Supporting Information for:

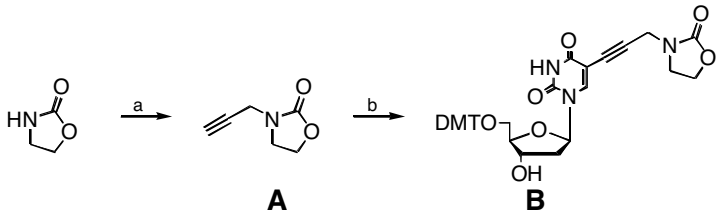
Adaptability at a protein-DNA interface: re-engineering the engrailed homeodomain to recognize an unnatural nucleotide

Matthew D. Simon and Kevan M. Shokat*

Department of Chemistry, University of California, Berkeley, California 94720 and Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94143

General. Unless otherwise noted, all reactions were performed under a N_2 or Ar atmosphere. Unless otherwise noted, reagents were purchased from commercial suppliers and used without further purification. Flash chromatography was performed using Merck 60 Å 230-400 mesh silica gel. Infrared spectra were taken on a Perkin Elmer series fourier transform infrared spectrometer. IR spectra were recorded neat (for oils) and as films from CH_2Cl_2 or $CHCl_3$ (for crystalline compounds) on a NaCl disk. UV analysis was performed on a UVIKON XL spectrometer unless otherwise noted. Mass spectrometry was performed by the University of California, San Francisco or Berkeley, mass spectrometry facilities. Matrix assisted laser desorption ionization-mass spectrometry was performed with David King at the University of California, Berkeley HHMI mass spectrometry facility using a Bruker Reflex III spectrometer. Elemental analyses were performed by the University of California, Berkeley Microanalytical facility on a Perkin Elmer 2400 Series II CHNO/S Analyzer. All NMR spectra were obtained using Bruker AMX-300, AMX-400 or DRX-500 MHz spectrometers. All ^1H, ^{13}C NMR experiments are reported in ∂ relative to TMS. When available, oligonucleotides were purchased from Operon with HPLC purification.

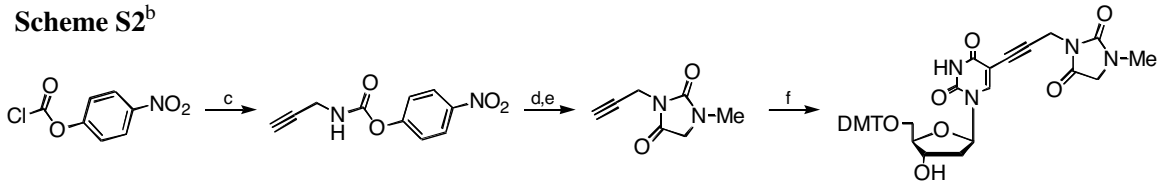
Scheme S1[a]



A **B**

[a]Conditions: (a)NaH, propargyl bromide; (b) $Pd(PPh_3)4$, $^iPr_2NH$, CuI, DMT-5IdU

Scheme S2[b]



C **D** **E**

[b]Conditinos: (c) propargyl amine, NMM; (d) sarcosine, TEA; (e) AcOH; (f) $Pd(PPh_3)4$, $^iPr_2NH$, CuI, DMT-5IdU

1-(prop-2-ynyl)-oxazolidine-2-one (A). A dry flask was charged with NaH (60% dispersion in oil, 0.92 g, 25 mmol, 1.1 equiv) and dry DMF (50 mL). Subsequently,

