J|A|C|S
COMMUNICATIONS

Published on Web 06/11/2004

Adaptability at a Protein−DNA Interface: Re-engineering the Engrailed Homeodomain to Recognize an Unnatural Nucleotide

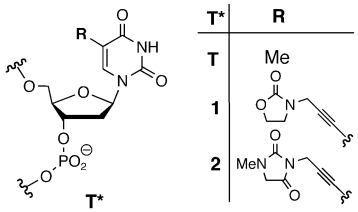
Matthew D. Simon and Kevan M. Shokat*

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

Received April 1, 2004; E-mail: shokat@cmp.ucsf.edu

Beyond sequences of DNA, biological information is stored and transmitted through noncovalent binding interactions including complementary protein−protein and protein−DNA interfaces. Some proteins contain inherent adaptability that allows them to be remodeled to bind a diverse array of ligands (e.g., antibodies).[1] However, it remains unclear how amenable other proteins, such as DNA-binding domains (DBDs), are toward re-engineering. Previous work has demonstrated that DBDs can be remodeled to bind alternative naturally occurring DNA sequences.[2] We asked if it would be possible to re-engineer a DBD to specifically recognize an unnatural, highly modified DNA structure, thereby probing the nascent adaptability in the protein−DNA interface. Herein we report the remodeling of one such DBD, engrailed Q50K homeodomain (HD_0), to specifically recognize the unnatural DNA duplex, TAA**1**CC_{20}, without compromising the high affinity and specificity of natural homeodomain−DNA interactions.

The homeodomain (HD) is a 60-amino acid DBD that has remained highly conserved over 500 million years of evolution.[3] The HD is composed of a flexible N-terminal arm followed by three α helices, the third of which contacts the DNA in the major groove as shown for HD_0 in Figure 1.[4] Binding of HD_0 to DNA is well characterized biochemically[5] and structurally.[4] HD_0 binds a nonpalindromic consensus site,[6] TAATCC, in which I47 contacts the C5-methyl of T4 in TAATCC. We reasoned that replacing the C5-methyl of thymine with a propynyl-oxazolidinone (TAA**1**CC_{20}) would preserve the B-form DNA structure[7] but would occlude HD_0 binding due to a clash between the modified nucleoside and I47. Modeling studies suggested that a large-to-small mutation at position 47 might relieve this clash, creating a recess on the protein surface to accommodate the propynyl protrusion in the major grove of the DNA.



We first synthesized several deoxythymidine analogues (T*) and incorporated them into TAAT*CC_{20} using conventional solid-phase DNA synthesis. We tested the binding of these strands to HD_0 using electrophoretic mobility shift assays (EMSA). As expected, the binding of HD_0 was greatly reduced toward TAA**1**CC_{20} as compared with that toward $TAATCC_{20}$ (Figure 2B). In fact, binding to the modified DNA was sufficiently weak that saturation was not observed even at high concentrations of HD_0. It is unlikely the loss of binding is caused by perturbation to the B-form DNA structure; propynyl substitutions are tolerated at the C5 position of thymidine[7] and thermal denaturation experiments demonstrated that this modification did not destabilize duplex DNA within experimental error.[8]

