J|A|C|S COMMUNICATIONS

Published on Web 02/27/2003

Miniature Homeodomains: High Specificity without an N-Terminal Arm

Jin Kim Montclare[†] and Alanna Schepartz*[,†,‡]

Department of Chemistry and Department of Molecular, Cellular and Developmental Biology, Yale University, P.O. Box 208217, New Haven, Connecticut 06520-8107

Received September 20, 2002; E-mail: Alanna.Schepartz@yale.edu

Recently, we described a strategy for the design of miniature proteins[1] that bind DNA and protein surfaces with high affinity and selectivity.[2−4] This strategy, which we call protein grafting, involves identifying the set of residues from an α-helix required for macromolecular recognition by a natural protein—the functional epitope—and presenting them in the same arrangement on the small, stable protein scaffold, avian pancreatic polypeptide (aPP) (Figure 1A).[5,6] The stability and function of a miniature protein designed in this way can often be optimized by combinatorial evolution.[3,7] Previously, high-affinity DNA recognition could be achieved only when the miniature protein contained the *complete* functional epitope, that is, when it included every energetically significant DNA contact residue.[2] Here we demonstrate that miniature proteins can achieve high affinity and selectivity for DNA, by design, even when the functional epitope is incomplete.

Homeodomains are highly conserved transcription factors with diverse roles in eukaryotic development.[8,9] Homeodomains bind DNA using a bipartite structural unit: residues on helix 3 of a helix−turn−helix motif[10,11] provide energetically significant contacts in the major groove, whereas residues along a flexible N-terminal arm provide energetically significant contacts in the minor groove.[12−14] In the case of the Q50K *engrailed* homeodomain (Q50K), major and minor groove contacts are made to distinct base pairs within the QRE site $T_1A_2A_3T_4C_5C_6$ (Figure 1A).[15−18] Residues K_{46}, I_{47}, W_{48}, K_{50}, N_{51}, R_{53}, K_{55}, K_{57}, and K_{58} on helix 3 contact A_3, T_4, G_{-5}, and G_{-6} within the major groove, while residues R_3, R_5, and T_6 along the N-terminal arm contact T_1 and T_{-2} within the minor groove (Figure 1).[15−19]

Our design of a miniature homeodomain began by aligning the sequences of Q50K helix 3 and the aPP α-helix (Figure 1B).[5,6] In alignment #1, DNA contact residues from helix 3 are presented near the N terminus of the aPP α-helix, while in alignment #2, they are shifted one helical turn closer to the C-terminus. Although alignment #1 lacked conflicts between residues implicated in QRE binding and aPP folding, it was disfavored by an expected poor backbone superposition between helix 3 and the N terminus of the aPP α-helix (RMSD = 1.31 Å). Alignment #2, while characterized by favorable backbone superposition (RMSD = 0.47 Å), possessed conflicts between three DNA contact residues on helix 3 (W_{48}, K_{55} and K_{58}) and three folding residues along the aPP α-helix (F_{20}, Y_{27}, and V_{30}). Using alignment #2 and favoring DNA binding over folding at two of three conflict positions (K_{55} and K_{58}), we designed PPeng4 as a starting point for a subsequent library of variants; helix 3 and eng4 were prepared for purposes of comparison (Figure 1B).

Quantitative electrophoretic mobility shift experiments were performed to determine the affinities of PPeng4, eng4, and helix 3 for DNA containing a QRE target site (QRE_{20}; TAATCC).[2,3] To our surprise, the PPeng4·QRE_{20} complex was exceptionally stable, with an equilibrium dissociation constant of 17 ± 8 nM at 25 °C

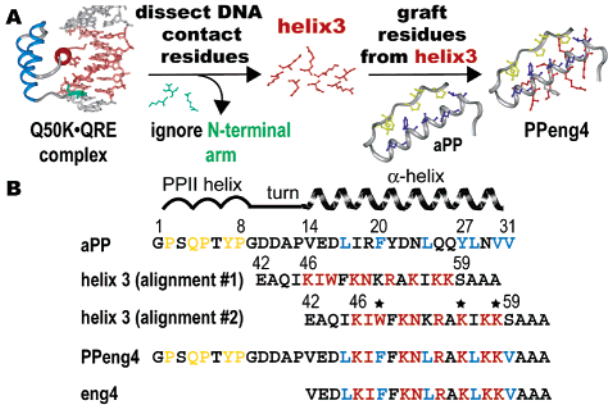

Figure 1. (A) Design of PPeng4. (B) Two alignments of aPP with residues 42−59 (helix 3) of Q50K. Q50K residues that contact the major groove are in red; residues along the PPII and α-helices of aPP that contribute to protein stability are in yellow or blue, respectively. Stars identify conflicts between residues required for DNA binding and aPP folding.