oxazolidine-2-one (2.02 g, 23.0 mmol, 1 equiv) was added and the solution was stirred for 1 h at 0 °C. A solution of propargyl bromide (80% in toluene, 3.5 mL, 35 mmol, 1.5 equiv) in DMF (10 mL) was added over 6 min and the reaction was warmed to rt. After 35 min, the reaction was carefully quenched with water (1 mL) and subsequently extracted with $CHCl_3$. The combined $CHCl_3$ extracts were washed with water and then saturated brine. The organic layer was dried over Na_2SO_4 and concentrated *in vacuo*. The crude product was purified by chromatography (SiO_2, 2.5% MeOH/$CHCl_3$) to yield compound **A** (1.24 g, 11 mmol, 43%) as a pale-yellow oil after drying *in vacuo*. IR (thin film): 3283, 2986, 2922, 2119 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$): δ 4.29 (2H, t, J = 6.9 Hz), 3.99 (2H, d, J = 2.4 Hz), 3.59 (2H, t, J = 6.9 Hz), 2.28 (1H, t, 2.4 Hz). ^{13}C NMR (100 MHz, $CDCl_3$): δ 157.8, 77.6, 74.3, 62.0, 43.7, 33.8. HRMS *calcd* for $C_6H_7NO_2$: 125.0477. Found: 125.0474.

***p*-Nitrophenyl *N*-(prop-2ynyl)carbamate (C).** *para*-Nitrophenylchloroformate (3.05 g, 15.1 mmol, 1 equiv) was dissolved in THF (80 mL) and the mixture was cooled to –55 °C (CH_3CN/acetone/CO_2). After 10 min, propargylamine (1.0 g, 18 mmol, 1.2 equiv) was added dropwise over 10 min. After stirring for 45 min at –55 °C, the solution was filtered through a plug of silica gel (ca. 1 g) over Celite. The solids were rinsed with THF (3 x 40 mL) and the combined eluent was concentrated. The resulting residue was dissolved in hot hexanes/ethyl acetate (~2:1) and the solution was cooled to rt. After 1 h, the mixture was cooled to –10 °C. Light yellow crystals were isolated by filtration, rinsed with hexanes (200 mL) and dried to yield compound **C** (2.58 g). A second crop of product was obtained from the mother liquor (combined: 2.92 g, 13.4 mmol, 89% total). mp 123-125 °C; IR (thin film): 3283, 3205, 2123 cm^{-1}. ^1H NMR (300 MHz, $CDCl_3$): δ 8.25 (2H, d, J = 3.3 Hz), 7.34 (2H, d, J = 3.3 Hz), 4.09 (2H, dd, J = 2.4, 5.7 Hz), 2.33 (1H, t, J = 3.3 Hz). ^{13}C NMR (75 MHz, CD): δ 155.6, 152.7, 145.0, 78.6, 72.5, 31.1. Recrystalization of this material from toluene/pentanes yielded **C** sufficiently pure for elemental analysis. Anal. *calcd* for $C_{10}H_8N_2O_4$: C,54.55; H, 3.66; N, 12.74. Found: C, 54.27; H, 3.56; N,12.57.

1-methyl-3-(prop-2-ynyl) hydantoin (D). A flask was charged with **C** (660 mg, 3.00 mmol, 1 equiv), sarcosine (534 mg, 6.00 mmol, 2 equiv) and THF (30 mL). Freshly distilled TEA (1.3 mL, 9 mmol, 3 equiv) was added and the solution was allowed to stir for 48 h at rt. The reaction was concentrated and acetic acid (10 mL) was added. The solution was allowed to stir for 30 min. and then the solution was concentrated. The resulting yellow oil was partitioned between water (10 mL) and CH_2Cl_2 (20 mL). The aqueous layer was rinsed with CH_2Cl_2 (3 x 20 mL). The organic solutions were combined, dried over $MgSO_4$ and concentrated. The yellow oil was chromatagraphed with CH_2Cl_2/MeOH (SiO_2, 2% - 10% MeOH gradient) and then with ethyl acetate/hexanes (SiO_2, 50% - 100% ethyl acetate gradient). The product containing fractions were concentrated to yield **D** as a clear oil (0.170 g, 1.12 mmol, 37%) after drying. IR (thin film): 3259, 2936, 2125, 1778, 1709 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$): δ 4.19 (2H, d, J = 2 Hz), 3.85 (2H, s), 2.95 (3H, s), 2.19 (1H, t, J = 2 Hz). ^{13}C NMR (100 MHz, $CDCl_3$): δ 168.7, 155.4, 71.5, 51.7, 29.6, 27.9. HRMS *calcd* for $C_7H_8N_2O_2$: 152.0588. Found 152.0588.

