

## Highly Specific DNA Recognition by a Designed Miniature Protein

Neal J. Zondlo and Alanna Schepartz\*

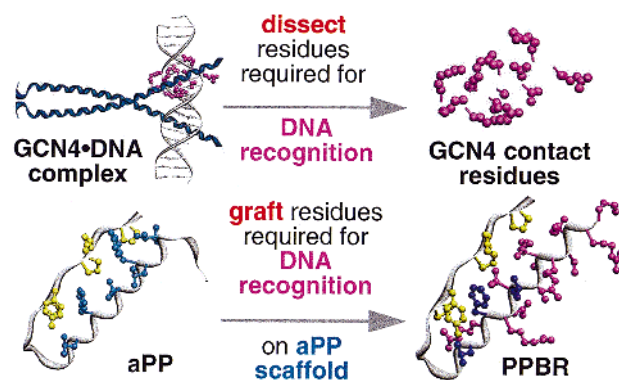
Department of Chemistry  
Yale University, P.O. Box 208107  
New Haven Connecticut 06520-8107

Received March 25, 1999

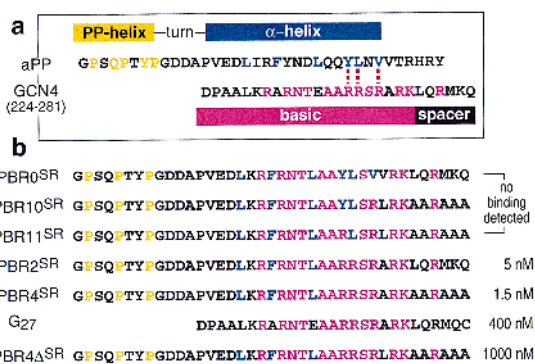
Many DNA-binding proteins employ a short  $\alpha$ -helix as their primary DNA recognition element.<sup>1–5</sup> When removed from their natural contexts, however, most  $\alpha$ -helices neither fold nor bind DNA with high affinity.<sup>6,7</sup> Here we describe a strategy for the design of miniature proteins that present a solvent-exposed  $\alpha$ -helix able to recognize DNA with high affinity and specificity. This strategy involves dissecting those  $\alpha$ -helical residues required for DNA recognition from their native protein context and grafting them on a stable, miniature protein core (Figure 1). This strategy was used to design a 42-amino acid protein displaying the DNA-contact residues of a single basic region-leucine zipper (bZIP) protein. This miniature protein bound DNA in the low nanomolar concentration range under physiological solution conditions and exhibited greater sequence specificity than a natural bZIP protein. This strategy may represent a general approach to the design of small, folded proteins that recognize nucleic acid and protein targets with high affinity and specificity.

Our design process began with avian pancreatic polypeptide (aPP), a small, well-folded protein of known structure consisting of a single  $\alpha$ -helix stabilized by hydrophobic interaction with a type II polyproline helix (Figure 1).<sup>8–10</sup> Since formation of the aPP core requires residues on only one face of the aPP  $\alpha$ -helix (shown in blue), the opposite, solvent-exposed face of the  $\alpha$ -helix is available for recognition of other macromolecular targets. By grafting various combinations of those thirteen residues (shown in pink) used by GCN4 to recognize the CRE half site (ATGAC, hsCRE)<sup>1,2</sup> on the solvent-exposed  $\alpha$ -helical face of aPP, we generated a series of polyproline helix-basic region (PPBR<sup>SR</sup>) molecules containing most or all of the DNA-contact residues of GCN4 and most or all of the folding residues of aPP (Figure 2).

This procedure generated three positions of conflict, where essential DNA-contact and aPP-folding residues occupied a single position on the helix (Figure 2). No significant DNA binding was detected with peptides PPBR0<sup>SR</sup>, PPBR10<sup>SR</sup>, and PPBR11<sup>SR</sup>, which lacked one or more of these DNA-contact residues. High-affinity DNA binding was observed with a peptide that contained these three residues: The equilibrium dissociation constant ( $K_d$ ) of the PPBR2<sup>SR</sup>·hsCRE complex was 5 nM under conditions of physiological ionic strength (Figure 3). DNA affinity was enhanced further by selective alanine substitutions that increased the overall  $\alpha$ -helical propensity of the peptide, producing the



**Figure 1.** Protein grafting strategy for the design of miniature DNA-binding proteins.



**Figure 2.** (A) Alignment of the aPP and the GCN4 basic-spacer segment sequences used to guide protein design. Essential DNA-contact residues within GCN4 are in pink; essential folding residues within aPP are in yellow or blue. Conflict positions are indicated by a dashed line. (B) Peptides used and their affinities for hsCRE<sub>24</sub>. Equilibrium dissociation constants of stable PPBR<sup>SR</sup>·hsCRE complexes are listed at right. All peptides except G<sub>56</sub><sup>28</sup> and G<sub>27</sub> contained GGC sequences at their carboxyl termini;<sup>29</sup> G<sub>27</sub> contained a single cysteine. The carboxy-terminal cysteine was alkylated with bromoacetamide to study protein monomers (PPBR<sup>SR</sup>, G<sub>27</sub>) or oxidized to study disulfide-linked dimers (PPBR<sup>SS</sup>).

