/T**Z** (-1.66 kcal/mol) is compared to AC/TG (-1.28 kcal/mol), with a $\Delta\Delta G°_{37}$ of -0.38 kcal/mol. On average, an **Z**:**P** substitution is -0.17 kcal/mol more stable than the similar NN that has a C:G pair. This suggests that each C:G to **Z**:**P** substitution would make the duplex -0.34 kcal/mol more stable. However, the average masks the significant sequence dependence observed. Depending on the sequence context, a single C:G to **Z**:**P** substation would have $\Delta\Delta G°_{37}$ vary from +0.84 to -0.86 kcal/mol. This large range is likely due to the different dipole moment sizes and directions for **Z**:**P** compared to C:G, thus impacting the trends in dipole-dipole contributions to base pair stacking. Tandem **Z**:**P** NN also show a large sequence dependence that is different than that for tandem C-G NN (e.g. **ZP**/**PZ** vs. **PZ**/**ZP** differ by -1.27 kcal/mol, while CG/GC vs. GC/CG differ by +0.07 kcal/mol). This is further evidence for stacking differences of **Z**:**P** compared to C:G.

Table S5 compares the NN trends for NN containing a **S**:**B** pair compared to the similar NN with T:A pair. For example, the NN A**B**/T**S** (-1.24 kcal/mol) is compared to AA/TT (-1.00 kcal/mol), with a $\Delta\Delta G°_{37}$ of -0.24 kcal/mol. On average, an **S**:**B** substitution is -0.53 kcal/mol more stable than the similar NN that has a T:A pair. This suggests that each T:A to **S**:**B** substitution would make the duplex -1.06 kcal/mol more stable. However, the average masks the significant sequence dependence observed. Depending on the sequence context, a single T:A to **S**:**B** substitution has $\Delta\Delta G°_{37}$ vary from -0.32 to -1.58 kcal/mol. This large range is likely due to the different dipole moment sizes and directions for S:B compared to T:A, thus impacting the trends in dipole-dipole contributions to base pair stacking. Tandem **S**:**B** NN also show a small sequence dependence that is slightly different than that for tandem T:A NN (e.g. **SB**/**BS** vs. **BS**/**SB** differ by -0.01 kcal/mol, while TA/AT vs. AT/TA differ by +0.30 kcal/mol). This is further evidence for stacking differences of **S**:**B** compared to T:A.

Table S8 compares the NN trends for NN containing tandem **S**:**B** and **Z**:**P** pairs compared to the similar NN with C:G pairs. For example, the NN **SZ**/**BP** (-2.29 kcal/mol) is compared to CC/GG (-1.84 kcal/mol), with a $\Delta\Delta G°_{37}$ of -0.45 kcal/mol. On average, the tandem **S**:**B**, **Z**:**P** NN are -0.08 kcal/mol more stable than their C:G counterparts. Interestingly, the tandem **S**:**B**, **Z**:**P** NN show only a weak sequence dependence, with $\Delta G°_{37}$ values ranging from -1.85 to -2.29 kcal/mol.

Transcription of hachimoji DNA to give hachimoji RNA, and its purification

The sense and antisense DNA strands were made by solid phase synthesis as described above, as was done with the oligonucleotides used in melting temperature studies. The ribonucleoside triphosphates r**B**TP, r**S**TP, r**P**TP and r**Z**TP were synthesized as reported previously (*22*, *29*).

Templates were annealed by independently mixing equimolar ratios of top and bottom strands in 20 mM NaCl, 40 mM Tris pH 7.8, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM dithiothreitol, heating to 90°C for 1 min followed by cooling to room temperature at 0.1°C/sec.

Transcriptions contained 0.2 μM template, 40 mM Tris (pH 7.8), 20 mM NaCl, 2 mM spermidine, 10 mM dithiothreitol, 2 mM each of appropriate rNTPs, 0.2 μCi/μL alpha labeled triphosphate (α^{32}P-CTP or α^{32}P-GTP) and either 16 mM (standard transcription with rGTP, rATP, rCTP, UTP), 18 mM (control transcriptions with one AEGIS ribotriphosphate and rGTP, rATP, rCTP, UTP) or 24 mM $MgCl_2$ (hachimoji transcriptions with rGTP, rATP, rCTP, UTP, r**P**TP, r**Z**TP, r**B**TP, and r**S**TP). Transcriptions were initiated with the addition of T7 RNA polymerase (0.175 μg/μL). After incubation (37°C, 16 hours), samples were processed with phenol:chloroform:isoamyl alcohol (24:24:1) to remove proteins followed by ethanol precipitation and polyacrylamide gel purification (7 M urea). RNA was extracted in 0.3 M sodium acetate, recovered by ethanol precipitation, quantified by UV spectroscopy and used in subsequent analysis. T7 RNA polymerase variants and control and aptamer sequences are outlined in **Table S11** and **Table S12**, respectively.

Model T2-templates (T2N) that have only one non-standard nucleotide (**N**)

5′-GGC GTA ATA CGA CTC ACT ATA GGG AGT GTT GTA TTT GGN CAA T T T
3′-CCG CAT TAT GCT GAG TGA TAT CCC TCA CAA CAT AAA CCN GTT AmAmA

Oligonucleotides containing a single non-standard hachimoji component for polymerase screen

To identify T7 RNA polymerases able to incorporate hachimoji ribonucleotides by transcription of hachimoji DNA, a series of template-promoter sequences were designed having a single hachimoji non-standard component, followed by a unique cytidine. In each, transcription with alpha ^{32}P-labeled CTP would create an RNA product with a bridging ^{32}P -labeled phosphate. Upon digestion with ribonuclease T2, this phosphate would end up bonded at the 3'-OH of the hachimoji nucleoside-3'-phosphate.

Preliminary run-off experiments with T7 RNA polymerase variants

Wild-type T7 RNA polymerase readily incorporated **B**TP opposite template d**S**, **P**TP opposite template d**Z**, and **Z**TP opposite template d**P**. However, as we reported in 1993 with a variant of **S** *(22)*, wild-type T7 RNA did not readily incorporate **S**TP opposite template d**B**. We interpreted this structurally, based on the fact that of all the nucleosides, natural and unnatural that are incorporated into the hachimoji system, only **S** has a nucleobase that fails to present electron density to the minor groove. Instead of an oxygen (exocyclic) or a ring purine nitrogen (N3), the heterocycle of **S** presents an amino group and a hydrogen bond donor to the minor groove.

In the past, we have engineered, DNA polymerases and reverse transcriptase to accept nucleobases such as **S** without substantial loss of fidelity, thermostability, or catalytic efficiency *(10, 30)*. Thus, we sought to generate RNA polymerase variants to accept **S**TP.

Here, we relied on a literature from laboratories seeking to improve the ability of T7 RNA polymerase to accept 2'-O-methyl ribonucleoside triphosphates and be more thermostable, all

while losing neither fidelity nor catalytic efficiency. The results of screening of several of these for their ability to incorporate **S**TP opposite template d**B** are shown in **Figure S7**.

Here, internal P32 labeling was done so that we could observe all of the failure sequences. While many of the variants had improved ability to incorporate **S**TP opposite template d**B**, the triple mutant Y639F H784A P266L (FAL) stood out. It shows no detectible deficit in fidelity and only modestly slower rate. However, it is thermostable and accepts 2'-O-methyl ribonucleoside triphosphates (*23*). Essentially all experiments were done with this variant.

The T7 duplex was prepared by mixing equimolar ratios of top strand (T2-template**S**) and bottom strand (T2-template**S**-comp) in 1X transcription buffer (20 mM NaCl, 40 mM Tris pH 7.8, 6 mM $MgCl_2$, 2 mM spermidine, and 10 mM DTT), heating to 90°C, and then slowly cooling to room temperature.

The transcription reaction mixtures contained a final concentration of 1 μM annealed template duplex, 1X transcription buffer (20 mM NaCl, 40 mM Tris pH 7.8, 18 mM $MgCl_2$, 2 mM spermidine, and 10 mM DTT), 2 mM of each rNTP (in minus assays r**S**TP was omitted), 0.5 μCi/μL α^{32}P-GTP and T7 RNA polymerase (0.025 μg/μL or 0.175 μg/μL of the FAL variant). Reaction mixtures were incubated at 37°C for 2 hours and overnight. Reactions were quenched with 3X formamide quench buffer and samples were resolved on a 20% PAGE.

PAGE (20%) was used to resolve transcription products using internal labeling. Wild-type and mutant T7 RNA polymerases were tested in the absence and presence of r**S**TP to show pausing and rescue. Full length product is a 24mer, **S** is at position 18 with pausing most prominent at position 17, depending on the variant. Specifically, the T7, FA and FAL mutants show pausing in the absence of r**S**TP and various levels of rescue in the presence of r**S**TP.

Variants of T7 RNA polymerase produce full length product (24mer) in the absence and presence of r**S**TP. A prominent "failure" band appears in the absence of r**S**TP with these variants, suggesting that the polymerase has difficulty further extending a primer with a mismatched end, where one of the standard nucleotides is mismatched against template d**B**. This pausing is rescued in the presence of r**S**TP. As seen in **Figure S7**, the failure band is most prominent at position 17, one base before the AEGIS site where **S** is called for; rescue increases with wild-type, FA and FAL mutants.

Sample Preparation for fluorescence and CD assays of spinach variants

Purified RNA was refolded in buffer containing 10 mM Tris (pH 7.5), 100 mM K^+ and 5 mM Mg^{2+} (the RNA sample was heated at 90˚C for one minute in water followed by incubation on ice for 3 minutes and at room temperature for 5 minutes, then diluted into folding buffer and heated at 55 ˚C for 30 minutes followed by incubation on ice for 5 minutes). Two equivalents of the Spinach fluorophore, DFHBI ((5Z)-5-[(3,5-difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2,3-dimethyl- 4H-imidazol-4-one, (Z)-4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1H-imidazol-5(4H)-one) were added, and the mixture was incubated at 37˚C for 30 min.

Fluorescence spectroscopy of spinach variants

Fluorescence emission was measured using a Fluorolog-3 spectrofluorometer equipped with a thermo-controller (Horiba Inc.) at an excitation wavelength of 468 nm and emission range of 490-520 nm with a slit width of 5 nm; reported results are the average of three consecutive independent measurements, and each measurement was performed in triplicate (the data points represent the average of these triplicate measurements). Experiments were performed at 25˚C.

Data were normalized and plotted in Microsoft Excel. The concentrations of RNA and fluorophore were 5 μM and 20 μM, respectively.

Circular Dichroism (CD) Spectroscopy of spinach variants

Measurements were performed using a JASCO-1500 CD Spectrometer. Three scans from 180 to 320 nm at 1 nm intervals were accumulated with a scan rate of 100 nm $min^{-1}$and averaged (Path length=1 mm). Experiments were performed at 25˚C. Data were plotted in Microsoft Excel. The concentrations of RNA and DFHBI were 5 μM and 20 μM, respectively. Maxima at ~260 nm suggests a parallel quadruplex, as usually seen in RNA. Spinach has a maximum around 264 nm as seen from our truncated construct *(31)*. **Z** near the quadruplex does not hinder quadruplex formation (**Fig. S8**).

HPLC analysis of RNA transcripts containing hachimoji **S**.

The ability of the FAL variant to incorporate **S**TP opposite template d**B** was shown first using HPLC analysis. Here, product RNA was treated with RNase T2, which does near total degradation to give nucleoside 3'-phosphates. These were separated by ion exchange HPLC using ammonium bicarbonate gradients (aqueous, from 0 to 200 mM).

S-3'-phosphate as authentic was prepared as described below. Protected N-dimethylformamidyl 5'-DMT 2'-TBDMS r**S** phosphoramidite (0.1 mmol) (ChemGenes) was treated with 3-hydroxypropionitrile (0.2 mmol) in presence of 5-ethylthio-1H-tetrazole (0.2 mmol) in acetonitrile (0.8 mL) for 3 hours. The reaction mixture was evaporated and then 6 mL of 0.02 M I_2 in tetrahydrofuran/pyridine/water was added. After 10 minutes, 1.0 mL of 5 % aq. Na_2SO_3 was added and then the reaction mixture was evaporated. After the residue was co-evaporated with toluene twice, 80% AcOH was added and stirred for 20 min. The reaction mixture was evaporated and co-evaporated with toluene three times. To the residue, conc. NH_4OH:ethanol (3:1) was added and stirred at 55 ºC overnight. After evaporation, 0.3 mL of DMSO and 0.3 mL of TEA•3HF were added and stirred at 55 ºC for 3 hours. The reaction mixture was evaporated and then dissolved in water. The solution was washed with ethyl acetate (EtOAc).

The following extinction coefficients at 260 nm were used to quantitate the nucleoside 3'-phosphates from the HPLC trace, estimated at pH 7.0.

15400, A-3'-phosphate (4 exemplars in the hachimoji transcript), relative absorbance 62600
11700, G-3'-phosphate (8 exemplars in the hachimoji transcript), relative absorbance 93600
7300, C-3'-phosphate (1 exemplars in the hachimoji transcript), relative absorbance 7300
8800, U-3'-phosphate (10 exemplars in the hachimoji transcript), relative absorbance 88000
6300, **S**-3'-phosphate (1 exemplar in the hachimoji transcript), relative absorbance 6300

The reference peaks show significant variance, of course. However, the average using C+A, U, and G as three independent internal standards indicates that 1.2 ± 0.4 equivalents of **S** are incorporated per transcript prepared by the FAL variant of T7 RNA polymerase; no **S** was detectable by HPLC using the wild-type T7 RNA polymerase. This indicates the value of protein engineering to create hachimoji transcripts.

Peak	Expected	Expected ratio (S = 1)	Observed peak	Observed ratio	Amount of **S** calculated using the index peak as reference
C+A	69900	11	0.065	16	0.68
G	93600	15	0.037	9	1.7
U	88000	14	0.043	11	1.3
average					1.2 ± 0.4 equivalents of **S** per transcript

HPLC analysis of hachimoji spinach containing **S**, **B**, **Z** and **P**.

Notwithstanding the overlapping of peaks, HPLC proved to be a useful way to identify hachimoji **Z**-3-phosphate in RNase T2 digests. **Figure S10** shows an HPLC analysis of hachimoji spinach RNA transcripts containing hachimoji **S**-3'-phosphate, **P**-3'-phosphate, **B**-3'-phosphate, and **Z**-3'-phosphate. Possibly because **Z** has a nitro group, it comes late in the ion exchange HPLC (ammonium bicarbonate gradient, as before).

B-3'-phosphate as authentic standard was prepared as described below. To a solution of 3'-AMP (5 μmol) (Santa Cruz Biotechnology) in aqueous saturated $NaHCO_3$ (12.5 μL) was added meta-chloroperbenzoic acid in MeOH (1.2 M, 25 μL). The mixture was stirred for 6 h at room temperature; H_2O (125 μL) was then added. After 10 min at 0 °C, precipitates were removed by filtration through a syringe filter. The filtrate was purified by DNA-pak ion exchange HPLC (eluted by a linear gradient of 0–0.5M ammonium bicarbonate buffer) to yield the fractions containing the desired nucleoside 3'-monophosphate N-oxide. The fractions were collected and concentrated *in vacuo*. 3'-AMP N-oxide (45 mg, 123 μmol from 144 μmol of 3'-AMP) was obtained. This (123 μmol, 45 mg) was dissolved in water (2 mL) and the solution was placed in a quartz cuvette. The mixture was irradiated for 16 h (loss of absorbance at 232 nm indicated completion of reaction). The solution was filtered through a 0.45 micron filter and the material was purified by HPLC (DNA-pak ion-exchange column). The product was eluted by a linear gradient of 0–0.5M ammonium bicarbonate buffer to yield the fractions containing the desired IsoG nucleoside 3'-monophosphate. The product was 3'-isoGMP (3.5 mg, 9.6 μmol from 123 μmol of 3'-AMP N-oxide).