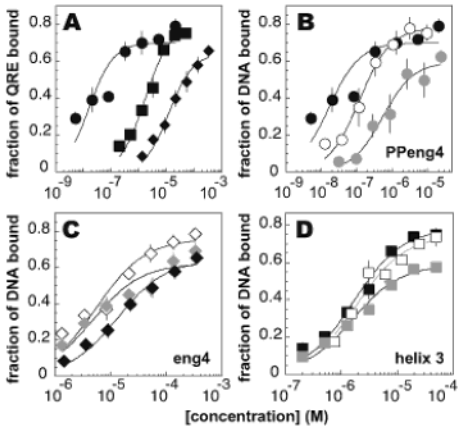

Figure 2. The affinity (A) and specificity (B−D) of DNA recognition by PPeng4, eng4 and helix 3. Plots illustrating (A) the fraction of QRE_{20} (TAATCC) bound by PPeng4 (●), helix 3 (■), and eng4 (◆) and the fraction of QRE_{20} (black), ERE_{20} (white; TAATTA), and MRE_{20} (gray; CCATCC) bound by (B) PPeng4, (C) eng4, and (D) helix 3 at 25 °C. All points represent the average of at least three trials. Error bars denote the standard error.[20] Binding reactions were performed and analyzed as described.[2,3,22]

(Figure 2A). The QRE complex of PPeng4 was 750 times more stable than the corresponding complex of eng4, which lacked the PPII helix and β-turn in PPeng4 (K_d = 13 ± 2 μM), and 100 times more stable that the corresponding complex of helix 3 (K_d = 1.7 ± 0.3 μM). These comparisons indicate that the PPII helix and β-turn in PPeng4 increased QRE_{20} affinity by as much as 2.8 kcal·mol^{-1}. Previous work had suggested that miniature proteins lacking even one DNA contact residue would fail to detectably bind DNA, even at 4 °C.[2] PPeng4 achieved high QRE affinity at 25 °C despite the absence of one DNA contact residue from helix 3 (W_{48}), and, more importantly, three from the N-terminal arm.[14−18]

[†] Department of Chemistry.
[‡] Department of Molecular, Cellular and Developmental Biology.

10.1021/ja028628s CCC: $25.00

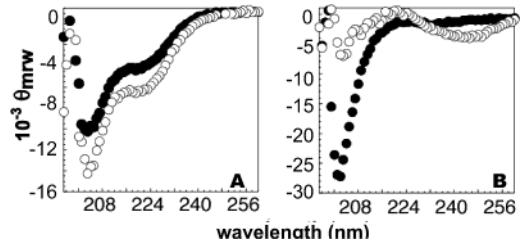

Figure 3. CD analysis of (A) PPeng4 (25 °C) and (B) helix 3 (4 °C) at 5 μM in the absence (black) or presence (white) of 5 μM QRE_{20}.[21,22]

Circular dichroism (CD) experiments were performed to characterize the secondary structure of PPeng4 in the presence and absence of QRE_{20}. The CD spectrum of PPeng4 revealed significant α-helix in the absence of QRE_{20}[21] (Figure 3A), with a mean residue ellipticity (Θ_{mrw}) at 222 nm of −3951 deg·cm^2·dmol^{-1}.[22] Addition of 1 equiv of QRE_{20} amplified the CD signal at 222 nm to −6239 deg·cm^2·dmol^{-1}. By contrast, the CD spectrum of helix 3 (Figure 3B) showed no α-helical structure in the absence or presence of QRE_{20}.[21] These experiments suggest that the affinity of PPeng4 for QRE_{20} resulted from significant pre-organization of the helix 3 functional epitope. In this case, a well-folded miniature protein capable of high-affinity DNA recognition was achieved by design, without combinatorial evolution[3,7]

The DNA specificity of PPeng4 was investigated using two sequences used previously to examine homeodomain specificity: ERE_{20} and MRE_{20} (Figure 2B).[15,18,22] ERE_{20} (TAATTA) differed from QRE_{20} at two base pairs contacted by K_{50} and K_{46} within helix 3, whereas MRE_{20} (CCATCC) differed at two base pairs contacted by R_3 and R_5 on the N-terminal arm.[15,17,18] Neither eng4 nor helix 3 discriminated between QRE_{20} and either MRE_{20} or ERE_{20}. In each case, the specificity ratio *R* (defined as the ratio of the equilibrium dissociation constants of mutant and specific complexes) was near unity. Eng4 and helix 3 bound ERE_{20} with affinities of 12 ± 2 and 1.9 ± 0.4 μM (R = 1 and 0.9, respectively) and bound MRE_{20} with affinities of 4.9 ± 1.6 and 1.6 ± 0.2 μM (R = 0.4 and 1, respectively) (Figure 2C, D). By contrast, PPeng4 discriminated QRE_{20} from ERE_{20} and, especially, from MRE_{20}. PPeng4 preferred QRE_{20} to ERE_{20} (K_d = 120 ± 20 nM) with a specificity ratio of 7; the *R* value reported for Q50K is 36[18] (Figure 2B). However, PPeng4 preferred QRE_{20} to MRE_{20} (540 ± 160 nM) by a factor of 32, a value only 2-fold lower than that reported for Q50K (R = 64).[18] Remarkably, PPeng4 accurately specified base pairs T_1 and T_{-2} of the QRE target site despite the absence of those residues that contact these base pairs within the Q50K·QRE complex,[15,16] residues that contribute >3.8 kcal·mol^{-1} of binding free energy.[17] This observation implies that base pairs T_1 and T_{-2} are specified indirectly by PPeng4 (and perhaps Q50K *engrailed*) via interactions between helix 3 and the adjacent DNA major groove. Our results indicate that the PPII helix and β-turn in PPeng4 enhance *both* affinity and specificity, using 13 amino acids to replace the remaining 43 residues—more than two-thirds—of a bipartite homeodomain motif. More broadly, the success of the PPeng4 design suggests that structural pre-organization can effectively compensate for the free energy of lost protein−DNA contacts. In this case it has been possible to miniaturize both the recognition surface and the structural framework of a globular protein fold.

