Concerted Evolution of Structure and Function in a Miniature Protein

Jason W. Chin and Alanna Schepartz*

*Department of Chemistry, Yale University
New Haven, Connecticut 06520-8107*

Unstructured peptides merely garnished with functional groups used for recognition rarely, if ever, possess function. By contrast, presentation of the same functional groups by a globular protein can result in a high affinity ligand. Recently we described[1] a strategy for the design of miniature functional proteins[2,3] in which DNA-binding residues from the bZIP protein GCN4 were grafted onto the solvent-exposed α-helix of avian pancreatic polypeptide (aPP).[4] aPP contains a close-packed hydrophobic core comprised of six residues from the α-helix (L17, F20, L24, Y27, L28, V30, and V31) and four residues from a type II polyproline (PPII) helix (G1, P2, Q4, P5, and Y7) (Figure 1).[4] In our first grafted peptide (PPBR4),[1] Y27, L28, and V30 were changed to arginine to preserve DNA affinity, and V31 was changed to alanine to improve helical propensity. Although PPBR4 bound DNA well at 4 °C, these mutations destroyed the aPP core. As a result, PPBR4 exhibited only nascent helicity at 4 °C and no evidence of DNA binding at ambient temperature. Mutations at positions 27, 28, 30, and 31 failed to produce molecules that were *both* helical and bound DNA.[5]

We reasoned that some combinations of amino acids at positions 1, 2, 4, 5, and 7 of PPBR4 might rebuild the damaged core and pre-organize the α-helix into a conformation closer to that required for DNA binding. If so, peptides containing such sequences should bind specific DNA with high affinity. We conceived to locate appropriately pre-organized miniature proteins by affinity purification of a library of PPBR4 analogues containing mutations in the N-terminal PPII helix.

Two phage libraries were created to identify appropriately folded PPBR4 analogues (Figure 1). The members of libraries A and B differ from PPBR4 at 3 (library A) or 4 (library B) positions on the PPII helix. The proline residues retained at positions 2 and 5 of library A are highly conserved among PP-fold proteins.[8] We anticipated that retention of these two prolines would effectively constrain the conformational space available to library A members[9] and that most would contain N-terminal PPII helices. Such conformational constraints are absent in library B, acknowledging that there may be many ways to stabilize DNA-bound α-helices.[10] Since the amino acids at positions 2 and 5 of library B are not restricted to proline, we anticipated that this library would sample a larger fraction of available $\varphi-\psi$ space.
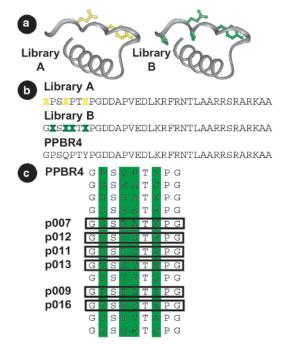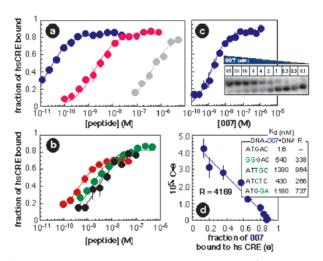

Figure 1. (a) Residues of PPBR4 targeted for variation mapped onto the crystal structure of aPP.[4] Side chains varied in library A are in yellow, those varied in library B are in green. (b) Sequences of PPBR4 and the two libraries. Residues varied are indicated by an X. Each position was randomized at the DNA level using the NNS codon scheme.[6] (c) Sequences of the N-terminal amino acids deduced from the DNA sequences of library B clones after three selection rounds. Peptides containing the boxed sequences followed by the remaining residues of PPBR4 were synthesized and their properties investigated *in vitro*.