PPBR4<sup>SR</sup>·hsCRE<sub>24</sub> complex whose  $K_d$  was 1.5 nM under identical conditions. Formation of the PPBR4<sup>SR</sup>·hsCRE<sub>24</sub> complex was unaffected by high concentrations of poly (dIdC)·(dIdC) or a scrambled CRE site (NON), indicating that the high stability of PPBR4<sup>SR</sup>·hsCRE<sub>24</sub> was not due primarily to nonspecific ionic interactions. Circular dichroism experiments indicated that like bZIP peptides,<sup>6,11</sup> PPBR4<sup>SR</sup> attained a fully  $\alpha$ -helical conformation only in the presence of specific DNA.<sup>12</sup> Although others have described monopartite DNA recognition by basic segment peptides, the affinities reported have been only moderate (60 nM–3  $\mu$ M), and the complexes are stable only in very low ionic strength buffers.<sup>13,14</sup> PPBR4<sup>SR</sup> represents the first example of high affinity, monopartite, major groove recognition at physiological ionic strength.

To examine the contribution of hydrophobic core formation to PPBR4<sup>SR</sup>·hsCRE<sub>24</sub> complex stability, we studied peptides G<sub>27</sub>

(1) O'Neil, K. T.; Hoess, R. H.; DeGrado, W. F. *Science* 1990, 249, 774.

(2) The CD spectrum of PPBR4<sup>SR</sup> was unchanged between 1 and 20  $\mu$ M, indicating that no detectable changes in secondary structure occurred in this range. Addition of hsCRE DNA significantly increased the  $\alpha$ -helix content of PPBR4<sup>SR</sup>; smaller changes were observed upon addition of hsCEBP DNA (see Supporting Information).

(3) Park, C.; Campbell, J. L.; Goddard, W. A., III *J. Am. Chem. Soc.* 1996, 118, 4235.

(4) Morii, T.; Yamane, J.; Aizawa, Y.; Makino, K.; Sugiura, Y. *J. Am. Chem. Soc.* 1996, 118, 10011.

\* To whom correspondence should be addressed.

(1) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* 1992, 71, 1223.

(2) König, P.; Richmond, T. *J. Mol. Biol.* 1993, 233, 139.

(3) Aggarwal, A. K.; Rodgers, D. W.; Drottler, M.; Ptashne, M.; Harrison, S. C. *Science* 1988, 242.

(4) Ferre-D'Amare, A. R.; Prendergast, G. C.; Ziff, E. B.; Burley, S. K. *Nature (London)* 1993, 363, 38.

(5) Kissinger, C. R.; Liu, B.; Martin-Blanco, E.; Kornberg, T. B.; Pabo, C. O. *Cell* 1990, 63, 579.

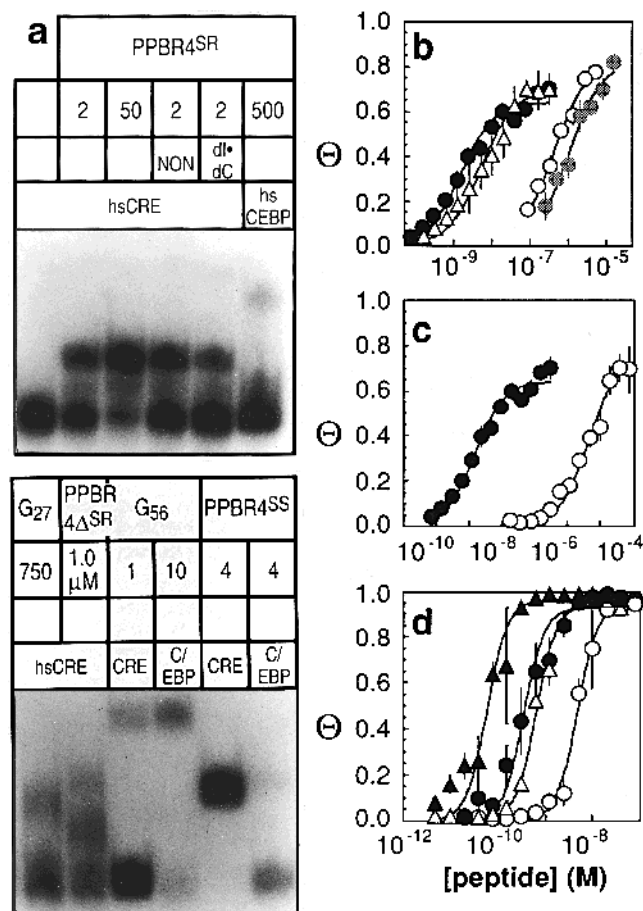
(6) Weiss, M. A.; Ellenberger, T.; Wobbe, C. R.; Lee, J. P.; Harrison, S. C.; Struhl, K. *Nature (London)* 1990, 347, 575.