Z-3'-phosphate as authentic standard was prepared as described below. Protected O-NPE N-acetyl 5'-DMT 2'-thiocarbamoylpyrrolidine sulfone r**Z** phosphoramidite (0.1 mmol) (*29*) was treated with 3-hydroxypropionitrile (0.2 mmol) in presence of 5-ethylthio-1H-tetrazole (0.2 mmol) in acetonitrile (0.8 mL) for 3 hours. The reaction mixture was evaporated and then 6 mL of 0.02 M iodine in THF/pyridine/water was added. After 10 minutes, 1.0 mL of 5 % aq. Na_2SO_3 was added and then the reaction mixture was evaporated. After the residue was co-evaporated with toluene twice, 80% AcOH was added and stirred for 20 min. The reaction mixtures were evaporated and co-evaporated with toluene three times. The residue was treated with 1.0 M DBU in acetonitrile (2 mL) overnight. After evaporation, the residue was treated with 2.0 mL of ethylenediamine for 2 hours. The reaction mixture was evaporated and then dissolved in water. The solution was washed with EtOAc.

P-3'-phosphate as authentic standard was prepared as described below. Protected N-dibutylformamidine 2'-thiocarbamoylpyrrolidine sulfone r**P** phosphoramidite (0.1 mmol) (*29*) was treated with 3-hydroxypropionitrile (0.2 mmol) in presence of 5-ethylthio-1H-tetrazole (0.2 mmol) in acetonitrile (0.8 mL) for 3 hours. The reaction mixture was evaporated and then 6 mL of 0.02 M Iodine in THF/pyridine/water was added. After 10 minutes, 1.0 mL of 5 % aq. Na_2SO_3 was added and then the reaction mixture was evaporated. After the residue was co-evaporated with toluene twice, 80% AcOH was added and stirred for 20 min. The reaction mixtures were evaporated and co-evaporated with toluene three times. The residue was treated with 2.0 mL of ethylenediamine for 2 hours. The reaction mixture was evaporated and then dissolved in water. The solution was washed with EtOAc.

The **Z**-3'-phosphate comes late, as confirmed by its co-elution with authentic material.

At 260 nm, **B**-3'-phosphate gives a relatively weak signal; its λ_{max} is at 290 nm; 260 nm is close to an absorbance minimum. Of course, **B** can affirmatively identified in this digest because of this unusual UV absorbance property, as well as its co-elution with authentic **B**-3'-phosphate

made by chemical synthesis. **S**-3'-phosphate and **P**-3'-phosphate co-elute by HPLC (as do A-3'-phosphate and C-3'-phosphate). Thus, the digestion product mixture shows all eight hachimoji nucleotides in the transcript made by the FAL variant of T7 RNA polymerase. However, the presence of A or C and **S** or **P** remains incompletely established, due to co-elution in the HPLC. This required us to develop alternative tools to do analysis of hachimoji RN, as described below.

Label shift analysis

Classical two dimensional thin layer chromatography (2D-TLC) was adapted to allow analysis of hachimoji RNA. In these assays, one of four standard RNA triphosphates is introduced into a transcription mixture with an alpha-^{32}P label. This leads to a product with a bridging ^{32}P phosphate. Subsequent hydrolysis by RNase T2 generates a mixture of nucleoside 3'-phosphates, where the 3'-nucleotide immediately preceding in the sequence carries a ^{32}P-label. The mixture of nucleoside 3'-phosphates is then resolved by chromatography to determine the adjacency patterns of the system.

Thus, in the various model products, incorporation of alpha-^{32}P labeled CTP will generate only a labeled **S**-3'-^{32}P product, only **Z**-3'-^{32}P product, only **P**-3'-^{32}P product, or only **B**-3'-^{32}P product, depending on the template. These are resolved by chromatography.

Following incubation with T7 RNA polymerase, transcripts were PAGE purified, digested with T2 ribonuclease (100 units) (Worthington Biochemical Corporation) by incubation at 37°C for 16 hours in 15 mM sodium acetate buffer pH 4.5. For each sample, approximately 1500 cpm was spotted on polyethyleneimine plates (PEI: 20 cm x 20 cm plates) (Sorbent Technologies), dried under flowing air, and run in the first dimension with first dimension buffer. The plate was then dried, turned 90°, and then run in the second dimension buffer. The plates were dried, exposed to a phosphorimager screen and analyzed with a Biorad Personal Molecular Imager (PMI) System.

As a primary solvent system for two dimensional thin layer chromatography, isobutyric acid-ammonia-water (66:1:33 v/v/v) for the first dimension and 2-propanol-HCl-water (70:15:15 v/v/v) for the second dimension were used.

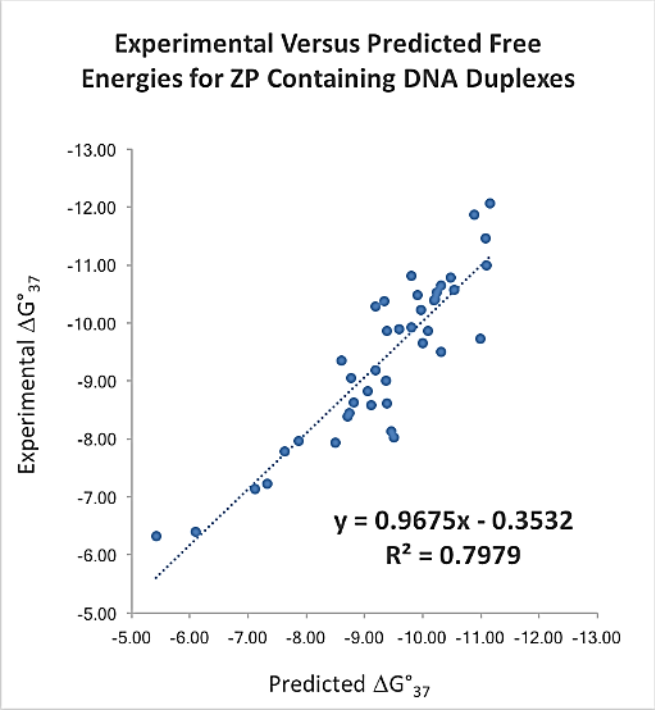
As a secondary solvent system for two dimensional thin layer chromatography, isobutyric acid-ammonia-water (66:1:33 v/v/v) for the first dimension and aqueous saturated ammonium sulfate (~140 g) with ammonium bisulfate (~ 0.8 g) in 200 mL, pH 3.2 for the second dimension were used.

Methods relating to crystal structure determination

Crystallization, data collection, structure determination, and refinement of Hachimoji DNA host-guest complexes. The N-terminal fragment of Moloney murine leukemia virus reverse transcriptase was expressed and purified as previously described *(32)*. **PB**, **P**C, and **PP** 16 mer DNA oligonucleotides were resuspended in 10 mM HEPES (pH 7.0) with 10 mM $MgCl_2$ to give final concentrations of 5 mM (2.5 mM duplex) and annealed at 70 °C for 10 min and then gradually cooled to room temperature. A 3.2 mM stock solution of the protein was diluted to 1.4 mM using 50 mM MES (pH 6.0) and 0.3 M NaCl and then further diluted to 0.65 mM in 100 mM HEPES (pH 7.5) and 0.3 M NaCl. The 0.65 mM stock of CF24 was used to prepare protein-DNA complexes. Protein-DNA complexes were prepared at a ratio of 1: 2 respectively (0.43 mM protein : 0.86 mM DNA) for **PB**, **P**C, and **PP** as well as a control oligonucleotide (5'-CTTAATTCGAATTAAG) and incubated on ice for 1 h. Crystals of host-guest complexes were grown using the hanging drop vapor diffusion method and ultimately microseeded using crystals

that had self-nucleated in initial screens. Drops contained 1 μL of precipitant or precipitant with microseeds from host-guest crystals and 1 μL protein-DNA complex in reservoir containing 8 % PEG 4000, 5 mM magnesium acetate and 50 mM ADA (pH 6.5). For data collection, crystals were cryoprotected in 9 % PEG 4000, 5 mM magnesium acetate, 100 mM HEPES (pH 7.5) and 20 % ethylene glycol before flash freezing in liquid nitrogen.

Host-guest complex datasets for **PB**, **P**C, and **PP** were collected at the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, beamline BL14-1 using a Dectris EIGER 16 M detector to 1.7 Å, 1.6 Å, and 1.7 Å, respectively, by Accelero Biostructures Inc. All three host-guest complexes crystallized in space group P21212 with one protein molecule and one half of the 16mer duplex oligonucleotide in the asymmetric unit of the crystal. The datasets were processed, scaled, and merged using XDS *(33)* and AIMLESS *(34)*. PHASER *(35)* was used for molecular replacement using the protein model from PDB ID 4XO0 as the search model. Thus, electron density for DNA was unbiased in initial electron density maps. Initial model fitting and addition of waters was done in COOT *(36)* followed by refinement in REFMAC *(37)*. The DNA models were built in COOT starting with the terminal 3 base pairs in the structures and refined in REFMAC. Geometric restraint files were created in eLBOW-PHENIX for **P**, **Z**, **S**, and **B**. Once the unnatural base pairs were modeled in COOT, an iterative process of refinement in PHENIX *(38)* and model building in COOT was used to complete the model. Ultimately, the model was refined as an intact single-strand 16-mer to ensure that the DNA backbone was continuous from which the 8mer duplex model was then created. Coordinates have been deposited in the PDB, **PB** (6MIG), **P**C (6MIH), and **PP** (6MIK).



Dimer	$\Delta G°_{37}$ kcal/mol	σ
A**P**/T**Z**	-1.66	0.10
A**Z**/T**P**	-1.62	0.10
C**P**/G**Z**	-2.31	0.10
C**Z**/G**P**	-2.26	0.09
G**P**/C**Z**	-2.10	0.11
G**Z**/C**P**	-1.77	0.11
T**P**/A**Z**	-1.40	0.08
T**Z**/A**P**	-1.73	0.09
PP/**ZZ**	-2.28	0.14
ZP/**PZ**	-2.88	0.23
PZ/**ZP**	-1.61	0.22
Init.w/Term **ZP**	+1.48	0.08

Dimer	$\Delta H°$ kcal/mol	σ
A**P**/T**Z**	-5.6	1.3
A**Z**/T**P**	-6.7	1.4
C**P**/G**Z**	-9.1	1.3
C**Z**/G**P**	-8.6	1.2
G**P**/C**Z**	-6.1	1.3
G**Z**/C**P**	-2.7	1.3
T**P**/A**Z**	-5.5	1.2
T**Z**/A**P**	-9.8	1.3
PP/**ZZ**	-11.5	1.1
ZP/**PZ**	-10.5	2.9
PZ/**ZP**	-6.5	2.8
Init.w/Term **ZP**	+5.5	1.1

Fig. S1.

Plot of the experimental free energy changes, $\Delta G°_{37}$, versus the predicted free energy changes for all 41 duplexes in this study (data in **Table S3** and **S4**). NN parameters and standard errors, σ, for **Z-P** containing NN dimers were derived by SVD and standard error propagation.

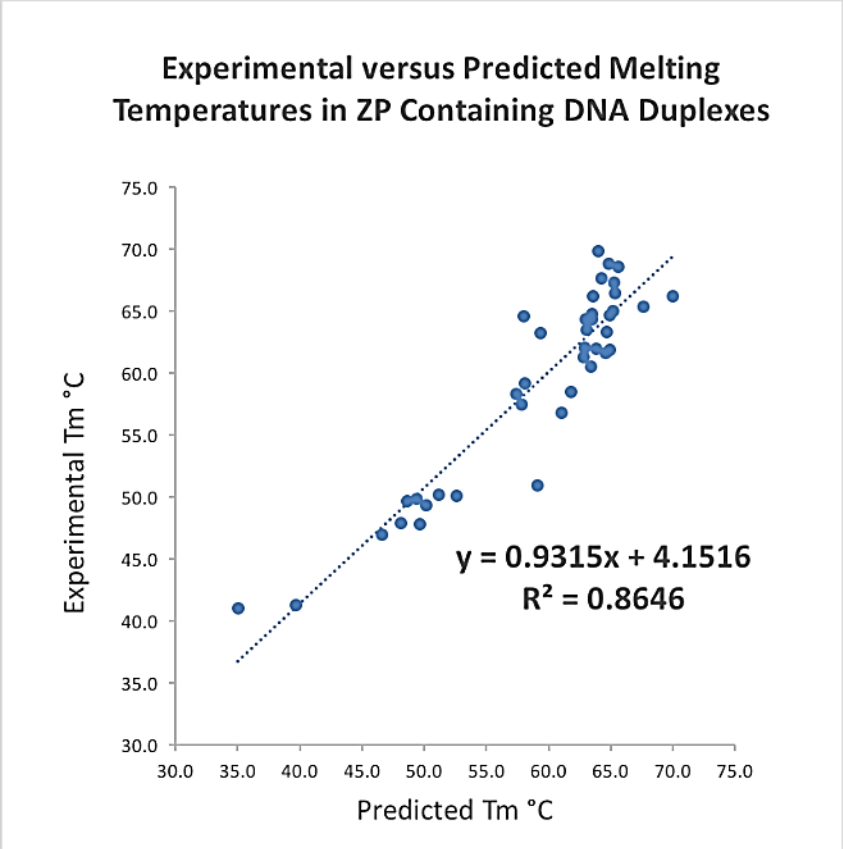
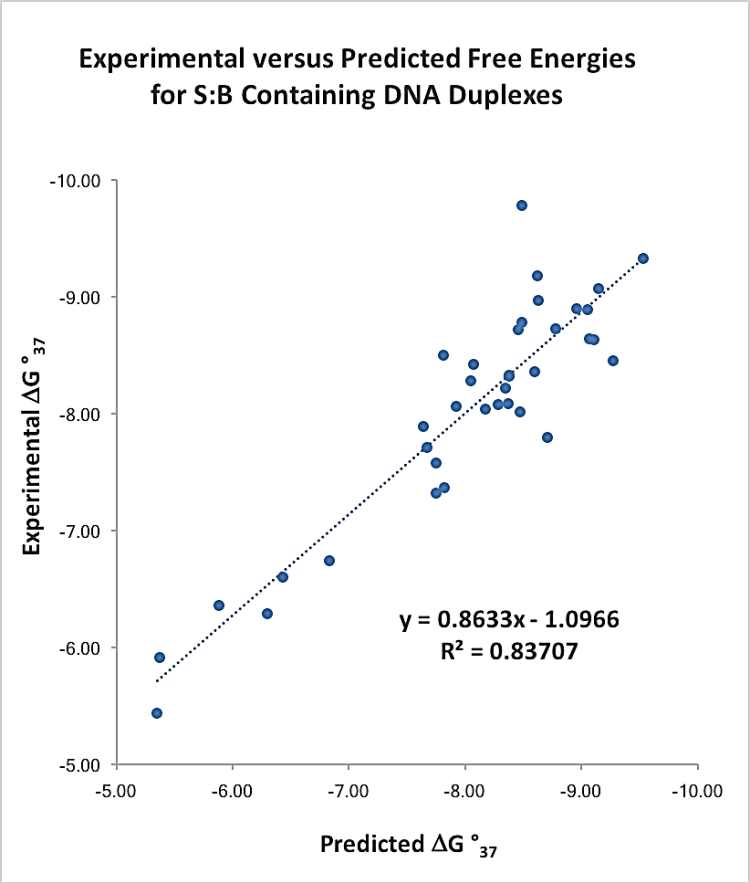


Fig. S2.