Acknowledgment. We thank the HHMI Biopolymer/Keck Foundation Biotechnology Resource Laboratory at the Yale University School of Medicine for peptide and DNA synthesis and analysis and the NIH (GM 59843) for support of this work. J.K.M. was supported by an NSF predoctoral fellowship.

Supporting Information Available: Experimental details (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) For leading references to previous work on miniature proteins that are well-folded or bind macromolecular targets, see: (a) Braisted, A. C.; Wells, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5688−5692. (b) Lombardi, A.; Bryson, J. W.; Ghirlanda, G.; DeGrado, W. F. *J. Am. Chem. Soc.* **1997**, *119*, 12378−12379. (c) Imperiali, B.; Ottesen, J. J. *Biopolymers* **1998**, *47*, 23−29. (d) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. *Annu. Rev. Biochem.* **1999**, *68*, 779−819. (e) Skerra, A. *J. Mol. Recognit.* **2000**, *13*, 167−187. (f) Cochran, A. G.; Skelton, N. J.; A., S. M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5578−5583. (g) Beerli, R. R.; Barbas, C. F. *Nat. Biotechnol.* **2002**, *20*, 135−141.
(2) Zondlo, N. J.; Schepartz, A. *J. Am. Chem. Soc.* **1999**, *121*, 6938−6939.
(3) Chin, J. W.; Schepartz, A. *J. Am. Chem. Soc.* **2001**, *123*, 2929−2930.
(4) Chin, J. W.; Schepartz, A. *Angew. Chem., Int. Ed.* **2001**, *40*, 3806−3809.
(5) Blundell, T. L.; Pitts, J. E.; Tickle, I. C.; Wood, S. P.; Wu, C.-W. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4175−4179.
(6) Glover, I.; Haneef, I.; Pitts, J.; Wood, S.; Moss, D.; Tickle, I.; Blundell, T. *Biopolymers* **1983**, *22*, 293−304.
(7) Chin, J. W.; Grotzfeld, R. M.; Fabian, M. A.; Schepartz, A. *Biorg. Med. Chem. Lett.* **2001**, *12*, 1501−1505.
(8) Biggin, M. D.; McGinnis, W. *Developments* **1997**, *124*, 4425−4433.
(9) Baxter, S. M.; Gontrum, D. M.; Phillips, C. L.; Roth, A. F.; Dahlquist, F. W. *Biochemistry* **1994**, *33*, 15309−15320.
(10) Wintjens, R.; Rooman, M. *J. Mol. Biol.* **1996**, *262*, 294−313.
(11) Gehring, W. J.; Affolter, M.; Burglin, T. *Annu. Rev. Biochem.* **1994**, *63*, 487−526.
(12) Liu, B. S.; Kissinger, C. R.; Pabo, C. O.; Martin-Blanco, E.; Kornberg, T. B. *Biochem. Biophys. Res. Commun.* **1990**, *171*, 257−259.
(13) Clarke, N. D.; Kissinger, C. R.; Desjarlais, J.; Gilliland, G. L.; Pabo, C. O. *Protein Sci.* **1994**, *3*, 1779−1787.
(14) Kissinger, C. R.; Liu, B.; Martin-Blanco, E.; Kornberg, T. B.; Pabo, C. O. *Cell* **1990**, *63*, 579−590.
(15) Tucker-Kellogg, L.; Rould, M. A.; Chambers, K. A.; Ades, S. E.; Sauer, R. T.; Pabo, C. O. *Structure* **1997**, *5*, 1047−1054.
(16) Fraenkel, E.; Rould, M. A.; Chambers, K. A.; Pabo, C. O. *J. Mol. Biol.* **1998**, *284*, 351−361.
(17) Ades, S. E.; Sauer, R. T. *Biochemistry* **1994**, *33*, 9187−9194.
(18) Ades, S. E.; Sauer, R. T. *Biochemistry* **1995**, *34*, 14601−14608.
(19) Residue T_6 on the Q50K N-terminal arm contacts the A_2A_3 internuclear phosphate.
(20) Sokal, R. R.; Rohlf, F. J. *Introduction to biostatistics*; 2nd ed.; W. H. Freeman and Company: New York, 1987.
(21) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry*; Freeman: New York, 1980.
(22) See Supporting Information for details.

JA028628S