Synthesis of alkynyl nucleosides. 5'-Dimethoxytrityl-5-iodouridine was prepared as according to Ahmadian *et al.*[S1] Under an argon atmosphere, a flask was charged with this DMT-protected nucleoside (400 mg, 0.61 mmol), an alkynyl heterocycle (1.22 mmol, 2 equiv), $Pd(PPh_3)_4$ (71 mg, 0.06 mmol, 0.1 equiv), CuI (29 mg, 0.15 mmol, 0.25 equiv), diisopropylamine (404 μL, 5 equiv) and THF (4 mL). The pale yellow solution was stirred for 1 hr at rt. The solution was then filtered through a small plug of AlO_2 (ca. 500 mg). The plug was rinsed with THF (50 mL) and the combined solutions were dried on a rotary evaporator. The resulting yellow foam was chromatographed with $MeOH/CH_2Cl_2$ including 0.5% pyridine (SiO_2, 2-5% MeOH gradient). The product containing fractions were combined and dried under vacuum.

Nucleoside B. Yield: 89%. ^{1}H NMR (500 MHz, $CDCl_3$): δ 2.30 (1H, m), 2.35 (1H, s), 2.52 (1H, m), 3.00 (2H, m), 3.25 (1H, m), 3.50 (1H, m), 3.79 (6H, s), 3.84 (2H, m), 3.99 (2H, app s), 4.09 (1H, app d, J = 2 Hz), 4.52 (1H, m), 6.35 (1H, app t, J = 6.5 Hz), 6.84 (4H, m), 7.28 (9H, m), 8.22 (1H, s), 9.52 (1H, br s). ^{13}C NMR (125 MHz, $CDCl_3$): δ 34.63, 41.66, 43.41, 55.26, 61.84, 63.27, 71.95, 75.83, 85.69, 86.64, 86.90, 86.94, 99.38, 113.28, 123.87, 125.25, 126.97, 127.76, 128.04, 128.18, 128.99, 129.89, 129.94, 135.32, 135.55, 142.87, 144.62, 149.20, 157.62, 158.63, 161.58. HRMS *calcd* for $C_{36}H_{36}N_3O_9$: 654.2452. Found 654.2462.

Nucleoside E. Yield: 83%. ^{1}H NMR (500 MHz, $CDCl_3$): δ 2.24 (1H, m), 2.35 (1H, s), 2.48 (1H, m), 2.87 (3H, s), 3.38 (2H, app d, J = 3 Hz), 3.63 (1H, app d, J = 17 Hz), 3.70 (1H, app d, J = 17 Hz), 3.79 (6H, s), 4.06 (1H, m), 4.19 (1H, d, J = 17.5 Hz), 4.29 (1H, d, J = 17.5 Hz), 4.43 (1H, m), 6.22 (1H, app t, J = 6.5 Hz), 6.86 (4H, m), 7.30 (9H, m), 7.99 (1H, s), 8.30 (1H, br s). ^{13}C NMR (125 MHz, $CDCl_3$): δ 28.75, 29.52, 41.24, 51.50, 55.25, 63.47, 72.08, 73.88, 85.80, 86.35, 86.81, 86.96, 99.22, 113.30, 125.25, 126.93, 127.82, 128.02, 128.18, 128.99, 129.09, 129.94, 129.99, 135.43, 135.52, 143.16, 144.58, 149.11, 155.40, 158.57, 161.23, 168.80. HRMS *calcd* for $C_{37}H_{37}N_4O_9$: 681.2561. Found 681.2558.