Plot of the experimental melting temperatures vs. the predicted T_m's for 41 **Z-P** containing DNA duplexes (data in **Table S3** and **S4**). These data were calculated as described in the Methods. All T_m's were calculated using a total oligonucleotide concentration of $1x10^{-4}$ M.



Dimer	$\Delta G°_{37}$ kcal/mol	σ
A**B**/T**S**	-1.24	0.09
A**S**/T**B**	-1.30	0.09
C**B**/G**S**	-1.92	0.09
C**S**/G**B**	-1.44	0.09
G**B**/C**S**	-1.70	0.10
G**S**/C**B**	-1.78	0.10
T**B**/A**S**	-1.37	0.08
T**S**/A**B**	-1.36	0.08
BB/**SS**	-1.79	0.08
SB/**BS**	-1.68	0.20
BS/**SB**	-1.67	0.20
Init.w/Term **SB**	1.48	0.07

Dimer	ΔH° kcal/mol	σ
A**B**/T**S**	-7.5	1.3
A**S**/T**B**	-5.3	1.3
C**B**/G**S**	-7.3	1.1
C**S**/G**B**	-7.0	1.1
G**B**/C**S**	-7.1	1.3
G**S**/C**B**	-6.2	1.3
T**B**/A**S**	-7.6	1.1
T**S**/A**B**	-7.8	1.1
BB/**SS**	-5.0	0.9
SB/**BS**	-4.3	2.6
BS/SB	-5.6	2.7
Init.w/Term **SB**	5.9	1.0

Fig. S3.

Plot of the experimental free energy changes, $\Delta G°_{37}$, versus the predicted free energy changes for all 37 duplexes in this study (data in **Table S6** and **S7**). NN parameters and standard errors, σ, for **S**-**B** containing NN dimers were derived by SVD and standard error propagation.

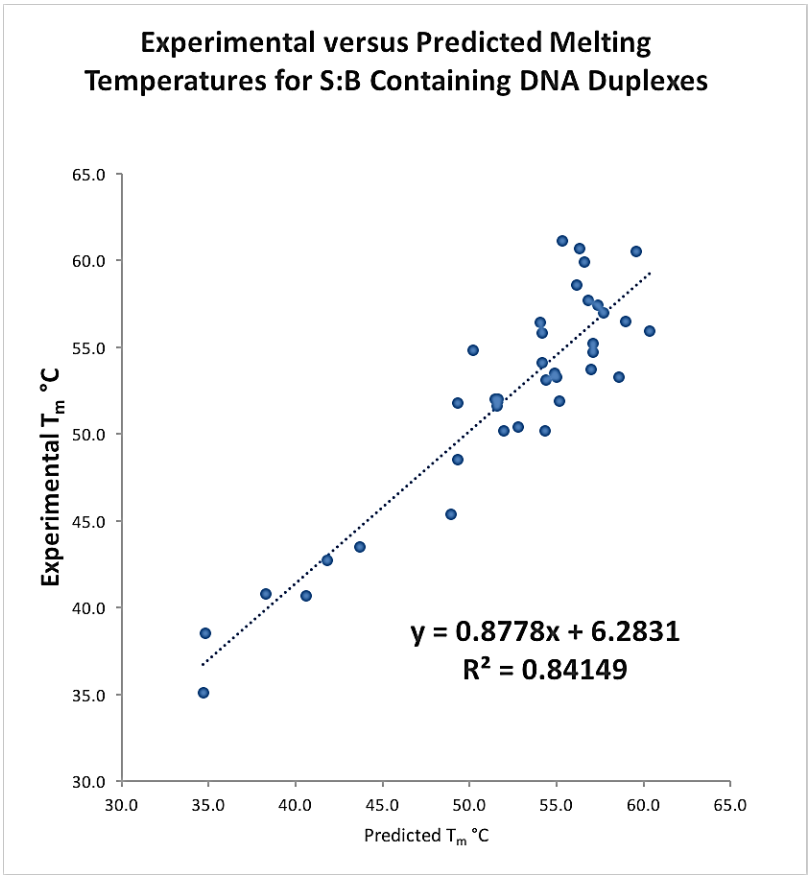
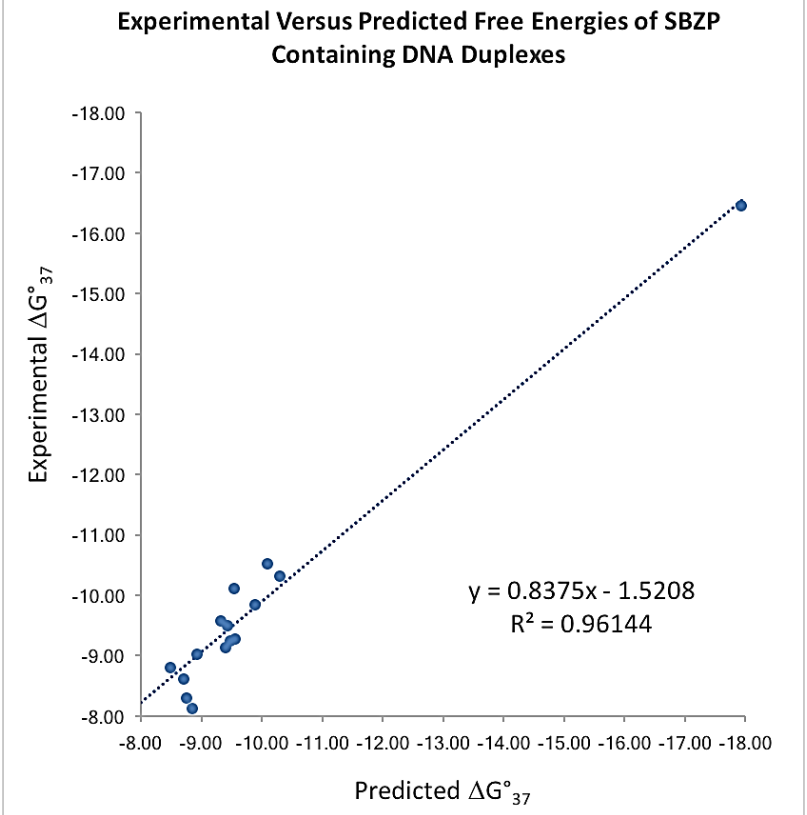


Fig. S4.

Plot of the experimental melting temperatures vs. predicted melting temperatures for all 37 **S**-**B** containing DNA duplexes (data in **Table S6** and **S7**). These data were calculated as described in the Methods. All Tm's were calculated using a total oligonucleotide concentration of $1x10^{-4}$ M.



Dimer	$\Delta G°_{37}$ kcal/mol
SZ/BP	-2.29 ± 0.11
SP/BZ	-2.29 ± 0.08
ZS/PB	-1.86 ± 0.09
BZ/SP	-1.90 ± 0.12
Dimer	$\Delta H°$ kcal/mol
SZ/BP	-8.7 ± 1.1
SP/BZ	-9.8 ± 0.9
ZS/PB	-6.5 ± 1.0
BZ/SP	-4.5 ± 1.2

Fig. S5.

Experimental vs. predicted free energies of **SBZP**-containing hachimoji DNA duplexes. Plotted are experimental free energy changes ($\Delta G°_{37}$) versus predicted free energy changes for all 15 duplexes in this study (data in **Tables S9** and **S10**). Parameters for dinucleotide pairing affinity and standard errors, σ, for dinucleotides containing **P** and **Z** dimers were derived by singular value decomposition and standard error propagation.

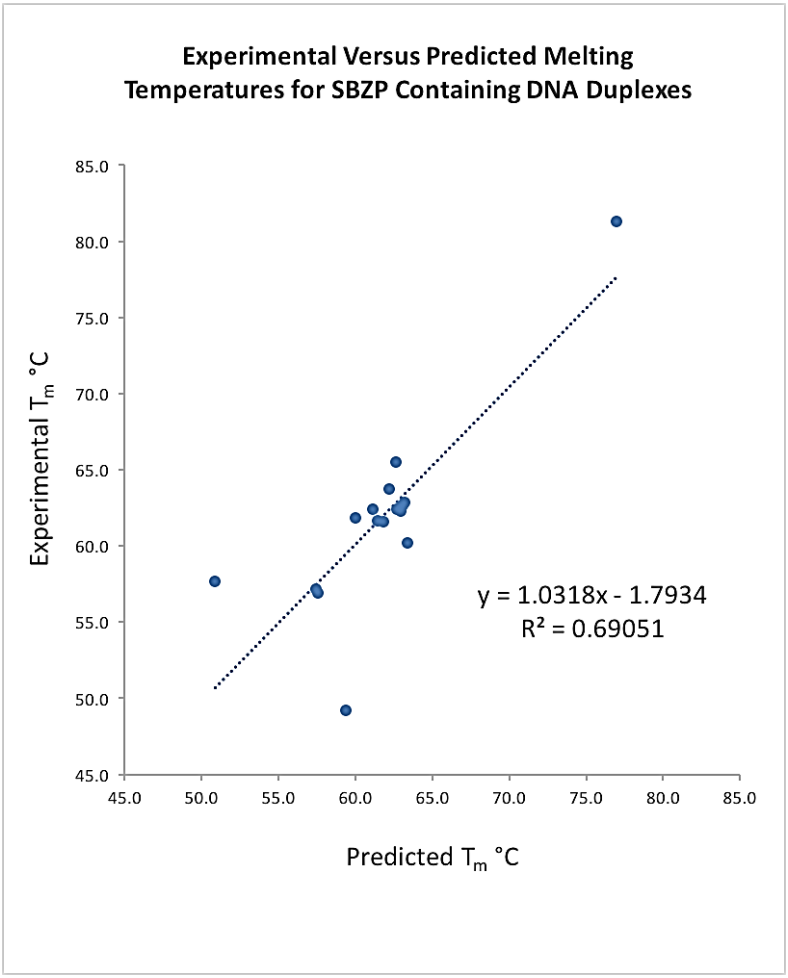


Fig. S6.

Plot of experimental melting temperatures vs. predicted melting temperatures for all 15 **S-B** and **Z-P** dinucleotides in DNA duplexes (data in **Table S9** and **S10**), calculated as described in the Methods. All T_m's were calculated using a total oligonucleotide concentration of $1x10^{-4}$ M.

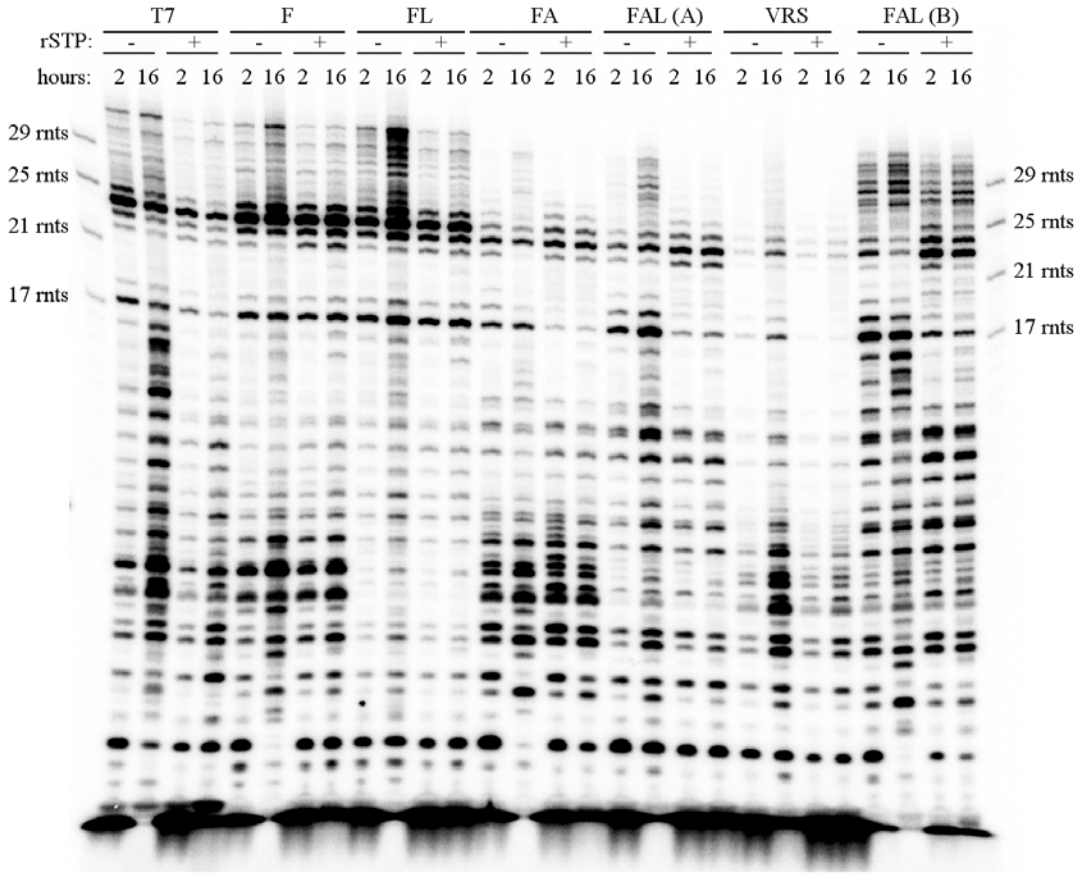


Fig. S7.

PAGE (20%) showing transcription products with internal labeling. Wild type and mutant T7 RNA polymerases were tested in the absence and presence of r**S**TP for their ability to generate the RNA product T2**S**; they show different levels of pausing and rescue. Full length product is a 24mer, **S** is at position 18; pausing is most prominent at position 17. T7, FA and FAL mutants show pausing in the absence of r**S**TP and various levels of rescue in the presence of r**S**TP.