Phage were sorted for three rounds on the basis of their ability to bind an oligonucleotide duplex containing the sequence ATGAC (hsCRE). To favor identification of sequences that bound hsCRE with high affinity at ambient temperature, two rounds of selection at 4 °C were followed by a single round at room temperature. By the final round, library A phage were retained at a level only comparable to PPBR4 phage and were not considered further. Library B phage were retained at a level comparable to PPBR4 phage after the first round, but at levels 15–16 times greater than PPBR4 phage after the subsequent two rounds. Twelve library B clones were sequenced (Figure 1c) after round three. Six sequences (p007, p009, p011, p012, p013, and p016) were synthesized, and the DNA-binding properties of four were analyzed in detail.
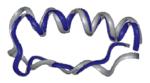
Quantitative electrophoretic mobility shift experiments were performed to assess the DNA affinities of p007, p011, p012, and p016 (Figure 2). All peptides tested bound hsCRE as well or better than did PPBR4 or G27 (the isolated basic region of GCN4). At 4 °C, p011 and p012 bound hsCRE with affinities of 1.5 ± 0.2 nM and 2.5 ± 0.5 nM, whereas p016 bound hsCRE with an affinity of 300 ± 60 pM. Of particular interest is p007, which bound hsCRE to form an exceptionally stable complex with a dissociation constant of 23 ± 1.2 pM (Figure 2a). This peptide bound specific DNA approximately 100 times better than did PPBR4 ($K_d = 1.9 \pm 0.2$ nM) and approximately 20 000 times better than did G27 ($K_d = 410 \pm 53$ nM).[1] Moreover, at 25 °C p007 bound hsCRE with an affinity of 1.6 ± 0.1 nM (Figure 2c). Neither PPBR4 nor G27 showed evidence of DNA binding at this temperature. P007 binds specific DNA considerably more tightly than two fingers from the Tramtrack zinc finger protein, which binds 5 bp of DNA with an affinity of 400 nM.[11]

(1) Zondlo, N. J.; Schepartz, A. *J. Am. Chem. Soc.* **1999**, *121*, 6938.
(2) Cunningham, B. C.; Wells, J. A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 457. Braisted, A. C.; Wells, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5688.
(3) Vita, C.; Roumestand, C.; Toma, F.; Menez, A. *Proc. Natl. Acad. U.S.A.* **1995**, *92*, 6404. Lombardi, A.; Bryson, J. W.; Ghirlanda, G.; DeGrado, W. F. *J. Am. Chem. Soc.* **1997**, *119*, 12378. Vita, C. et al.*, Proc. Natl. Acad. U.S.A.* **1999**, *96*, 13091. McColl, D. J.; Honchell, C. D.; Frankel, A. D. *Proc. Natl. Acad. U.S.A.* **1999**, *96*, 9521. Domingues, H.; Cregut, D.; Sebald, W.; Oschkinat, H.; Serrano, L. *Nat. Struct. Biol.* **1999**, *6*, 652. Segal, D. J.; Barbas, C. F., III. *Curr. Opin. Chem. Biol.* **2000**, *4*, 34.
(4) 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.
(5) Zondlo, N. Ph.D. Thesis, Yale University, 1999.
(6) Kay, B. K.; Winter, J.; McCafferty, J. In *Phage Display of Peptides and Proteins*; Academic Press: San Diego, 1995.
(7) Greisman, H. A.; Pabo, C. O. *Science* **1997**, *275*, 657.
(8) Li, X.; Sutcliffe, M. J.; Schwartz, T. W.; Dobson, C. M. *Biochemistry* **1992**, *31*, 1245.
(9) Adzubei, A. A.; Sternberg, M. J. E. *Protein Sci.* **1994**, *3*, 2395.
(10) Baranger, A. M. *Curr. Opin. Chem. Biol.* **1998**, *2*, 18.

Figure 2. Plots illustrating the fraction of hsCRE bound (θ) as a function of (a) p007 (blue ●), PPBR4 (pink ●), and G27[1] (gray ●) and (b) p016 (red ●), p011 (green ●), and p012 (black ●) concentration at 4 °C. Each point represents the average of at least three trials. Error bars show the standard error. (c) Autoradiogram and plot illustrating the fraction of hsCRE bound (Θ) by p007 at 25 °C. (d) Plot illustrating the relative affinity of p007 for γ-[^{32}P] hsCRE vs CT DNA. The molar concentration of competing binding sites on CT DNA (C) was estimated from the concentration of DNA in mg/mL and the molecular weight of a base pair assuming that every base on either DNA strand represents the start of a competitor site. The inset shows the affinity of p007 for four duplexes that differ from hsCRE at two of five base pairs. Binding reactions were performed and analyzed as described.[1,7]

The specificity of DNA binding was investigated by determining the affinity of p007 for several duplex oligonucleotides containing two bp changes within the five bp hsCRE sequence. P007 was extremely discriminating, exhibiting specificity ratios R (defined as the ratio of the dissociation constants of mutated and specific complexes) between 200 and 800 ($\Delta\Delta G = -3.3 - 4.0$ kcal·mol^{-1}). This high level of discrimination was observed across the entire five bp hsCRE sequence (Figure 2d, inset), suggesting that no single interaction dominated the free energy of the p007·hsCRE complex and that the binding energy is partitioned across the entire protein−DNA interface. By contrast, at 4 °C PPBR4 discriminates poorly ($\Delta\Delta G = -1.7$ kcal·mol^{-1}) against sequences possessing mutations at the 5′ terminus of hsCRE. We suggest that the exceptional specificity of p007 results from structural imprinting of the correctly folded peptide in complex with hsCRE by functional selection.