DNA synthesis. DNA strands TAA**1**CC_{20} and TAA**2**CC_{20} were synthesized on an ABI DNA synthesizer employing nucleosides **B** and **E** in addition to conventional DNA synthesis reagents (Glen Research). Before synthesis, **B** and **E** were converted to the respective phosphoramidites, precipitated, dried over night and used without further purification.[S2] The strands were synthesized without the final DMT-deprotection for purification using Poly Pak II cartridges (Glen Research). After this initial purification, the DNA was further purified by reverse phase HPLC (Hamilton PRP-1 column using 0.1 M triethylammonium acetate and a 5 – 25% MeCN gradient). The purified DNA was analyzed by PAGE (20% polyacrylamide-UREA gel, visualized by UV-shadowing, see Figure S1) and MALDI-TOF MS (Bruker Reflex III, positive ion, reflectron mode, matrix: saturated 3-hydroxypicolinic acid solution in 50% MeCN combined 9:1 with 50 mg/mL ammonium citrate).

TAA1CC$_{20}$. For PAGE analysis, see Figure S1 (lane 2). MALDI-TOF: *calcd* for $[M+H]^+$: 6144.1; found: 6143.9. *Calcd* for $[M+2H]^{2+}$: 3072.5; found: 3072.8.

TAA2CC$_{20}$. For PAGE analysis, see Figure S1 (lane 3). MALDI-TOF: *calcd* for $[M+H]^+$: 6171.1; found: 6171.2. *Calcd* for $[M+2H]^{2+}$: 3086.0; found: 3086.3.

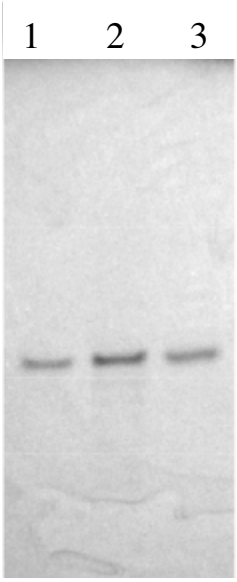


Figure S1. PAGE analysis of TAAT*CC_{20} strands. Lane 1: TAATCC_{20} (Operon); Lane 2: TAA**1**CC_{20}; Lane 3: TAA**2**CC_{20}.

Thermal denaturation experiments. Melting points of DNA duplexes were determined on a Cary-100 UV-Vis spectrophotometer using 1 cm quartz cells under the following conditions: [DNA] = 1.1 μM, 150 mM KCl 20 mM HEPES•KOH pH 7.5, 1 mM EDTA. Each experiment was performed in triplicate. The melting temperature is defined as the temperature at the maximum of the derivative of the melting curve. Melting temperatures: T_m (TAATTA_{20}) = 66.00 ± 0.01 °C; T_m (TAA**1**TA_{20}) = 66.01 ± 0.02 °C; T_m (TAA**2**TA_{20}) = 65.11 ± 0.02 °C.

Expression and purification of HDs. Expression and purification of the HDs were carried out using maltose binding fusions essentially as described.[S6] Mutations were introduced using QuikChange mutagenesis. Concentrations of the purified proteins were determined using the calculated extinction coefficient (74,250 $M^{-1}cm^{-1}$) by diluting the protein solutions 1:10 into 8 M guanidinium•HCl and measuring their absorbance at 280 nm in triplicate. The purity of the proteins was analyzed by denaturing PAGE (6 - 18% polyacrylamide, SDS/Tris-Glycine, Gradipore), staining with coomassie blue (see Figure S2).