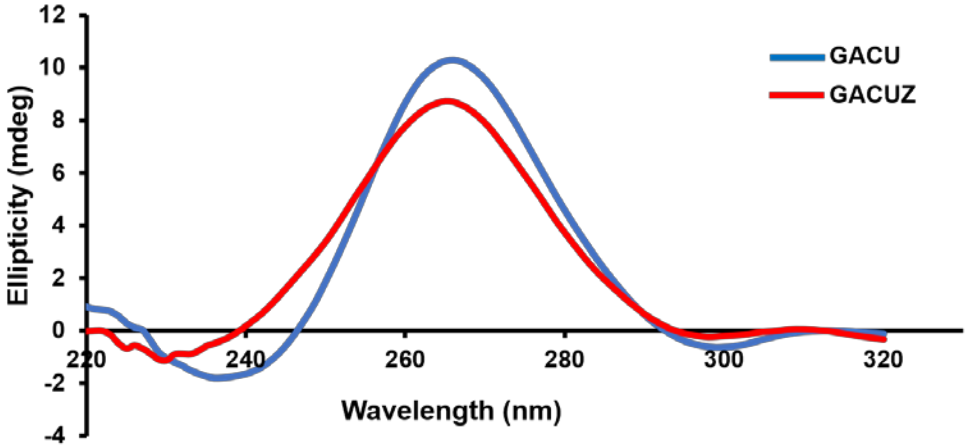


Fig. S8.

CD spectra of native Spinach (blue) and a variant with **Z** at position 50, near the fluor binding sites, with further changes at position 53 (A to G) and position 29 (U to C) to restore the base triple seen in the crystal structure. A maximum at ~260 nm often indicates the presence of a parallel G-quadruplex in RNA *(31)*. These similar CD signatures suggest that all variants locally fold into a G-quadruplex motif, which is essential for fluorophore binding and fluorescence activation.

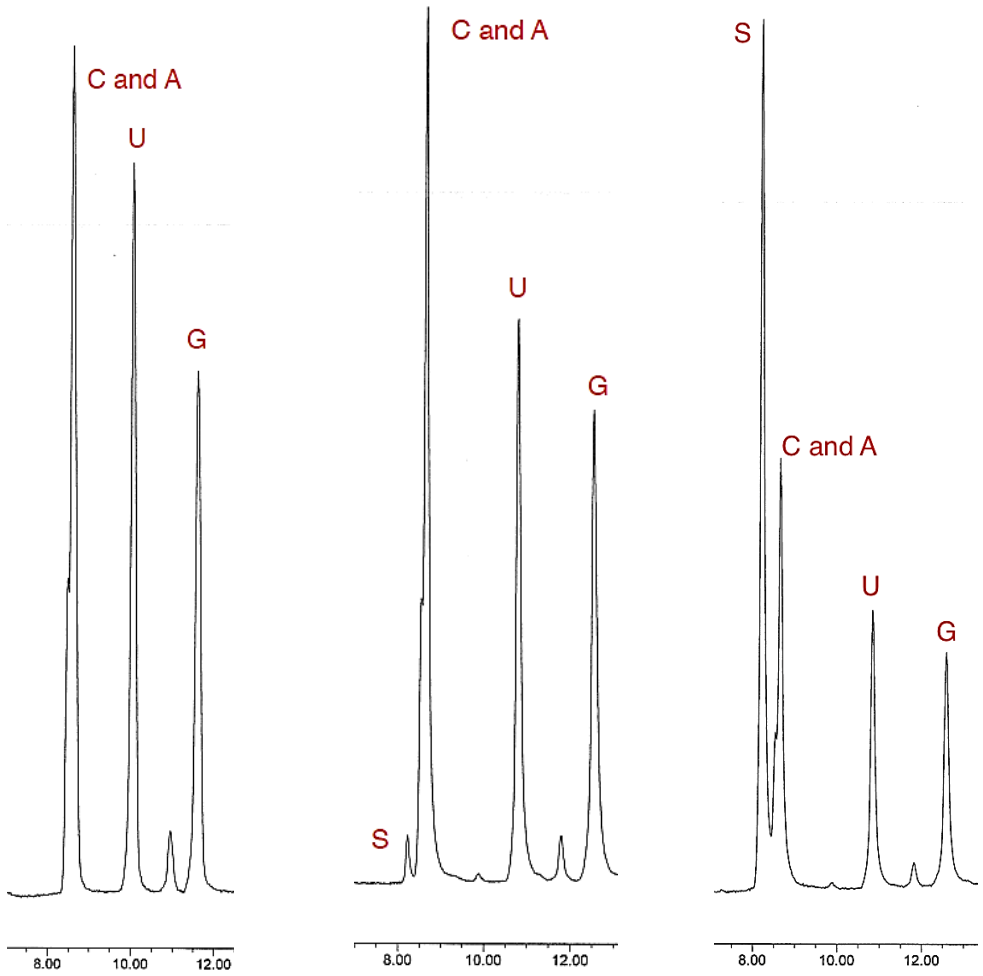


Fig. S9.

HPLC (ammonium bicarbonate, 0 to 200 mM) traces of the rN-3'-monophosphates recovered by RNase T2 digestion of the RNA made by attempts with different RNA polymerase variants to make hachimoji spinach. (**Left**) Trace from RNA made via transcription using wild-type T7 RNA polymerase. Note absence of **S**-3'-P. (**Center**) Trace from RNA made via transcription using the FAL variant of T7 RNA polymerase. Note detectable presence of **S**-3'-P, notwithstanding its low extinction coefficient and its expected presence in the transcript as only one exemplar. (**Right**). Trace from RNA made via transcription using the FAL variant T7 RNA polymerase, with co-injection of the authentic r**S**-3'-monophosphate. The expected hachimoji transcript is: GGG AGU GUU GUA UUU GG**S** CAA UUU, with one **S** relative to 5 {A+C}, 8 G, and 10 U. Using the extinction coefficients above, 1.2 ± 0.4 **S** nucleotides were incorporated into the transcript by the FAL variant of T7 RNA polymerase.
The increased intensity of the U peak relative to the others in the left trace (made with wild-type T7 RNA polymerase) is consistent with a model whereby wild-type T7 RNA polymerase incorporates U opposite template d**B**, due to the presence of a minor tautomer of d**B** that is Watson-Crick complementary to U. This would generate a transcript with 11 U's instead of 10 U's. The peak at 11.5 min is unassigned.

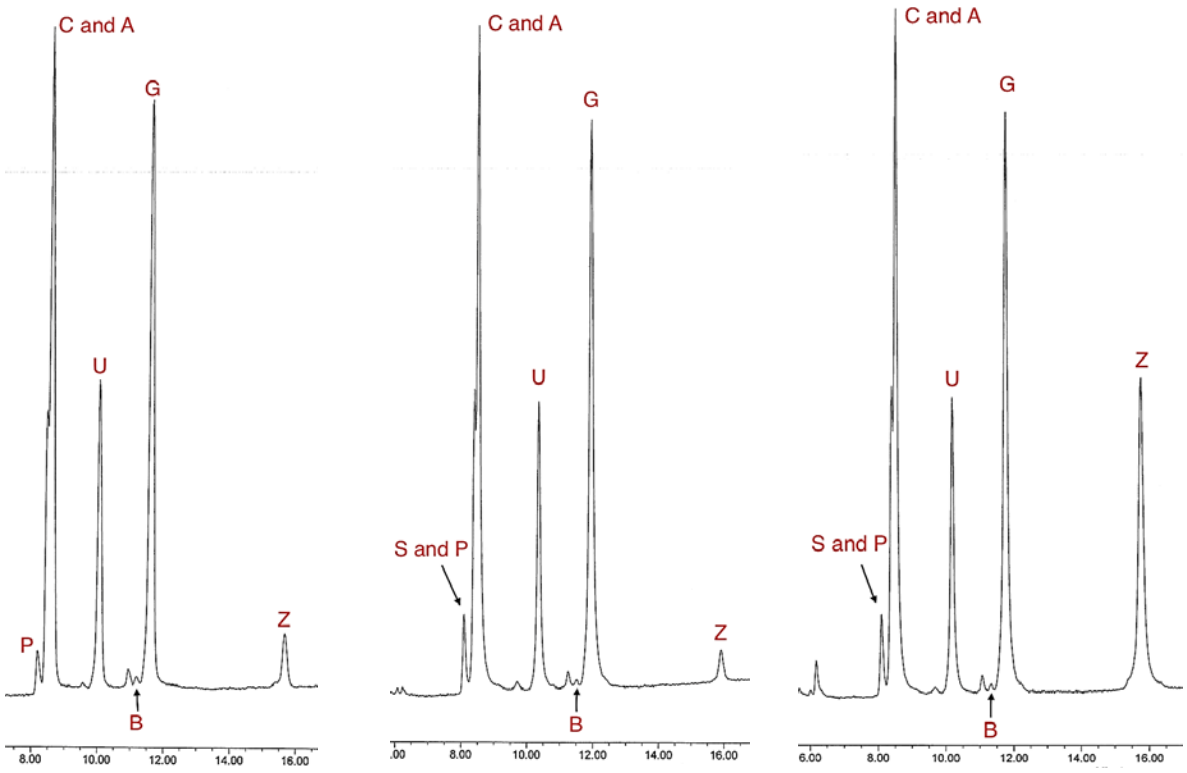


Fig. S10.

HPLC trace (ammonium bicarbonate, 0 to 200 mM) of rN-3'-monophosphates recovered by RNase T2 digestion of the spinach aptamer made by transcription of a hachimoji template. (**Left**) Products from the aptamer made by wild-type T7 RNA polymerase; it does not contain S-3'-**P**, as confirmed by TLC. (**Center**) Products from the aptamer made by the FAL variant of T7 RNA polymerase containing all eight hachimoji components (G, A, C, T, **Z**, **P**, **S**, and **B**). (**Right**) Products from the aptamer made by the FAL variant of T7 RNA polymerase with co-injection of the authentic r**Z**-3'-monophosphate.

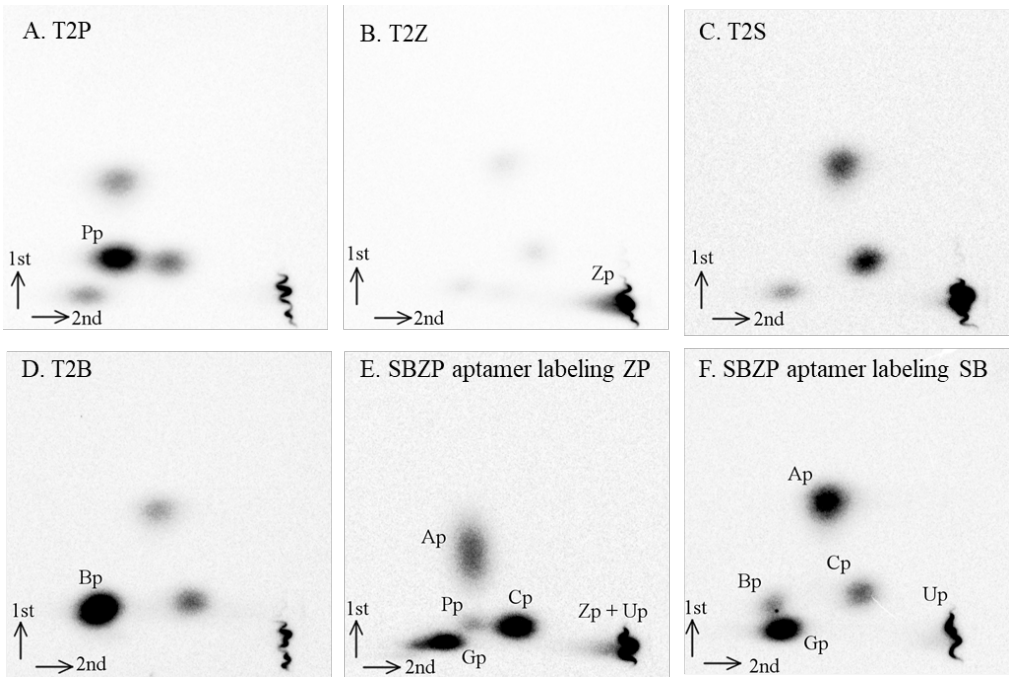


Fig. S11.

2D-TLC of RNase T2 digests of labeled test sequences (panels A-D) and spinach (panels E and F) made with wild-type T7 RNA polymerase in primary solvent system. (A) With template giving a product containing **P** as the only hachimoji non-standard nucleotide, generates **P**-3'-^{32}P (**P**p) after digestion. (B) With template giving a product containing **Z** as the only hachimoji non-standard nucleotide, generates **Z**-3'-^{32}P (**Z**p) after digestion, which runs with U-3'-P. (C) With template that produces a product containing **S** as the only hachimoji non-standard nucleotide, wild-type T7 RNA polymerase apparently does not incorporate **S**TP (absence of **S**-3'- ^{32}P which would run to the right of C-3'-P in this solvent system, shown below). (D) With template giving a product containing **B** as the only hachimoji non-standard nucleotide, generates **B**-3'-^{32}P (**B**p) after digestion. (E) Transcript of the spinach aptamer using alpha-^{32}P-GTP, which nearest neighbor labels all four standard nucleotides, as well as **Z** and **P**. After digestion, evidence of incorporation comes from the appearance of the corresponding **Z**-3'-^{32}P (**Z**p) and **P**-3'-^{32}P (**P**p). Since **Z**-3'-P does not separate convincingly in this system, its presence in the spinach aptamer was confirmed by HPLC (**Figure S10**), and in a second buffer system (shown below). (F) Transcript of the spinach aptamer using alpha-^{32}P-CTP, which nearest neighbor labels all four standard nucleotides, as well as **S** and **B**. After digestion, evidence of incorporation of **B**TP comes from the appearance of the corresponding **B**-3'-^{32}P (**B**p). However, essentially no amount of radioactivity is attributable to **S**-3'-^{32}P. This suggests the need to use a variant of T7 RNA polymerase to allow the preparation of hachimoji RNA from hachimoji DNA by transcription. In addition, a secondary TLC system was required to resolve all eight 3'-phosphates arising from all eight components of the hachimoji system.

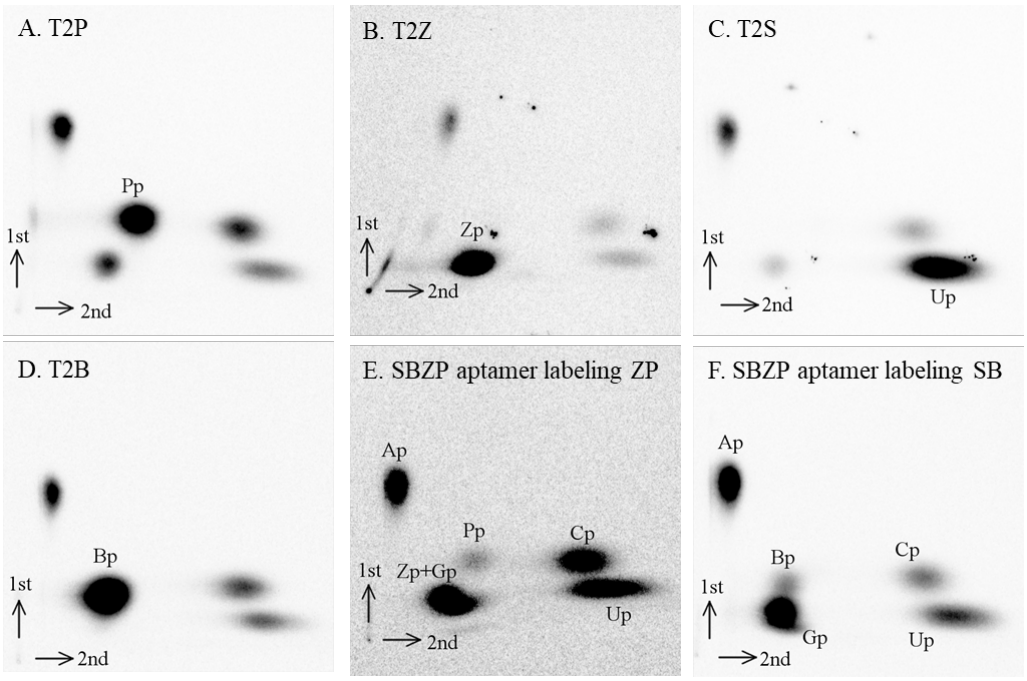


Fig. S12.