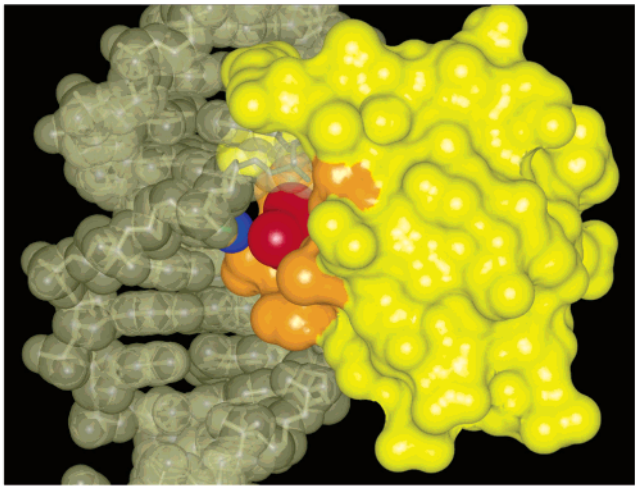


Figure 1. HD_0 (yellow) bound to DNA (gray) modeled from the reported crystal structure.[4] The C5-methyl from T4 of TAATCC (blue) contacts I47 (red) of HD_0. The residues randomized in this study are shown in orange. The sequences of the HDs employed in this study are listed, including engrailed HD Q50K (HD_o), the corresponding I47A mutant (HD_A), the library used (HD_X), and the phage-selected HD (HD_ϕ).

To efficiently determine the relative binding to these different DNA strands, we devised a novel EMSA in which equimolar concentrations of the modified and unmodified DNA directly compete for binding to the HD. The DNA strands were differentially labeled with either Cy3 or Cy5 fluorophores. After incubation of the HD with saturating concentrations of both DNA strands ([DNA] $\gg K_d$), the bound DNA was separated from the unbound DNA on a polyacrylamide gel. From the shifted band (corresponding to the HD−DNA complex), the ratio of Cy3/Cy5 fluorescence approximates the ratio of K_d values for the two different DNA strands (Figure 2A). Consistent with results from the conventional EMSA, this assay demonstrated that HD_0 binds with greater than 60-fold specificity to $TAATCC_{20}$ over TAA**1**CC_{20} (Figure 2, lane 3).

We next tested whether an I47A mutation (HD_A) could relieve the steric clash between I47 and the modified base. Although HD_A demonstrates higher affinity for TAA**1**CC_{20} than HD_0 (Figure 2B), HD_A binds more tightly to $TAATCC_{20}$ than to TAA**1**CC_{20} (Figure

10.1021/ja048113w CCC: $27.50

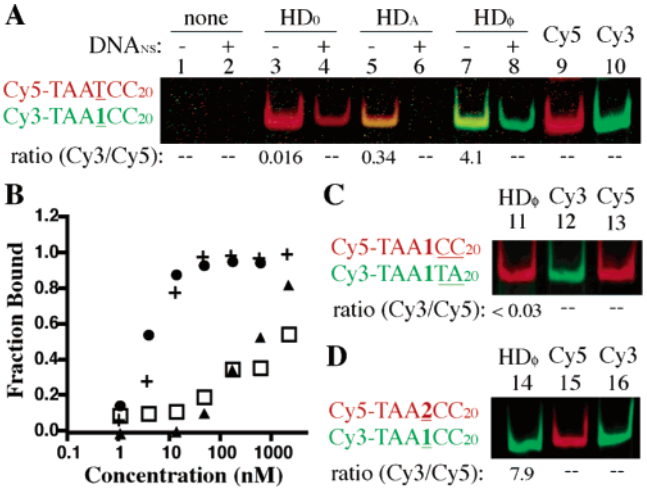


Figure 2. EMSA analysis of HD binding. (A) Cy3/Cy5 competition EMSA demonstrates that HD_0 binds specifically to $TAATCC_{20}$ (red, lane 3), whereas the HD_A is less selective (lane 5) and HD_ϕ prefers $TAA\mathbf{1}CC_{20}$ (green, lane 7). Competition with DNA_{ns} (salmon-sperm DNA, 1 mg/mL) demonstrates HD_0 and HD_ϕ bind with high specificity (lanes 4, 8) but HD_A does not (lane 6). Standards are shown in lanes 9 and 10. (B) Conventional EMSA analysis of HDs binding to $TAA\mathbf{1}CC_{20}$ (□) HD_0; (▲) HD_A; (●) HD_ϕ; (+) positive control, HD_0 binding to $TAATCC_{20}$. Cy3/Cy5 competition EMSA demonstrates HD_ϕ binds preferentially to $TAA\mathbf{1}CC_{20}$ over $TAA\mathbf{1}TA_{20}$ (C) and to $TAA\mathbf{1}CC_{20}$ over $TAA\mathbf{2}CC_{20}$ (D). For conditions see Supporting Information.