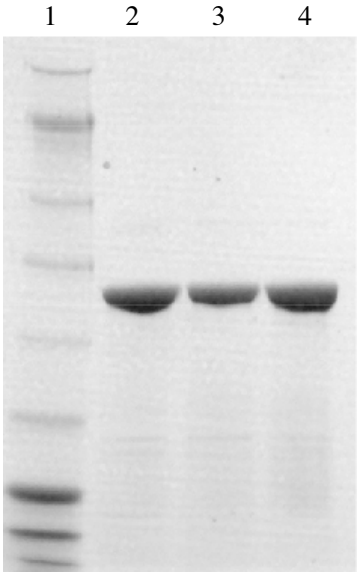


Figure S2. 8-16% PAGE analysis of MBP-HD fusions. Lane 1: SeeBlue Plus2 pre-stained standards; Lane 2: HD_0; Lane 3: HD_A; Lane 4: HD_ϕ.

Phage library construction. Similar to previous work,[S4-S6] a phagmid library was constructed using a modified form of Kunkel mutagenesis. The modified Kunkel reaction was performed using a primer constructed using split-and-pool DNA synthesis (codon randomization)[S3] with degenerate at codons at positions encoding amino acid 43, 44, 45, 46, 48, 49, 50, 51, 52 and 54. Position 47 was fixed as either Ala or Gly. The degenerate oligonucleotide was constructed on an ABI DNA synthesizer with two active columns. For each degenerate codon, the solid support was pooled by suspension in dichloroethane,10-20% of the support was removed to the "minor" column to receive the degenerate codon and the remaining 80-90% was added to the "major" column to receive the wild-type codon (with a wobble base whenever possible to facilitate identification of sister clones). The resulting sequence was 5'-AGCGAGTTGGGCCTGAACGAG(GCN/NNS)-(CAR/NNS)(AAR/NNS)GSN(TGG/NNS)(TTY/NNS)(CAR/NNS)(AAY/NNS)(AAR/NNS)CGG(GCN/NNS)AAGATCAAGAAGT-3' where N = A, T, C, or G; S = C or G; R = A or G; Y = T or C and the parenthetical codons represent the major/minor sequence incorporated at each position.

The degenerate oligonucleotide was synthesized DMT-ON and initially purified using a Sep Pak II cartridge (Glen Research) according to the manufactures instructions. Then the degenerate oligonucleotide was further purified by excision from a urea-PAGE gel. The purified oligonucleotide was used in a modified Kunkel reaction as previously reported.[S4,S5] The reaction mixture was transformed into XL1-Blue super competent cells (Stratagene) to yield 1.8 x 10^7 AU (ampicillin resistant units). After transformation, the cells were grown at 37 °C with shaking in 2YT media supplemented with ampicillin and helper phage (K07, 10^{10} phage/mL). The phage library was isolated *via* precipitation as described.[S4-S6]

Selection. Phage solutions (10 μL) were incubated with biotinylated TAA**1**CC_{20} (60 nM) in 100 μL binding buffer (100 mM KCl, 20 mM HEPES•KOH pH 7.5, 5 mM $MgCl_2$, 5%

glycerol, 0.1 mg/mL BSA) with competitor salmon-sperm DNA (250 ng/mL). The incubations were mixed gently in 1.5 mL tubes on a roller for 3 h and then added to 0.1 mg pre-rinsed magnetic beads (Dynal M270) suspended in 400 μL selection wash buffer (binding buffer with additional 0.1% Triton-X) in a fresh 1.5 mL microcentrifuge tube. The beads were incubated with the phage solution on a roller for 30 min at room temperature. Using a magnetic stand (Dynal), the beads were captured and the buffer removed by aspiration. The beads were rinsed with selection wash buffer (10x), resuspending the beads during each rinse by flicking the tubes removed from the magnetic stand.