2D-TLC of RNase T2 digests of labeled test sequences (panels A-D) and spinach (panels E and F) made with wild-type T7 RNA polymerase in secondary solvent system. (A) With template giving a product containing **P** as the only hachimoji non-standard nucleotide, generates **P**-3'-^{32}P (**P**p) after digestion. (B) With template giving a product containing **Z** as the only hachimoji non-standard nucleotide, generates **Z**-3'-^{32}P (**Z**p) after digestion, which now runs much slower, separate from U-3'P. (C) With template that produces a product containing **S** as the only hachimoji non-standard nucleotide, essentially no amount of radioactivity is attributable to **S**-3'-^{32}P. (D) With template giving a product containing **B** as the only hachimoji non-standard nucleotide, generates **B**-3'-^{32}P (**B**p) after digestion. (E) Transcript of the spinach aptamer using alpha-^{32}P-GTP, which nearest neighbor labels all four standard nucleotides, as well as **Z** and **P**. After digestion, evidence of incorporation comes from the appearance of the corresponding **Z**-3'-^{32}P (**Z**p) and **P**-3'-^{32}P (**P**p). (F) Transcript of the spinach aptamer using alpha-^{32}P-CTP, which nearest neighbor labels all four standard nucleotides, as well as **S** and **B**. After digestion, evidence of incorporation of **B**TP comes from the appearance of the corresponding **B**-3'-^{32}P (**B**p). However, essentially no radioactivity are attributable to **S**-3'-^{32}P. This again suggests the need to use a variant of T7 RNA polymerase to allow the preparation of hachimoji RNA from hachimoji DNA by transcription.

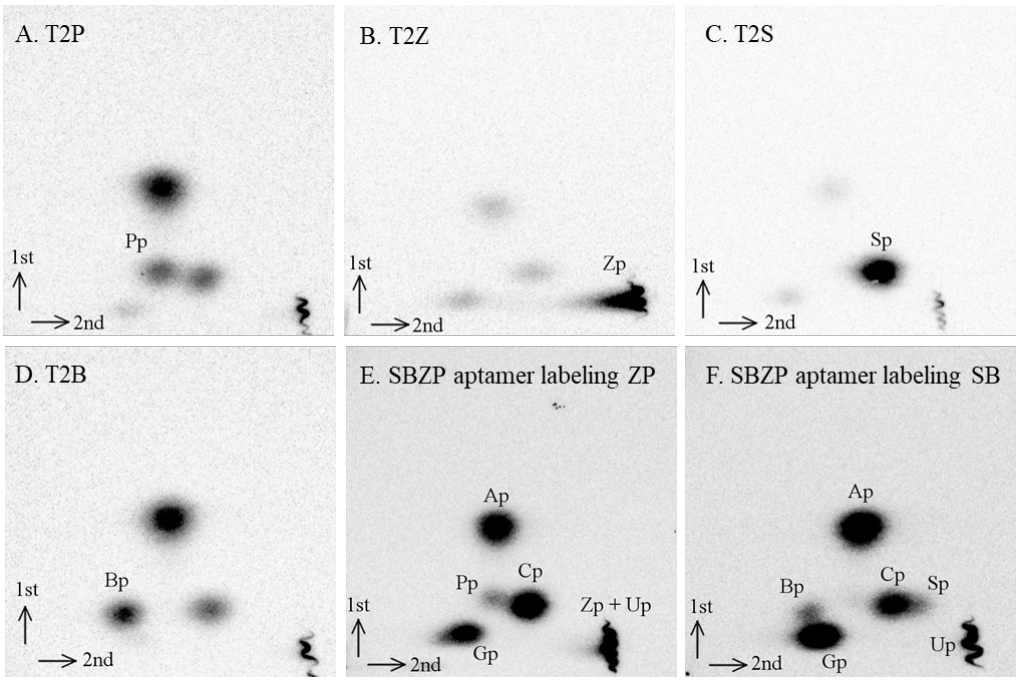


Fig. S13.

2D-TLC of RNase T2 digests of labeled test sequences (panels A-D) and spinach (panels E and F) made with FAL variant of T7 RNA polymerase in primary solvent system. (A) With template giving a product containing **P** as the only hachimoji non-standard nucleotide, generates **P**-3'-^{32}P (**P**p) after digestion. (B) With template giving a product containing **Z** as the only hachimoji non-standard nucleotide, generates **Z**-3'-^{32}P (**Z**p) after digestion, which runs with U-3'-P. (C) With template that produces a product containing **S** as the only hachimoji non-standard nucleotide, **S**-3'- ^{32}P is now clearly present. (D) With template giving a product containing **B** as the only hachimoji non-standard nucleotide, generates **B**-3'-^{32}P (**B**p) after digestion. (E) Transcript of the spinach aptamer using alpha-^{32}P-GTP, which nearest neighbor labels all four standard nucleotides, as well as **Z** and **P**. After digestion, evidence of incorporation comes from the appearance of the corresponding **Z**-3'-^{32}P (**Z**p) and **P**-3'-^{32}P (**P**p). Since **Z**-3'-P does not separate convincingly in this system, its presence in the spinach aptamer was confirmed by HPLC (**Figure S10**). (F) With the spinach aptamer labeled with C-alpha-^{32}P-triphosphate, label of G, A, C, U, **S**, and **B** is expected. All six spots are seen in the amounts approximately as expected.

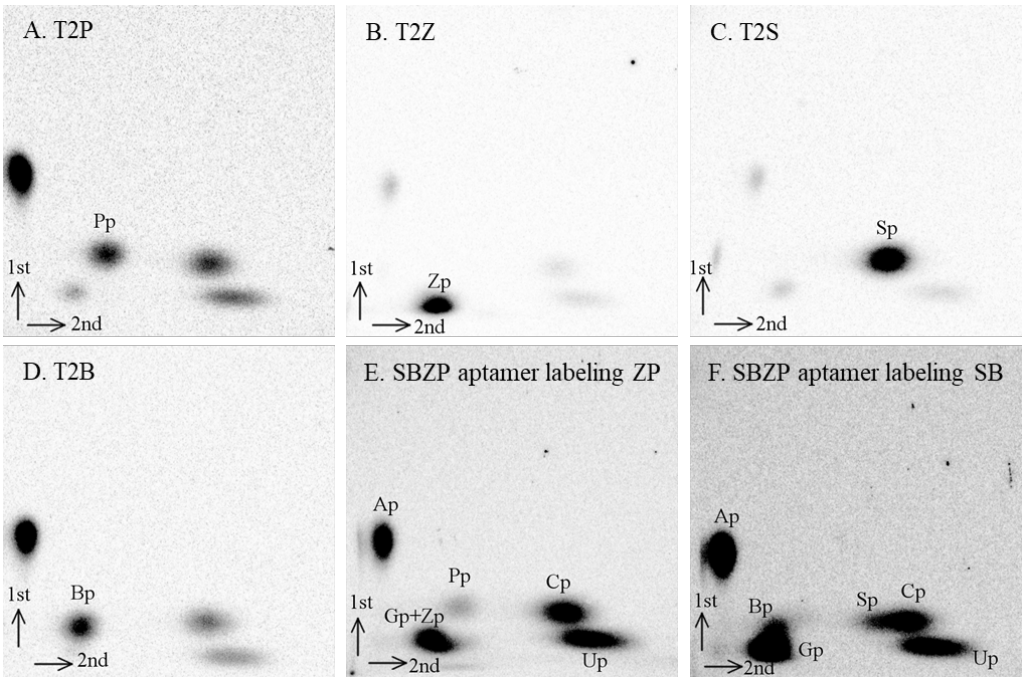


Fig. S14.

2D-TLC of RNase T2 digests of labeled test sequences (panels A-D) and spinach (panels E and F) made with FAL variant of T7 RNA polymerase in secondary solvent system. (A) With template giving a product containing **P** as the only hachimoji non-standard nucleotide, generates **P**-3'-^{32}P (**P**p) after digestion. (B) With template giving a product containing **Z** as the only hachimoji non-standard nucleotide, generates **Z**-3'-^{32}P (**Z**p) after digestion. (C) With template that produces a product containing **S** as the only hachimoji non-standard nucleotide, **S**-^{32}P-phosphate is again clearly present. (D) With template giving a product containing **B** as the only hachimoji non-standard nucleotide, generates **B**-3'-^{32}P (**B**p) after digestion. (E) Transcript of the spinach aptamer using alpha-^{32}P-GTP, which nearest neighbor labels all four standard nucleotides, as well as **Z** and **P**. After digestion, evidence of incorporation comes from the appearance of the corresponding **P**-3'-^{32}P (**P**p). **Z**-3'-^{32}P (**Z**p) running near G-3'-P; its incorporation was confirmed by HPLC (**Figure S10**). (F) With the spinach aptamer labeled with alpha-^{32}P-CTP, label of G, A, C, U, **S**, and **B** is expected. All six spots are seen with **B**-3'-^{32}P (**B**p) running above G-3'-P and **S**-3'-P running to the left of C-3'-P.

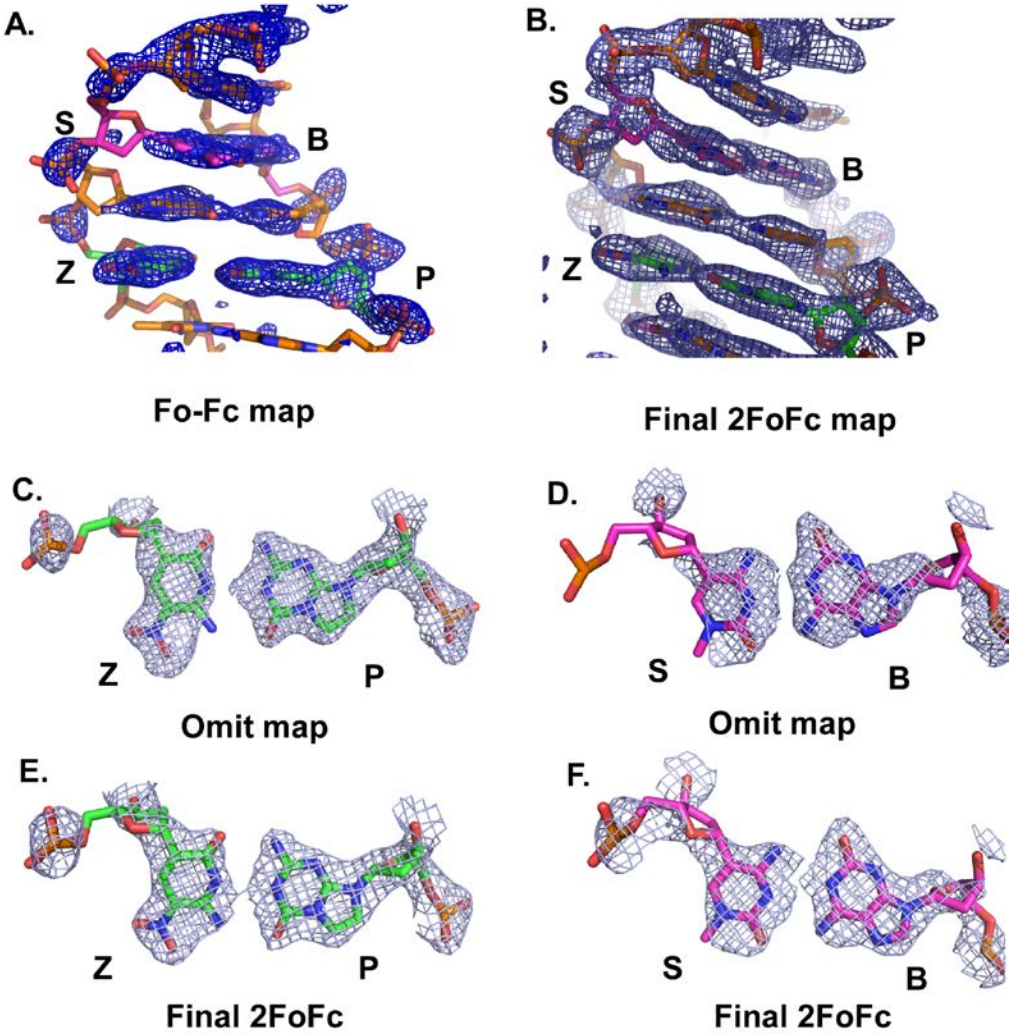


Fig. S15.

Representative electron density maps for **Z**:**P** and **S**:**B** pairs in the PC structure. (**A)** An Fo-Fc electron density map phased from a model including the protein model and the first 4 base pairs is shown contoured at 3 sigma in blue mesh and superimposed on the final refined model for the last 4 bp of the unique 8-mer DNA duplex starting with the **Z:P** pair. The initial molecular replacement phasing for the structure were calculated using the protein model alone. The DNA model was built into unbiased difference electron density maps. At the contour level shown, strong density is apparent for the phosphate groups and the nucleobases as expected. (**B)** The final 2Fo-Fc electron density map contoured at 1 sigma in blue mesh is shown for the last 4 bp shown in (A) with the final refined model. Simulated annealing omit maps (2mFo-DFc) for the PC structure were calculated in PHENI1(*38, 39*) omitting nucleobase pairs **Z5:P12** contoured at 1.5 sigma in blue mesh (**C**) or **B7:S10** contoured at 1.4 sigma in blue mesh (**D**). Final 2Fo-Fc electron density maps for **Z5:P12** contoured at 1.4 sigma in blue mesh (**E**) or **B7:S10** contoured at 1.3 sigma in blue mesh (**F**). Hydrogen bonding distances in the **Z:P** pair are between 2.8 -3.0 Å (**G**) and for **S:B** between 2.9-3.3 Å (**H**). The longer hydrogen bond in the **S:B** pair between atoms O2 and N2 in the PC structure is also found in the PB structure in the same position. In the **S8:B9** pair in the PP structure, the distance between the O2 and N2 atoms is 3.86 Å. The **Z:P** pairs in the PP structure exhibit a longer hydrogen bond (3.3 Å) between N4 and O6 atoms.