2A, lane 5). Furthermore, compared with HD_0, HD_A exhibited a ~10-fold reduction in binding affinity.[8] Finally, HD_A lacks high sequence specificity as evidenced by the loss of the bound complex upon treatment with salmon-sperm DNA at concentrations tolerated by HD_0 (lanes 4 versus lane 6). These data are consistent with a disruption of the highly tuned HD−DNA interface.

Although simple rational design proved unsuccessful, we sought to regain high specificity and affinity at the modified HD−DNA interface using phage display. Employing oligonucleotide-directed mutagenesis, a library was constructed with mutations in 11 codons in HD_0 (Figure 1, HD_X). The mutagenic oligonucleotide was constructed using a split-and-pool DNA synthesis technique (codon-randomization)[9] to bias the library toward members with four or fewer mutations.

Using this library and conditions similar to those previously reported,[10] we selected mutant HDs capable of binding to $TAA\mathbf{1}CC_{20}$. After three rounds of selection, a single clone (HD_ϕ) dominated the pool. DNA sequencing revealed that, in addition to a space creating I47G mutation, this clone contains two other mutations in the hydrophobic core of the protein (I45V and K52M). Presumably, these changes allow subtle repacking of the HD that improves binding to the unnatural DNA. Phage ELISA experiments demonstrated that HD_ϕ binds more tightly to $TAA\mathbf{1}CC_{20}$ than to $TAATCC_{20}$.[8]

To further characterize HD_ϕ, we overexpressed this HD in *Escherichia Coli* as a maltose binding protein fusion and examined its binding behavior by EMSA (Figure 2B). Consistent with results from the phage ELISA, we found that HD_ϕ binds tightly to $TAA\mathbf{1}CC_{20}$ ($K_d = 2.8 \pm 0.4$ nM). In fact, the affinity of HD_ϕ binding to $TAA\mathbf{1}CC_{20}$ is ~2-fold tighter than HD_0 binding to its target, $TAATCC_{20}$. Also consistent with the phage ELISA, HD_ϕ binds preferentially to $TAA\mathbf{1}CC_{20}$ over $TAATCC_{20}$ (Figure 2A, lane 7). Furthermore, HD_ϕ binding to $TAA\mathbf{1}CC_{20}$ is sequence specific, as demonstrated by a greater than 40-fold preference for binding to $TAA\mathbf{1}CC_{20}$ over $TAA\mathbf{1}TA_{20}$ (Figure 2C, lane 11) and competition experiments with salmon-sperm DNA (Figure 2A, lane 8).

To test if HD_ϕ is specific for the oxazolidinone presented in the major groove of the DNA, binding to $TAA\mathbf{1}CC_{20}$ was compared with binding to a hydantoin-substituted DNA, $TAA\mathbf{2}CC_{20}$ (Figure 2D, lane 14). Gratifyingly, HD_ϕ binds preferentially to $TAA\mathbf{1}CC_{20}$, supporting the conclusion that HD_ϕ makes contacts specifically tuned for the oxazolidinone-bearing DNA.

These data demonstrate that HD_0 was re-engineered to specifically recognize $TAA\mathbf{1}CC_{20}$ with high affinity. Furthermore, the preference for $TAA\mathbf{1}CC_{20}$ over $TAA\mathbf{2}CC_{20}$ confirms that the engineered HD, beyond merely accommodating bulky substitution, demonstrates specificity that is sensitive to subtle changes in the chemical structure of the unnatural nucleotide. We conclude that the interface of the re-engineered HD binding to $TAA\mathbf{1}CC_{20}$ is highly tuned, thereby demonstrating a surprising level of nascent adaptability at the HD−DNA interface. This adaptability may prove useful to engineer novel functionality into the HD−DNA interface. For example, the engineered hole in HD_ϕ could be viewed as an oxazolidinone binding pocket, offering a new route to small-molecule control[11] of HD−DNA interactions.