The DNA-bound phage were eluted by resuspending the beads in 40 μL DNase I solution (1:20 in selection wash buffer, Roche RNase-free 10 u/μL) for 5 min. The beads were captured and the eluant was used to infect 1 mL of rapidly growing XL1-Blue cells (OD_{600} 0.2 – 1) for 20 min. To monitor the titer, serial dilutions (1:5) of the infected cells into LB media were plated on LB-Amp plates in 10 μL spots and incubated at 37 °C overnight. To propagate the eluted phage, the remaining cells were used to inoculate 2YT media (50 mL) supplemented with ampicillin and helper phage (K07, 10^{10} particle/mL). The phage were propagated overnight and isolated by precipitation.

The second round proceeded as the first except with higher stringency. The concentration of biotinylated TAA**1**CC_{20} was reduced to 30 nM and unbiotinylated $TAATCC_{20}$ (3 nM) was added to select against binding to unmodified DNA. In this round, the DNase I elution was allowed to proceed for only 1 min before capturing the beads.

In the third round, a Reacti-Bind NeutrAvidin Coated 96-well plate (Pierce) was used instead of magnetic beads. The plate was pre-coated for 30 min with 100 μL of biotinylated TAA**1**CC_{20} (60 nM). After rinsing with wash buffer (3X) the phage solution was added as a 1:30 dilution into 2YT supplemented with 100 mM KCl, 20 mM HEPES•KOH pH 7.5, 5 mM $MgCl_2$ and 0.015% Tween-20. After 1.5 h, the phage solutions were decanted and the wells washed with wash buffer (10X). The DNA-bound phage were eluted with 40 μL/well DNase I solution (1:20 in selection wash buffer, Roche RNase-free 10 u/μL) for 10 min. The infection and propagation were carried out as described for rounds 1 and 2. Excess phage from each round could be stored at –80 °C for later use with minimal loss of activity or increase in non-specific binding.

Phage ELISA. DNA-binding behavior of individual phage populations was examined using an ELISA in which phage bound to DNA-coated wells were detected using a HRP-conjugated anti-M13 antibody. Biotinylated-DNA solutions (60 nM DNA, 100 mM KCl, 20 mM HEPES pH 7.4, 5 mM $MgCl_2$) were added to each well (100 μL/well) of a Reacti-Bind NeutrAvidin Coated 96-well plate (Pierce) and incubated for 1 h at room temperature. The DNA solutions were decanted and the wells rinsed with wash buffer (3x) as above. Purified phage solutions (purified by ppt as described for the library above) were added as a 1:30 dilution into 2YT supplemented with 100 mM KCl, 20 mM HEPES•KOH pH 7.5, 5 mM $MgCl_2$ and 0.015% Tween-20. After 1.5 h, the phage solutions were decanted and the wells washed with wash buffer (10X)..

A 1:10,000 dilution of HRP-conjugated anti-M13 antibody (Amersham Life Science, 150 μL/well) was added in antibody buffer (stock buffer with 5% glycerol and 0.1 mg/mL BSA) and the plates were incubated with shaking at room temperature for 30

min. The antibody solution was decanted, and the plates were rinsed with wash buffer (5x) as before. The plates were developed for at least 10 min using ABTS solution (5 mg ABTS, 22 mL 50 mM citrate pH 5, 36 μL 30% H_2O_2) and the absorbance was recorded at 405 nm. See Figure S3 for results withphage displayed HD_ϕ binding to various DNA strands.

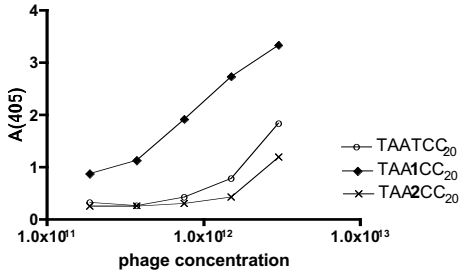


Figure S3. Phage ELISA of HD_ϕ binding to modified and unmodified DNA targets.