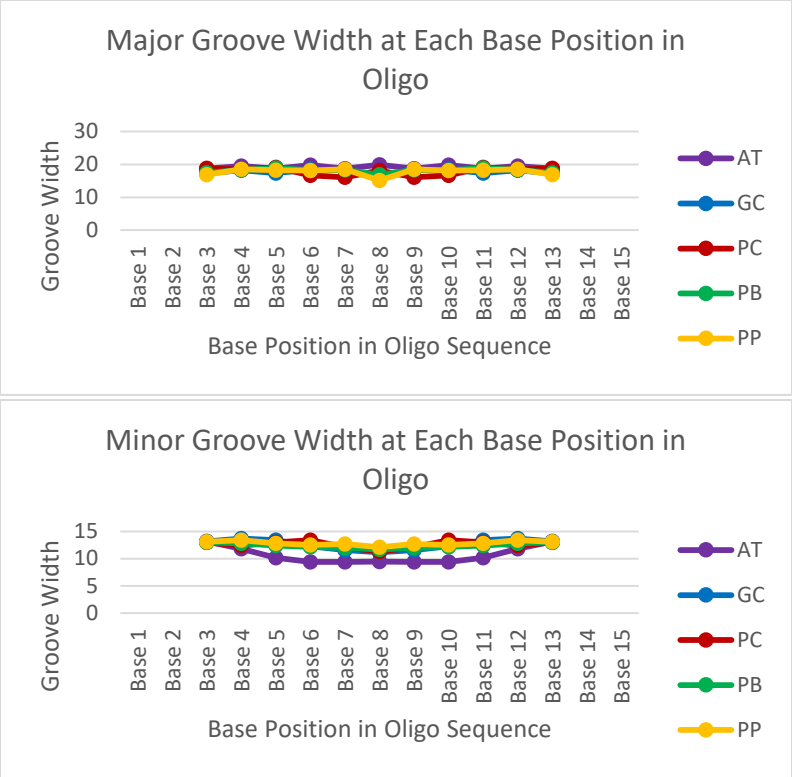


Fig. S16.

Comparison of major and minor groove widths calculated in 3DNA for hachimoji DNA structures **PB**, **P**C, and **PP** and natural AT and GC structures. The groove widths for hachimoji DNA are similar to that of the natural GC structure.

Table S1.

The sequences of 6-letter and 8-letter DNA oligonucleotides for thermodynamic parameters.

a) 6-letter oligonucleotides containing **S** and **B**.

Name	Sequence	Name	Sequence
SB-1	5’-GG**S** AT**B** CC	SB-20	5’-**S**AC TAG T**B**
SB-2	5’-GC**B** AT**S** GC	SB-21	5’-GAC **BS**G TC
SB-3	5’-GG**S** TA**B** CC	SB-22	5’-GAC **SB**G TC
SB-4	5’-GC**B** TA**S** GC	SB-23	5’-GTG **BS**C AC
SB-5	5’-GA**S** CG**B** TC	SB-24	5’-GTG **SB**C AC
SB-6	5’-GT**S** CG**B** AC	SB-25	5’-GCA **BS**T GC
SB-7	5’-GA**S** AT**B** TC	SB-26	5’-GCT **BS**A GC
SB-8	5’-GA**B** AT**S** TC	SB-27	5’-GCA **SB**T GC
SB-9	5’-GA**S** TA**B** TC	SB-28	5’-GCT **SB**A GC
SB-10	5’-GT**S** AT**B** AC	SB-29	5’-G**SS** AT**B B**C
SB-11	5’-G**B**C ATG **S**C	SB-30	5’-G**SS** TA**B B**C
SB-12	5’-G**S**G ATC **B**C	SB-31	5’-C**BB** AT**S S**G
SB-13	5’-C**S**C ATG **B**G	SB-32	5’-GA**S SBB** TC
SB-14	5’-C**S**T CGA **B**G	SB-33	5’-GT**B BSS** AC
SB-15	5’-C**S**A CGT **B**G	SB-34	5’-GGA **SB**T CC
SB-16	5’-C**B**G ATC **S**G	SB-35	5’-G**S**A CGT **B**C
SB-17	5’-**B**GC ATG C**S**	SB-36	5’-GGA **BS**T CC
SB-18	5’-**S**GC ATG C**B**	SB-37	5’-G**B**T CGA **S**C
SB-19	5’-**B**AC TAG T**S**		

b) 6-letter oligonucleotides containing **Z** and **P**.

Name	Sequence	Name	Sequence
ZP-1	5’-GG**Z** AT**P** CC	ZP-22	5’-GAC **ZP**G TC
ZP-2	5’-GC**P** AT**Z** GC	ZP-23	5’-GTG **PZ**C AC
ZP-3	5’-GG**Z** TA**P** CC	ZP-24	5’-GTG **ZP**C AC
ZP-4	5’-GC**P** TA**Z** GC	ZP-25	5’-GCA **PZ**T GC
ZP-5	5’-GA**Z** CG**P** TC	ZP-26	5’-GCT **PZ**A GC
ZP-6	5’-GT**Z** CG**P** AC	ZP-27	5’-GCA **ZP**T GC
ZP-7	5’-GA**Z** AT**P** TC	ZP-28	5’-GCT **ZP**A GC
ZP-8	5’-GA**P** AT**Z** TC	ZP-29	5’-G**ZZ** AT**P P**C
ZP-9	5’-GA**Z** TA**P** TC	ZP-30	5’-G**ZZ** TA**P P**C
ZP-10	5’-GT**Z** AT**P** AC	ZP-31	5’-C**PP** AT**Z Z**G
ZP-11	5’-G**P**C ATG **Z**C	ZP-32	5’-**P**GA CGT C**Z**
ZP-12	5’-G**Z**G ATC **P**C	ZP-33	5’-GGA **ZP**T CC
ZP-13	5’-C**Z**C ATG **P**G	ZP-34	5’-G**P**A CGT **Z**C
ZP-14	5’-C**Z**T CGA **P**G	ZP-35	5’-GGA **PZ**T CC
ZP-15	5’-C**Z**A CGT **P**G	ZP-36	5’-GCC A**P**T TAA 3’-CGG T**Z**A ATT
ZP-16	5’-C**P**G ATC **Z**G	ZP-37	5’-GC**Z** AGT TAA 3’-CG**P** TCA ATT
ZP-17	5’-**P**GC ATG C**Z**	ZP-38	5’-G**Z**C AGT TAA 3’-C**P**G TCA ATT
ZP-18	5’-**Z**GC ATG C**P**	ZP-39	5’-G**ZZ** AGT TAA 3’-C**PP** TCA ATT
ZP-19	5’-**P**AC TAG T**Z**	ZP-40	5’-GA**Z ZPP** TC
ZP-20	5’-**Z**AC TAG T**P**	ZP-41	5’-GT**P PZZ** AC
ZP-21	5’-GAC **PZ**G TC		

c) 8-letter oligonucleotides containing **S**, **B**, **Z** and **P**.

Name	Sequence	Name	Sequence
SBZP-1	5’-C**SZ** AT**P B**G	SBZP-9	5’-G**SZ** AT**P B**C
SBZP-2	5’-C**SP** AT**Z B**G	SBZP-10	5’-G**SP** AT**Z B**C
SBZP-3	5’-C**ZS** AT**B P**G	SBZP-11	5’-G**ZS** AT**B P**C
SBZP-4	5’-C**ZB** AT**S P**G	SBZP-12	5’-G**ZB** AT**S P**C
SBZP-5	5’-C**SZ** TA**P B**G	SBZP-13	5’-GA**Z SBP** TC
SBZP-6	5’-G**SP** TA**Z B**C	SBZP-14	5’-GT**P BSZ** AC
SBZP-7	5’-G**ZS** TA**B P**C	SBZP-15	5’-C**BZ** AT**P S**G
SBZP-8	5’-C**ZB** TA**S P**G	SBZP-16	5’-G**BZ** TA**P S**C

Table S2.

Comparison of **Z-P** to C-G NN Parameters.

	ΔG_{37} **with Z-P**	**Error** ΔG	ΔG_{37} **with C-G**	$\Delta\Delta G_{37}$
Single Internal Z-P pairs:				
AP/TZ	-1.66	0.10	-1.28	**-0.38**
AZ/TP	-1.62	0.10	-1.44	**-0.18**
CP/GZ	-2.32	0.10	-2.17	**-0.15**
CZ/GP	-2.26	0.09	-1.84	**-0.42**
GP/CZ	-2.10	0.11	-1.84	**-0.26**
GZ/CP	-1.78	0.11	-2.24	**+0.42**
TP/AZ	-1.40	0.08	-1.45	**+0.05**
TZ/AP	-1.73	0.09	-1.30	**-0.43**
Tandem Z-P pairs:				
PP/ZZ	-2.28	0.23	-1.84	**-0.44**
ZP/PZ	-2.88	0.22	-2.17	**-0.71**
PZ/ZP	-1.61	0.08	-2.24	**+0.63**
Average of Single **Z-P** substitutions				**-0.17**

*All results are in kcal/mol

Parameters with C-G pairs are from *(15)*.

Table S3.

Experimental and predicted free energy changes and T_m's for the 41 **Z**:**P** containing duplexes. The experimental values are the error-weighted average of thermodynamic values from curve fit and T_m^{-1} methods of analysis (see **Table S4**). These are the data plotted in **Figures S1** and **S2**.

Self-Complementary Duplex	**Pred. $\Delta G°_{37}$ (kcal/mol)**	**Expt. $\Delta G°_{37}$ (kcal/mol)**	**Pred. Tm ($1e^{-4}$M)**	**Exp. Tm ($1e^{-4}$M)**
GG**Z**AT**P**CC	-8.50	-7.93	61.0	56.8
GC**P**AT**Z**GC	-11.08	-11.46	65.4	66.4
GG**Z**TA**P**CC	-8.72	-8.38	62.9	62.0
GC**P**TA**Z**GC	-10.54	-10.57	65.2	64.9
GA**Z**CG**P**TC	-9.81	-9.92	63.5	64.3
GT**Z**CG**P**AC	-10.32	-10.65	63.5	64.7
GA**Z**AT**P**TC	-7.12	-7.13	46.6	47.0
GA**P**AT**Z**TC	-7.87	-7.96	50.1	49.3
GA**Z**TA**P**TC	-7.34	-7.23	48.1	47.9
GT**Z**AT**P**AC	-7.63	-7.78	48.6	49.6
G**P**CATG**Z**C	-9.12	-8.59	65.0	61.8
G**Z**GATC**P**C	-10.20	-10.39	70.0	66.2
C**Z**CATG**P**G	-10.09	-9.87	64.9	64.6
C**Z**TCGA**P**G	-10.20	-10.40	64.7	63.3
C**Z**ACGT**P**G	-9.97	-10.23	63.0	63.4
C**P**GATC**Z**G	-10.24	-10.52	63.0	64.3
PGCATGC**Z**	-9.38	-8.61	61.8	58.5
ZGCATGC**P**	-9.51	-8.02	63.8	61.9
PACTAGT**Z**	-6.10	-6.40	39.7	41.2
ZACTAGT**P**	-5.42	-6.33	35.0	41.0
GAC**PZ**GTC	-9.34	-10.38	58.0	64.6
GAC**ZP**GTC	-10.48	-10.79	63.6	66.2
GTG**PZ**CAC	-9.19	-10.28	59.4	63.2
GTG**ZP**CAC	-9.80	-10.81	65.2	67.2
GCA**PZ**TGC	-9.91	-10.48	63.4	64.4
GCT**PZ**AGC	-9.06	-8.83	58.1	59.1
GCA**ZP**TGC	-11.10	-10.99	67.6	65.3
GCT**ZP**AGC	-11.00	-9.73	64.6	61.6
G**ZZ**AT**PP**C	-9.38	-9.86	64.0	69.8
G**ZZ**TA**PP**C	-9.60	-9.90	65.6	68.5
C**PP**AT**ZZ**G	-11.16	-12.07	64.3	67.6
PGACGTC**Z**	-8.77	-9.05	57.4	58.3
GGA**ZP**TCC	-10.00	-9.66	62.9	61.3
G**P**ACGT**Z**C	-10.32	-9.51	63.4	60.5
GGA**PZ**TCC	-8.81	-8.63	57.9	57.4
GA**ZZPP**TC	-10.88	-11.87	64.9	68.7
GT**PPZZ**AC	-9.46	-8.13	59.1	50.9
non-self-complementary duplexes				
GCCA**P**TTAA CGGT**Z**AATT	-9.37	-9.01	52.6	50.1
GC**Z**AGTTAA	-9.18	-9.19	51.2	50.2

CG**P**TCAATT G**Z**CAGTTAA	-8.60	-9.35	49.4	49.8
C**P**GTCAATT G**ZZ**AGTTAA	-8.74	-8.45	49.6	47.8
C**PP**TCAATT				

Table S4.

The raw data obtained for melts of duplexes containing **Z**:**P** nearest-neighbors in 1 M NaCl, 10 mM Na_2HPO_4, and 0.5 mM Na_2EDTA, pH 7.00 buffer. All data were fit using the thermodynamic analysis software Meltwin v.3.5.