To investigate the possibility that DNA sequences other than these four might bind p007 tightly, we measured the affinity of p007 for calf thymus DNA (CT DNA) which possesses a potential binding site in every register on either DNA strand. The average specificity ratio for recognition of hsCRE in preference to any site in CT DNA was greater than 4100 (Figure 2d). This ratio is considerably greater than the number of potential competitor sites ($4^5 = 1024$).[7] Whereas the triple zinc finger construct Zif268 and variants thereof selected by phage display fail to uniquely specify one to two base pairs of their nine bp binding sites,[7] p007 completely specifies all five base pairs of its target sequence. In fact, even if each possible five base pair competitor site were

Figure 3. Residues 1−31 of the minimized mean structure of p007 (in blue) superimposed on the corresponding residues of aPP (in gray).

present at equal molarity to the target site, 80% of the p007 molecules would be bound to hsCRE, despite the effects of mass action.

Multidimensional NMR experiments allowed us to characterize the structure of p007 in greater detail. The backbone and side-chain connectivities in p007 were assigned on the basis of reasonably disperse NOESY spectra. The presence of amide−amide cross-peaks and cross-peaks between residues at positions i and $i+3$ and i and $i+4$ defined an α-helical conformation for residues 14−30. Eleven long-range NOEs between residues 8 and 17, 8 and 20, 7 and 20, 5 and 20, 4 and 27, 2 and 29, and 2 and 30 specify a folded structure that superimposes on residues 5−8 and 15−28 of aPP with a backbone rmsd of 1.6 Å. Thus, the main chain folds of p007 and aPP are remarkably similar, with residues 5, 7, and 8 proximal to residue 20 and residues 1 and 2 proximal to residue 30. As in previous studies of pancreatic fold polypeptides,[4] the PPII helix proposed for residues 1−8 of p007 is under-defined by the NMR data. However, in light of the similarity between the folds of aPP and p007, we suggest that p007 contains a structure similar to a PPII helix. If so, it is notable that p007 was selected from a pool of molecules in which the prolines were varied and thus possessed many backbone conformations. It is possible that PPII helices, because of their extended yet regular structure, represent a general solution to the problem of α-helix recognition (Figure 3).

We have shown that protein grafting,[1] in combination with combinatorial mutation and selection, provides rapid access to folded, functional, miniature proteins.[2,3,12] Protein grafting involves installing on a miniature scaffold those residues−the functional epitope−responsible for recognition in their native context. As a result, the conformational stability of the grafted protein is largely independent of factors that contribute to the stability of the *native* functional epitope. This observation suggests that protein grafting could provide ready access to miniature versions of other proteins that use an α-helix to recognize macromolecules. It also suggests that this access could be achieved in a manner irrespective of the tertiary complexities of these proteins. Further, the successful selection of a folded, active miniature protein supports the hypothesis that functional selection of an organized active site may have driven evolution of protein receptors, alleviating the need for an otherwise exhaustive search of sequence space.[13]

Supporting Information Available: 2D NOESY data for p007 (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA0056668

(11) Fairall, L.; Schwabe, J. W. R.; Chapman, L.; Finch, J. T.; Rhodes, D. *Nature* **1993**, *366*, 483. Fairall, L.; Harrison, S. D.; Travers, A. A.; Rhodes, D. *J. Mol. Biol.* **1992**, *226*, 349.

(12) Nygren, P. A.; Uhlén, M. *Curr. Opin. Struct. Biol.* **1997**, *7*, 463. Shogren-Knaak, M. A.; Imperiali, B. *Bioorg. Med. Chem.* **1999**, *7*, 1993.
(13) Yomo, T.; Saito, S.; Sasai, M. *Nat. Struct. Biol.* **1999**, *6*, 743.