Conventional electrophoretic mobility shift assays. Binding of the protein to DNA was determined essentially as described,[S6] except a Cy3-labeled probe was used instead of a radio-labeled probe. The appropriate 5'-Cy5-labeled complementary strand (Operon) was hybridized 1:1.5 with, TAAT*CC_{20}. The probes (0.5 nM) were combined with increasing concentrations of protein in assay buffer (50 mM KCl, 20 mM HEPES pH 7.6, 1 mM EDTA, 5% glycerol, 0.02% NP-40). The binding reactions were incubated for at least 90 min at rt and were then run on 8% 0.5 X TBE polyacrylamide gels (pre-run for at least 30 min. at 200 V, 4°C) for ~35 min at 200 V. The fluorescence was measured on a Typhoon scanner, integrating the bound and free bands to determine the fraction bound (Θ). Apparent equilibrium dissociation constants were determined by fitting the data to the equation: $\Theta = (x + 0.5 + K_d) - ((x + 0.5 + K_d)^2 - 2x)^{1/2}$ where $\Theta = cpm_{bound}/(cpm_{bound} + cpm_{free})$ and x is the concentration of the HD. See Figure S4 for additional binding curves not shown in Figure 2B.

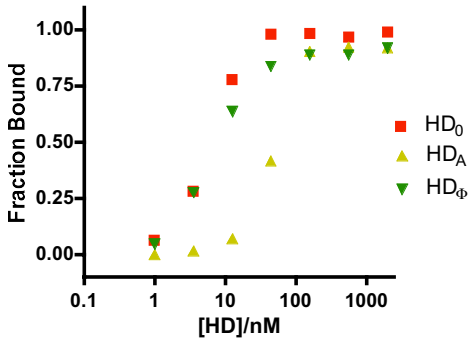


Figure S4. EMSA analysis of HDs binding to $TAATCC_{20}$.

	HD_0	HD_A	HD_ϕ
$TAATCC_{20}$	5.7 ± 1.1 nM	54 ± 12 nM	8.6 ± 1.4 nM
$TAA1CC_{20}$	> 500 nM	398 ± 38 nM	2.9 ± 0.4 nM

Table S1. Apparent equilibrium K_d values determined by EMSA.

Cy3/Cy5 competition EMSA analysis of HDs. Competition EMSA experiments were performed similar to the conventional EMSA. The HD (30 nM) was incubated with two competing probes (200 nM ea.), one Cy3-labeled, the other Cy5-labeled. These reactions were performed either in the presence or absence of competitor salmon-sperm DNA (1 mg/mL). Similar binding reactions were performed without competition to serve as Cy3 and Cy5 standards to facilitate quantification of ratio bound (each channel was normalized to its respective standard). The results were not dependent on the differential dyes used; a dye-flip experiment was performed generating essentially identical results (see Figure S5).

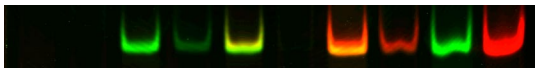


Figure S5. Dye flip experiment confirming the result seen in Figure 2A. Here TAA**1**CC_{20} was Cy5-labeled and $TAATCC_{20}$ was Cy3-labeled.

SI References

(S1) Ahmadian, M.; Zhang, P. M.; Bergstrom, D. E. *Nucleic Acids Res* **1998**, *26*, 3127-3135.
(S2) Sinha, N. D.; Biernat, J.; Mcmanus, J.; Koster, H. *Nucleic Acids Res* **1984**, *12*, 4539-4557.
(S3) Cormack, B. P.; Struhl, K. *Science* **1993**, *262*, 244-248.
(S4) Sidhu, S. S.; Lowman, H. B.; Cunningham, B. C.; Wells, J. A. In *Method Enzymol.*, **2000**; Vol. 328, pp 333-363.
(S5) Sidhu, S. S.; Weiss, G. A.; Wells, J. A. *J. Mol. Biol.* **2000**, *296*, 487-495.
(S6) Simon, M. D.; Sato K.; Weiss G.; Shokat, K. M. **2004** *submitted*