	T_m^{-1} Method			Curve Fit Method			
Duplex	**ΔH°**	**ΔS°**	**ΔG°$_{37}$**	**T_m**	**ΔH°**	**ΔS°**	**ΔG°$_{37}$**
	(kcal/mol)	**(cal/mol*K)**	**(kcal/mol)**	**(1e^{-4}M)**	**(kcal/mol)**	**(cal/mol*K)**	**(kcal/mol)**
GG**Z**AT**P**CC	-34.1 ± 8.6	-84.7 ± 26.8	-7.79 ± 0.67	56.8	-38.8 ± 5.0	-99.4 ± 15.5	-7.94 ± 0.20
GC**P**AT**Z**GC	-67.7 ± 5.8	-181.2 ± 17.6	-11.52 ± 0.40	66.4	-66.5 ± 3.3	-177.5 ± 10.0	-11.44 ± 0.22
GG**Z**TA**P**CC	-36.5 ± 12.1	-90.7 ± 36.8	-8.38 ± 1.38	62.0	-36.2 ± 4.7	-89.6 ± 14.0	-8.38 ± 0.33
GC**P**TA**Z**GC	-58.6 ± 1.7	-154.9 ± 5.0	-10.53 ± 0.11	64.9	-65.1 ± 5.4	-174.5 ± 16.3	-10.96 ± 0.36
GA**Z**CG**P**TC	-50.2 ± 5.0	-130.4 ± 15.2	-9.81 ± 0.28	64.3	-54.6 ± 4.8	-143.7 ± 14.5	-10.04 ± 0.30
GT**Z**CG**P**AC	-60.3 ± 1.0	-160.2 ± 3.2	-10.64 ± 0.06	64.7	-64.2 ± 3.5	-172.0 ± 10.6	-10.87 ± 0.24
GA**Z**AT**P**TC	-44.3 ± 12.4	-120.0 ± 39.3	-7.11 ± 0.88	47.0	-47.0 ± 4.6	-128.4 ± 14.2	-7.13 ± 0.18
GA**P**AT**Z**TC	-59.4 ± 4.0	-165.9 ± 12.7	-7.96 ± 0.08	49.3	-60.3 ± 6.3	-168.8 ± 19.8	-7.95 ± 0.15
GA**Z**TA**P**TC	-45.0 ± 5.3	-121.8 ± 17.0	-7.24 ± 0.17	47.9	-47.0 ± 7.0	-128.1 ± 22.6	-7.23 ± 0.06
GT**Z**AT**P**AC	-49.8 ± 3.9	-135.5 ± 12.5	-7.75 ± 0.10	49.6	-55.6 ± 2.6	-154.0 ± 8.3	-7.81 ± 0.10
G**P**CATG**Z**C	-38.4 ± 6.7	-96.3 ± 20.8	-8.50 ± 0.47	61.8	-40.9 ± 9.1	-103.9 ± 27.9	-8.68 ± 0.47
G**Z**GATC**P**C	-52.9 ± 5.7	-137.5 ± 17.3	-10.27 ± 0.34	66.2	-59.9 ± 9.4	-158.6 ± 28.5	-10.71 ± 0.55
C**Z**CATG**P**G	-49.8 ± 6.8	-129.1 ± 20.7	-9.77 ± 0.46	64.6	-52.3 ± 5.5	-136.5 ± 16.5	-9.93 ± 0.36
C**Z**TCGA**P**G	-57.6 ± 4.5	-152.7 ± 13.8	-10.25 ± 0.27	63.3	-64.5 ± 5.3	-173.5 ± 16.0	-10.65 ± 0.34
C**Z**ACGT**P**G	-55.1 ± 7.7	-145.0 ± 23.4	-10.08 ± 0.49	63.4	-60.4 ± 7.1	-161.2 ± 21.1	-10.40 ± 0.51
C**P**GATC**Z**G	-58.6 ± 1.9	-155.1 ± 5.7	-10.45 ± 0.11	64.3	-67.1 ± 4.4	-181.1 ± 13.3	-10.97 ± 0.29
PGCATGC**Z**	-40.2 ± 1.8	-101.9 ± 5.7	-8.60 ± 0.06	58.5	-45.3 ± 5.8	-117.6 ± 17.8	-8.81 ± 0.24
ZGCATGC**P**	-30.0 ± 7.1	-71.4 ± 22.0	-7.87 ± 0.59	61.9	-34.4 ± 9.9	-84.8 ± 31.2	-8.06 ± 0.31
PACTAGT**Z**	-50.8 ± 9.9	-143.0 ± 32.0	-6.42 ± 0.26	41.2	-53.6 ± 0.6	-152.0 ± 2.1	-6.40 ± 0.05
ZACTAGT**P**	-52.7 ± 8.1	-149.4 ± 26.5	-6.32 ± 0.25	41.0	-50.7 ± 3.5	-142.9 ± 11.6	-6.33 ± 0.10
GAC**PZ**GTC	-56.7 ± 5.9	-149.6 ± 17.9	-10.32 ± 0.38	64.6	-58.7 ± 6.4	-155.4 ± 19.2	-10.46 ± 0.41
GAC**ZP**GTC	-53.0 ± 9.9	-137.5 ± 29.6	-10.36 ± 0.82	66.2	-60.9 ± 4.6	-161.3 ± 13.8	-10.86 ± 0.33
GTG**PZ**CAC	-60.2 ± 3.0	-160.7 ± 9.1	-10.33 ± 0.17	63.2	-58.1 ± 2.9	-154.5 ± 8.7	-10.22 ± 0.18
GTG**ZP**CAC	-56.1 ± 4.9	-146.3 ± 14.9	-10.69 ± 0.34	67.2	-58.9 ± 4.0	-154.9 ± 11.9	-10.90 ± 0.30
GCA**PZ**TGC	-55.8 ± 6.4	-146.6 ± 20.0	-10.31 ± 0.40	64.4	-64.5 ± 8.2	-173.0 ± 24.4	-10.86 ± 0.59
GCT**PZ**AGC	-42.9 ± 6.7	-110.3 ± 20.9	-8.68 ± 0.34	59.1	-49.1 ± 4.1	-129.7 ± 12.5	-8.89 ± 0.21
GCA**ZP**TGC	-60.9 ± 4.7	-161.5 ± 14.3	-10.83 ± 0.30	65.3	-68.2 ± 6.4	-183.5 ± 19.4	-11.31 ± 0.42
GCT**ZP**AGC	-52.7 ± 4.6	-139.0 ± 14.2	-9.62 ± 0.24	61.6	-58.6 ± 5.5	-157.0 ± 16.8	-9.92 ± 0.32
G**ZZ**AT**PP**C	-44.5 ± 6.6	-111.7 ± 19.8	-9.90 ± 0.51	69.8	-42.3 ± 8.7	-104.6 ± 26.3	-9.82 ± 0.56
G**ZZ**TA**PP**C	-43.5 ± 7.4	-108.9 ± 22.4	-9.72 ± 0.57	68.5	-49.8 ± 9.9	-127.8 ± 29.6	-10.17 ± 0.69
C**PP**AT**ZZ**G	-70.9 ± 4.5	-189.6 ± 13.7	-12.04 ± 0.31	67.6	-71.5 ± 5.3	-191.5 ± 15.9	-12.11 ± 0.38
PGACGTC**Z**	-51.0 ± 4.1	-135.4 ± 12.6	-8.99 ± 0.18	58.3	-58.4 ± 7.9	-158.3 ± 24.2	-9.33 ± 0.38
GGA**ZP**TCC	-54.7 ± 6.1	-145.4 ± 18.6	-9.64 ± 0.34	61.3	-55.2 ± 6.9	-146.8 ± 21.0	-9.69 ± 0.36
G**P**ACGT**Z**C	-53.3 ± 1.9	-141.4 ± 5.8	-9.48 ± 0.08	60.5	-61.3 ± 4.8	-165.7 ± 14.5	-9.85 ± 0.29
GGA**PZ**TCC	-42.4 ± 12.9	-109.3 ± 40.4	-8.49 ± 0.78	57.4	-48.2 ± 3.5	-127.5 ± 11.2	-8.63 ± 0.07
GCCA**P**TTAA	-54.3 ± 6.2	-146.3 ± 19.5	-8.91 ± 0.18	50.1	-62.7 ± 3.0	-173.1 ± 9.5	-9.03 ± 0.09
GC**Z**AGTTAA	-60.3 ± 9.4	-165.0 ± 29.9	-9.10 ± 0.32	50.2	-69.9 ± 9.0	-195.6 ± 28.6	-9.21 ± 0.17
G**Z**CAGTTAA	-70.9 ± 3.3	-198.5 ± 10.5	-9.34 ± 0.06	49.8	-72.2 ± 2.1	-202.5 ± 6.8	-9.36 ± 0.05
G**ZZ**AGTTAA	-55.5 ± 1.8	-151.9 ± 5.7	-8.40 ± 0.04	47.8	-61.8 ± 3.0	-171.8 ± 9.5	-8.52 ± 0.05
GA**ZZPP**TC	-64.4 ± 6.0	-169.9 ± 18.0	-11.72 ± 0.48	68.7	-67.5 ± 3.6	-179.0 ± 10.7	-11.94 ± 0.33
GT**PPZZ**AC	-53.9 ± 6.4	-147.8 ± 20.2	-8.06 ± 0.24	50.9	-58.8 ± 4.7	-163.2 ± 14.8	-8.16 ± 0.17

Table S5.

Comparison of **S**:**B** to T:A NN Parameters.

	ΔG_{37} with S-B	Error ΔG	ΔG_{37} with T-A	$\Delta\Delta G_{37}$
Single Internal S-B pairs:				
AB/TS	-1.24	0.09	-1.00	**-0.24**
AS/TB	-1.30	0.09	-0.88	**-0.42**
CB/GS	-1.92	0.09	-1.45	**-0.47**
CS/GB	-1.44	0.09	-1.28	**-0.16**
GB/CS	-1.70	0.10	-1.30	**-0.40**
GS/CB	-1.78	0.10	-1.44	**-0.34**
TB/AS	-1.37	0.08	-0.58	**-0.79**
TS/AB	-1.36	0.08	-1.00	**-0.36**
Tandem S-B pairs:				
BB/SS	-1.79	0.08	-1.00	**-0.79**
SB/BS	-1.68	0.20	-0.58	**-1.10**
BS/SB	-1.67	0.20	-0.88	**-0.79**
Average of Single **S**:**B** substitutions				**-0.53**

*All results are in kcal/mol

Parameters with T:A pairs are from *(15)*.

Table S6.

The Experimental and Predicted Free Energy changes and T_m's for the 37 **S**:**B** containing duplexes. Experimental values are the error-weighted average of thermodynamic values from curve fit and T_m^{-1} methods of analysis (see **Table S7**). These are the data plotted in **Figures S3** and **S4**.

Self-Complementary Duplex	Pred. $\Delta G°_{37}$ (kcal/mol)	Expt. $\Delta G°_{37}$ (kcal/mol)	Pred. T_m ($1e^{-4}$M)	Exp. T_m ($1e^{-4}$M)
GG**S**AT**B**CC	-8.47	-8.02	55.2	51.9
GC**B**AT**S**GC	-9.53	-9.33	59.6	60.5
GG**S**TA**B**CC	-7.92	-8.06	51.5	52.0
GC**B**TA**S**GC	-9.11	-8.63	59.0	56.5
GA**S**CG**B**TC	-8.38	-8.32	54.2	55.8
GT**S**CG**B**AC	-8.78	-8.73	54.9	53.5
GA**S**AT**B**TC	-6.43	-6.60	41.8	42.7
GA**B**AT**S**TC	-6.30	-6.29	40.6	40.7
GA**S**TA**B**TC	-5.88	-6.36	38.3	40.8
GT**S**AT**B**AC	-6.83	-6.74	43.7	43.5
G**B**CATG**S**C	-8.35	-8.22	54.4	53.1
G**S**GATC**B**C	-8.49	-9.78	55.3	61.1
C**S**CATG**B**G	-7.67	-7.71	49.3	48.5
C**S**TCGA**B**G	-7.75	-7.32	48.9	45.4
C**S**ACGT**B**G	-8.28	-8.08	52.0	50.2
C**B**GATC**S**G	-7.81	-8.50	50.2	54.8
BGCATGC**S**	-7.75	-7.58	51.6	52.0
SGCATGC**B**	-8.71	-7.80	58.6	53.3
BACTAGT**S**	-5.35	-5.44	34.7	35.1
SACTAGT**B**	-5.37	-5.92	34.8	38.5
GAC**BS**GTC	-8.60	-8.36	55.0	53.3
GAC**SB**GTC	-7.64	-7.89	49.3	51.8
GTG**BS**CAC	-8.46	-8.72	54.1	56.4
GTG**SB**CAC	-8.62	-9.18	56.3	60.7
GCA**BS**TGC	-9.15	-9.07	57.1	54.7
GCT**BS**AGC	-9.07	-8.64	57.1	55.2
GCA**SB**TGC	-9.27	-8.45	60.4	55.9
GCT**SB**AGC	-9.05	-8.89	57.3	57.4
G**SS**AT**BB**C	-8.37	-8.09	57.0	53.7
G**SS**TA**BB**C	-7.82	-7.37	52.8	50.4
C**BB**AT**SS**G	-8.63	-8.97	57.7	57.0
GA**SSBB**TC	-8.07	-8.42	56.2	58.6
GT**BBSS**AC	-8.49	-8.78	56.6	59.9
GGA**SB**TCC	-8.17	-8.04	54.3	50.2
G**S**ACGT**B**C	-8.96	-8.90	56.8	57.7
GGA**BS**TCC	-8.05	-8.28	51.6	51.6
G**B**TCGA**S**C	-8.38	-8.33	54.2	54.1

Table S7.

The raw data obtained for duplex melts containing **S**:**B** nearest-neighbors in 1 M NaCl, 10 mM Na_2HPO_4, and 0.5 mM Na_2EDTA, pH 7.00 buffer. All data were fit using the thermodynamic analysis software Meltwin v.3.5.