Acknowledgment. We thank Prof. Carolyn Bertozzi for generously providing space in her laboratories, David King (HHMI MS facility), and the UCSF National Research Resource (NIH BRTP 01614). Funding was provided by NIH (RO1 EB-001987) and graduate fellowships from the NSF and the organic division of the ACS to M.D.S.

Supporting Information Available: Experimental procedures and characterizations. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) For a discussion of antibody plasticity, see: (a) Yin, J.; Beuscher, A. E.; Andryski, S. E.; Stevens, R. C.; Schultz, P. G. *J. Mol. Biol.* **2003**, *330*, 651−656. For a review see: (b) Schultz, P. G.; Yin, J.; Lerner, R. A. *Angew Chem., Int. Ed.* **2002**, *41*, 4427−4437. For examples of protein−protein interface remodeling see: (c) Atwell, S.; Ultsch, M.; DeVos, A. M.; Wells, J. A. *Science* **1997**, *278*, 1125−1128 and (d) Hiipakka, M.; Poikonen, K.; Saksela, K. *J. Mol. Biol.* **1999**, *293*, 1097−1106.
(2) For examples and reviews, see: (a) Greisman, H. A.; Pabo, C. O. *Science* **1997**, *275*, 657−661. (b) Segal, D. J.; Beerli, R. R.; Blancafort, P.; Dreier, B.; Effertz, K.; Huber, A.; Koksch, B.; Lund, C. V.; Magnenat, L.; Valente, D.; Barbas, C. F. *Biochemistry* **2003**, *42*, 2137−2148. (c) Beerli, R. R.; Barbas, C. F. *Nat. Biotechnol.* **2002**, *20*, 135−141. (d) Jamieson, A. C.; Miller, J. C.; Pabo, C. O. *Nat. Rev. Drug Discov.* **2003**, *2*, 361−368.
(3) For a review, see: (a) Gehring, W. J.; Qian, Y. Q.; Billeter, M.; Furukubotokunaga, K.; Schier, A. F.; Resendezperez, D.; Affolter, M.; Otting, G.; Wuthrich, K. *Cell* **1994**, *78*, 211−223. For discussion of the molecular evolution of HDs, see (b) Banerjee-Basu, S.; Baxevanis, A. D. *Nucleic Acids Res.* **2001**, *29*, 3258−3269.
(4) TuckerKellogg, L.; Rould, M. A.; Chambers, K. A.; Ades, S. E.; Sauer, R. T.; Pabo, C. O. *Structure* **1997**, *5*, 1047−1054.
(5) (a) Ades, S. E.; Sauer, R. T. *Biochemistry* **1994**, *33*, 9187−9194. (b) Ades, S. E.; Sauer, R. T. *Biochemistry* **1995**, *34*, 14601−14608.
(6) Wild-type engrailed HD binds the palindromic DNA sequence, TAATTA, thereby complicating the interpretation of biochemical experiments. To avoid this difficulty, we chose to work with the Q50K mutant which binds TAATCC.
(7) See: He, J. L.; Seela, F. *Nucleic Acids Res.* **2002**, *30*, 5485−5496 and references therein.
(8) See Supporting Information.
(9) Cormack, B. P.; Struhl, K. *Science* **1993**, *262*, 244−248.
(10) Simon, M. D.; Sato K.; Weiss G.; Shokat, K. M. Manuscript submitted.
(11) For examples of other approaches to engineer small-molecule control over protein−protein and protein−DNA interactions, see: (a) Arndt, H. D.; Hauschild, K. E.; Sullivan, D. P.; Lake, K.; Dervan, P. B.; Ansari, A. Z. *J. Am. Chem. Soc.* **2003**, *125*, 13322−13323. (b) Lin, Q.; Barbas, C. F.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 612−613. (c) Guo, Z. H.; Zhou, D. M.; Schultz, P. G. *Science* **2000**, *288*, 2042−2045.

JA048113W