	T_m^{-1} Method			Curve Fit Method			
Duplex	**ΔH°** **(kcal/mol)**	**ΔS°** **(cal/mol*K)**	**ΔG°$_{37}$** **(kcal/mol)**	**T_m** **($1e^{-4}$M)**	**ΔH°** **(kcal/mol)**	**ΔS°** **(cal/mol*K)**	**ΔG°$_{37}$** **(kcal/mol)**
GG**S**AT**B**CC	-50.6 ± 1.1	-137.4 ± 3.3	-8.02 ± 0.02	51.9	-55.9 ± 7.0	-154.1 ± 22.3	-8.06 ± 0.13
GC**B**AT**S**GC	-50.0 ± 7.2	-131.5 ± 22.1	-9.25 ± 0.43	60.5	-52.4 ± 3.0	-138.7 ± 8.9	-9.35 ± 0.21
GG**S**TA**B**CC	-51.1 ± 7.8	-138.8 ± 24.7	-8.05 ± 0.34	52.0	-52.9 ± 8.1	-144.4 ± 25.6	-8.06 ± 0.19
GC**B**TA**S**GC	-45.4 ± 5.9	-118.9 ± 18.3	-8.51 ± 0.31	56.5	-50.7 ± 3.6	-135.5 ± 10.9	-8.69 ± 0.23
GA**S**CG**B**TC	-44.7 ± 5.8	-117.3 ± 18.3	-8.27 ± 0.23	55.8	-48.2 ± 4.0	-128.4 ± 12.7	-8.33 ± 0.12
GT**S**CG**B**AC	-59.2 ± 8.2	-162.9 ± 25.8	-8.71 ± 0.32	53.5	-60.4 ± 6.8	-166.6 ± 21.3	-8.74 ± 0.20
GA**S**AT**B**TC	-45.7 ± 3.6	-125.9 ± 11.5	-6.62 ± 0.08	42.7	-50.6 ± 1.0	-141.9 ± 3.1	-6.58 ± 0.07
GA**B**AT**S**TC	-44.5 ± 3.8	-122.9 ± 12.6	-6.36 ± 0.11	40.7	-48.2 ± 1.9	-135.2 ± 6.0	-6.29 ± 0.03
GA**S**TA**B**TC	-47.6 ± 4.6	-132.9 ± 15.2	-6.42 ± 0.15	40.8	-52.3 ± 3.9	-148.1 ± 12.4	-6.35 ± 0.05
GT**S**AT**B**AC	-50.7 ± 3.7	-141.7 ± 11.9	-6.75 ± 0.05	43.5	-56.8 ± 3.7	-161.4 ± 12.0	-6.73 ± 0.05
G**B**CATG**S**C	-52.1 ± 1.9	-141.4 ± 6.1	-8.22 ± 0.04	53.1	-52.7 ± 3.4	-143.2 ± 10.7	-8.24 ± 0.07
G**S**GATC**B**C	-53.8 ± 6.0	-142.4 ± 18.4	-9.65 ± 0.35	61.1	-58.1 ± 5.0	-155.6 ± 15.2	-9.89 ± 0.32
C**S**CATG**B**G	-60.7 ± 9.5	-170.6 ± 30.4	-7.75 ± 0.27	48.5	-56.9 ± 2.8	-158.6 ± 9.0	-7.70 ± 0.09
C**S**TCGA**B**G	-60.8 ± 9.9	-172.4 ± 31.8	-7.34 ± 0.33	45.4	-62.2 ± 3.2	-177.0 ± 10.3	-7.32 ± 0.14
C**S**ACGT**B**G	-58.6 ± 6.8	-163.0 ± 21.5	-8.07 ± 0.19	50.2	-57.0 ± 8.6	-157.8 ± 27.4	-8.09 ± 0.13
C**B**GATC**S**G	-49.8 ± 6.0	-133.3 ± 18.7	-8.46 ± 0.24	54.8	-52.2 ± 2.5	-141.0 ± 8.0	-8.51 ± 0.09
BGCATGC**S**	-40.2 ± 3.2	-105.2 ± 10.2	-7.57 ± 0.10	52.0	-43.4 ± 3.4	-115.3 ± 10.8	-7.60 ± 0.11
SGCATGC**B**	-39.8 ± 4.0	-103.3 ± 12.6	-7.76 ± 0.13	53.3	-43.2 ± 1.9	-114.1 ± 6.0	-7.82 ± 0.10
BACTAGT**S**	-46.2 ± 4.1	-131.3 ± 13.7	-5.52 ± 0.16	35.1	-49.6 ± 1.5	-142.5 ± 4.7	-5.42 ± 0.07
SACTAGT**B**	-47.4 ± 9.3	-133.7 ± 30.5	-5.91 ± 0.40	38.5	-47.9 ± 4.6	-135.4 ± 15.1	-5.92 ± 0.10
GAC**BS**GTC	-47.7 ± 11.6	-117.0 ± 36.6	-8.33 ± 0.58	53.3	-55.8 ± 1.1	-121.4 ± 3.5	-8.36 ± 0.06
GAC**SB**GTC	-49.2 ± 4.0	-133.1 ± 12.8	-7.89 ± 0.12	51.8	-49.2 ± 1.6	-133.1 ± 4.8	-7.89 ± 0.09
GTG**BS**CAC	-47.4 ± 8.3	-125.3 ± 25.8	-8.57 ± 0.30	56.4	-52.3 ± 3.4	-140.6 ± 10.9	-8.73 ± 0.08
GTG**SB**CAC	-47.6 ± 2.1	-124.0 ± 6.6	-9.14 ± 0.10	60.7	-52.7 ± 3.7	-139.7 ± 11.3	-9.38 ± 0.21
GCA**BS**TGC	-59.1 ± 9.8	-161.6 ± 30.6	-9.00 ± 0.42	54.7	-62.1 ± 5.2	-171.2 ± 16.2	-9.09 ± 0.21
GCT**BS**AGC	-46.5 ± 14.3	-122.5 ± 44.6	-8.48 ± 0.80	55.2	-53.2 ± 4.1	-143.8 ± 12.9	-8.64 ± 0.10
GCA**SB**TGC	-46.2 ± 3.7	-122.1 ± 11.4	-8.37 ± 0.14	55.9	-49.3 ± 2.1	-131.7 ± 6.7	-8.48 ± 0.09
GCT**SB**AGC	-52.0 ± 4.8	-139.1 ± 15.0	-8.89 ± 0.22	57.4	-52.0 ± 3.7	-139.0 ± 11.2	-8.90 ± 0.24
G**SS**AT**BB**C	-41.7 ± 9.3	-108.8 ± 29.1	-7.94 ± 0.55	53.7	-48.6 ± 4.9	-130.4 ± 15.1	-8.12 ± 0.25
G**SS**TA**BB**C	-38.7 ± 13.0	-101.0 ± 41.2	-7.35 ± 1.08	50.4	-40.4 ± 5.1	-106.6 ± 15.9	-7.37 ± 0.13
C**BB**AT**SS**G	-52.5 ± 4.5	-140.7 ± 13.9	-8.90 ± 0.20	57.0	-56.1 ± 5.5	-151.6 ± 17.0	-9.07 ± 0.24
GA**SSBB**TC	-37.5 ± 9.1	-94.3 ± 28.3	-8.27 ± 0.52	58.6	-42.4 ± 1.8	-109.5 ± 5.5	-8.43 ± 0.14
GT**BBSS**AC	-42.3 ± 2.4	-108.2 ± 7.5	-8.73 ± 0.11	59.9	-48.3 ± 3.7	-126.8 ± 11.1	-9.01 ± 0.23
GGA**SB**TCC	-55.6 ± 4.9	-153.6 ± 15.6	-8.01 ± 0.15	50.2	-59.0 ± 4.5	-164.1 ± 14.1	-8.07 ± 0.14
G**S**ACGT**B**C	-46.8 ± 11.3	-122.8 ± 34.9	-8.75 ± 0.72	57.7	-51.7 ± 2.3	-138.0 ± 7.1	-8.91 ± 0.13
GGA**BS**TCC	-55.3 ± 4.5	-151.9 ± 14.2	-8.22 ± 0.12	51.6	-62.1 ± 4.8	-173.3 ± 15.3	-8.34 ± 0.13
G**B**TCGA**S**C	-45.1 ± 6.7	-119.1 ± 21.0	-8.19 ± 0.34	54.1	-51.2 ± 2.2	-138.2 ± 6.9	-8.35 ± 0.13

Table S8.

Comparison of tandem **S**:**B**, **Z**:**P** NN to C:G NN Parameters.

	ΔG_{37} **with S:B Z:P**	**Error** ΔG	ΔG_{37} **with C:G**	$\Delta\Delta G_{37}$
Tandem S:B-Z:P pairs:				
SZ/BP	-2.29	0.11	-1.84	**-0.45**
ZS/PB	-1.86	0.09	-1.84	**-0.02**
SP/BZ	-2.28	0.08	-2.17	**-0.11**
PS/ZB	-1.90	0.12	-2.17	**+0.27**
Average of **S**:**B**, **Z**:**P** to C:G substitutions				**-0.08**

*All results are in kcal/mol

Parameters with C:G pairs are from *(15)*.

Table S9.

Experimental and Predicted Free Energy changes and T_m's for 16 tandem **S**:**B**, **Z**:**P** containing duplexes used in plots in **Figures S5** and **S6**. The experimental values are the error-weighted average of thermodynamic values from curve fit and T_m^{-1} methods of analysis (see **Table S10**). * This 16mer duplex is non-two-state due to competing hairpin formation, slow kinetics, and limited high-temperature baseline. It is presented for validation purposes, but was not used to derive NN parameters.

Self-Complementary Duplex	**Pred. $\Delta G°_{37}$ (kcal/mol)**	**Expt. $\Delta G°_{37}$ (kcal/mol)**	**Pred. T_m ($1e^{-4}$M)**	**Exp. T_m ($1e^{-4}$M)**
C**SZ**AT**PB**G	-8.75	-8.30	57.6	56.9
C**SP**AT**ZB**G	-9.40	-9.14	57.5	57.2
C**ZS**AT**BP**G	-9.47	-9.24	61.2	62.4
C**ZB**AT**SP**G	-10.30	-10.32	63.1	62.7
G**SP**TA**ZB**C	-9.56	-9.27	61.8	61.6
G**ZS**TA**BP**C	-7.93	-8.58	50.9	57.6
C**ZB**TA**SP**G	-9.88	-9.85	63.0	62.3
G**SZ**AT**PB**C	-9.43	-9.50	63.4	60.2
G**SP**AT**ZB**C	-10.08	-10.51	62.2	63.7
G**ZS**AT**BP**C	-8.48	-8.81	60.0	61.9
G**ZB**AT**SP**C	-9.32	-9.57	62.7	62.4
GA**ZSBP**TC	-8.85	-8.13	59.4	49.1
GT**PBSZ**AC	-9.54	-10.12	62.6	65.5
C**BZ**AT**PS**G	-8.93	-9.03	63.2	62.8
G**BZ**TA**PS**C	-8.71	-8.61	61.5	61.7
CTTAT**PPSBZZ**ATAAG*	-17.93	-16.46	77.0	81.3

Table S10.

The raw data obtained for duplex melts containing **S**-**B**, **Z**-**P** nearest-neighbors in 1 M NaCl, 10 mM Na_2HPO_4, and 0.5 mM Na_2EDTA, pH 7.00 buffer. All data were fit using the thermodynamic analysis software Meltwin v.3.5.

* The 8mer did not give a discernable melting transition and thus is listed as NTS = non-two-state.

** This 16mer duplex is non-two-state due to competing hairpin formation, slow kinetics, and limited high-temperature baseline. It is presented for validation purposes, but was not used to derive NN parameters.

	T_m^{-1} Method			**Curve Fit Method**			
Duplex	**ΔH° (kcal/mol)**	**ΔS° (cal/mol*K)**	**$\Delta G°_{37}$ (kcal/mol)**	**T_m ($1e^{-4}$M)**	**ΔH° (kcal/mol)**	**ΔS° (cal/mol*K)**	**$\Delta G°_{37}$ (kcal/mol)**
C**SZ**AT**PB**G	-41.5 ± 10.1	-107.4 ± 31.3	-8.20 ± 0.69	56.7	-44.2 ± 6.1	-115.7 ± 18.7	-8.32 ± 0.34
C**SP**AT**ZB**G	-50.8 ± 14.1	-135.1 ± 43.0	-8.87 ± 1.11	57.1	-57.8 ± 6.2	-156.9 ± 18.9	-9.18 ± 0.42
C**ZS**AT**BP**G	-44.6 ± 4.6	-114.3 ± 14.0	-9.12 ± 0.23	62.2	-50.2 ± 5.1	-131.7 ± 15.4	-9.41 ± 0.27
C**ZB**AT**SP**G	-60.3 ± 4.7	-161.2 ± 14.4	-10.29 ± 0.27	62.7	-61.0 ± 4.6	-163.3 ± 13.9	-10.35 ± 0.30
C**SZ**TA**PB**G*		NTS		NTS		NTS	
G**SP**TA**ZB**C	-43.9 ± 16.1	-112.3 ± 49.3	-9.07 ± 1.23	61.6	-48.9 ± 1.8	-127.8 ± 5.5	-9.27 ± 0.10
G**ZS**TA**BP**C	-46.2 ± 5.8	-121.4 ± 18.1	-8.58 ± 0.30	57.8	-46.6 ± 2.8	-122.5 ± 9.1	-8.58 ± 0.10
C**ZB**TA**SP**G	-53.8 ± 5.9	-142.2 ± 18.1	-9.75 ± 0.36	61.9	-58.2 ± 8.1	-155.4 ± 24.7	-10.02 ± 0.47
G**SZ**AT**PB**C	-51.9 ± 14.7	-137.4 ± 44.6	-9.33 ± 1.24	60.0	-56.0 ± 9.3	-149.8 ± 28.2	-9.53 ± 0.55
G**SP**AT**ZB**C	-59.9 ± 3.4	-159.6 ± 10.2	-10.45 ± 0.20	63.6	-63.1 ± 4.8	-169.1 ± 14.4	-10.65 ± 0.29
G**ZS**AT**BP**C	-40.2 ± 5.3	-101.6 ± 16.4	-8.72 ± 0.28	61.4	-42.9 ± 3.2	-110.0 ± 9.9	-8.84 ± 0.17
G**ZB**AT**SP**C	-47.1 ± 4.5	-121.5 ± 14.0	-9.40 ± 0.24	61.9	-53.7 ± 3.3	-141.9 ± 9.9	-9.71 ± 0.22
GA**ZSBP**TC	-64.8 ± 5.4	-182.8 ± 17.2	-8.12 ± 0.12	61.9	-65.1 ± 1.7	-183.7 ± 5.3	-8.13 ± 0.06
GT**PBSZ**AC	-52.3 ± 2.6	-136.1 ± 7.9	-10.10 ± 0.14	61.9	-54.4 ± 4.9	-142.4 ± 14.9	-10.23 ± 0.33
C**BZ**AT**PS**G	-41.6 ± 14.2	-105.8 ± 42.9	-8.59 ± 1.16	61.9	-44.2 ± 7.5	-113.2 ± 22.5	-9.11 ± 0.49
G**BZ**TA**PS**C	-39.6 ± 1.3	-99.9 ± 3.8	-8.60 ± 0.06	61.9	-41.6 ± 3.2	-106.0 ± 9.8	-8.71 ± 0.16
CTTAT**PPSBZZ**ATAAG**	-83.9 ± 9.3	-218.1 ± 27.0	-16.22 ± 0.96	81.3	-83.6 ± 1.4	-225.2 ± 44.1	-16.46 ± 0.13

Table S11.

T7 RNA Polymerases Examined in this Research.

Variant	substrates incorporated	name
wild-type		
Y639F	2'-fluoro-2'-deoxyribose	F
Y639F P266L	2'-fluoro-2'-deoxyribose ribose lower activity, less early termination	FL
Y639F H784A	modified 2'-ribose derivatives	FA
Y639F H784A P266L	modified 2' ribose derivatives, lower activity, less early termination	FAL
G542V H772R H784S	2'-fluoro ribose	VRS

Table S12.

RNA generated by T7 RNA polymerase.

T2P: Incorporating a single hachimoji **P** 5’-GGG AGU GUU GUA UUU GG**P** CAA UUU Transcription with α^{32}P-CTP then digestion with RNase T2 yields labeled **P**-3'-phosphate.
T2Z: Incorporating a single hachimoji **Z** 5’-GGG AGU GUU GUA UUU GG**Z** CAA UUU Transcription with α^{32}P-CTP then digestion with RNase T2 yields labeled **Z**-3'-phosphate.
T2B: Incorporating a single hachimoji **B** 5’-GGG AGU GUU GUA UUU GG**B** CAA UUU Transcription with α^{32}P-CTP then digestion with RNase T2 yields labeled **B**-3'-phosphate.
T2S: Incorporating a single hachimoji **S** 5’-GGG AGU GUU GUA UUU GG**S** CAA UUU Transcription with α^{32}P-CTP then digestion with RNase T2 yields labeled **S**-3'-phosphate.
Spinach aptamer with four non-standard hachimoji components 5’-GGA CGC G**B**C **Z**GA AAU GGU GAA GGA CGG GUC CAG UGC GAA ACA CGC ACU GUU GAG UAG AGU GUG AGC UCC GUA ACU **P**G**S** CGC GUC-3’

Table S13.

Crystallographic Data for Hachimoji DNA.

	PB	**PC**	**PP**
Cell dimensions			
a (Å)	55.21	55.19	55.15
b (Å)	145.90	146.24	145.85
c (Å)	46.93	46.86	46.85
Space group	$P2_12_12$	$P2_12_12_1$	$P2_12_12$
Resolution (Å)	39.47 – 1.7	39.45 – 1.6	39,42 – 1.7
Total observations	187020	339883	189052
Unique reflections	42391	50952	42311
Completeness (%)	99.4 (98.5)	99.9 (99.7)	99.5 (99.3)
R_{merge}	0.054 (0.89)	0.055(1.34)	0.095 (0.84)
R_{pim}	0.028 (0.47)	0.023 (0.56)	0.050 (0.44)
I/σ	13.9 (1.7)	15.8 (1.3)	8.0 (1.6)
CC(1/2)	0.999 (0.62)	0.998(0.556)	0.992 (0.666)
Resolution	44.69 – 1.7	39.45 – 1.6	39.42 – 1.7
Refinement Statistics			
R_{work}/R_{free} (%)	22.2/25.6	21.5/23.4	21.3/24.7
No. atoms			
Protein	1989	2024	2024
DNA	329	329	332
Water	161	204	191
B-factors			
Protein	31.1	30.5	31.3
DNA	65.8	57.4	61.8
Water	31.1	32.2	32.3
R.m.s, deviations			
Bond lengths (Å)	0.007	0.007	0.007
Bond angles (°)	0.92	0.96	0.97

PDB IDs are as follows: **PB** (6MIG), **P**C (6MIH), and **PP** (6MIK).