(7) O'Neil, K. T.; Shuman, J. D.; Ampe, C.; DeGrado, W. F. *Biochemistry* 1991, 30, 9030.

(8) 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.

(9) Tonan, K.; Kawata, Y.; Hamaguchi, K. *Biochemistry* 1990, 29, 4424.

(10) Hermans, J., Jr.; Scheraga, H. A. *J. Am. Chem. Soc.* 1961, 83, 3283.



**Figure 3.** DNA recognition by miniature PPBR proteins. (A) Autoradiogram illustrating binding at the protein concentrations shown in nM (unless indicated otherwise) to  $^{32}\text{P}$  DNA (AGTGGAGXCTCGTGC): hsCRE<sub>24</sub>, X = ATGACAGCTA; hsCEBP<sub>24</sub>, X = ATTGCAGCTA; CRE<sub>24</sub>, X = ATGACGTCAT; CEBP<sub>24</sub>, X = ATTGCGCAAT in the presence or absence of unlabeled, nonspecific DNA (NON, X = TAAGGCCTAT) or poly (dIdC)·(dIdC).<sup>30,31</sup> (B) The fraction of hsCRE<sub>24</sub> bound (Θ) as a function of (●) PPBR<sub>4</sub><sup>SR</sup>; (Δ) PPBR<sub>2</sub><sup>SR</sup>; (○) G<sub>27</sub>; or (gray circle) PPBR<sub>4</sub>Δ<sup>SR</sup> concentration. (C) The relative affinity of PPBR<sub>4</sub><sup>SR</sup> for (●) hsCRE<sub>24</sub> and (○) hsCEBP<sub>24</sub>. (D) The relative affinity of G<sub>56</sub> for (●) hsCRE<sub>24</sub>, (○) hsCEBP<sub>24</sub>, (▲) CRE<sub>24</sub> and (Δ) CEBP<sub>24</sub>. Binding reactions were performed at 4 °C in phosphate-buffered saline [2.7 mM KCl, 137 mM NaCl, 4.3 mM Na<sub>2</sub>HPO<sub>4</sub>, and 1.4 mM KH<sub>2</sub>PO<sub>4</sub> (pH 7.3)] containing 5% glycerol, 1 mM EDTA, 0.4 mg/mL BSA, and 0.05% NP-40.

and PPBR<sub>4</sub>Δ<sup>SR</sup> (Figure 2). G<sub>27</sub> lacked the polyproline helix and turn, whereas PPBR<sub>4</sub>Δ<sup>SR</sup> contained D-tryptophan at position 4 and leucine at position 31. Modeling studies suggested that these substitutions would disrupt core formation by kinking the polyproline or the α-helix. The stabilities of the G<sub>27</sub>·hsCRE<sub>24</sub> and PPBR<sub>4</sub>Δ<sup>SR</sup>·hsCRE<sub>24</sub> complexes were 3.1 and 3.2 kcal·mol<sup>-1</sup> lower, respectively, than that of PPBR<sub>4</sub><sup>SR</sup>·hsCRE<sub>24</sub> complex (Figure 3). These data indicate that hydrophobic core formation stabilized the PPBR<sub>4</sub><sup>SR</sup>·hsCRE<sub>24</sub> complex by as much as 3 kcal·mol<sup>-1</sup>.

The sequence specificity of PPBR<sub>4</sub><sup>SR</sup> was examined by comparing its affinity for hsCRE<sub>24</sub> to that for hsCEBP<sub>24</sub>, a sequence containing the half-site recognized by C/EBP bZIP proteins (Figure 2).<sup>15</sup> This half-site (ATTGC) differs from the CRE half-site (ATGAC) by two base pairs and provides an excellent measure of base pair specificity.<sup>16,17</sup> PPBR<sub>4</sub><sup>SR</sup> displayed

remarkable specificity for hsCRE<sub>24</sub> (Figure 3). The specificity ratio  $K_{\text{rel}}$  ( $K_{\text{rel}} = K_{\text{d}}(\text{hsCRE})/K_{\text{d}}(\text{hsCEBP})$ ) describing preferred recognition of hsCRE<sub>24</sub> by PPBR<sub>4</sub><sup>SR</sup> was 2600 ( $\Delta\Delta G = -4.4$  kcal·mol<sup>-1</sup>). By contrast, G<sub>56</sub>, which comprised the bZIP element of GCN4, displayed low specificity. Specificity ratios of 118 and 180 were observed for binding of CRE<sub>24</sub> by G<sub>56</sub> in preference to CEBP<sub>24</sub> and hsCRE<sub>24</sub> in preference to hsCEBP<sub>24</sub>, ( $\Delta\Delta G = -2.6$  and  $-2.9$  kcal·mol<sup>-1</sup>), respectively. The relative specificities of G<sub>56</sub> and PPBR<sub>4</sub><sup>SR</sup> were most recognizable when one considered the concentration of each protein required to bind one-half of the two DNAs: For PPBR<sub>4</sub><sup>SR</sup>, this difference corresponded to a ratio of 2600, whereas for G<sub>56</sub>, it corresponded to a ratio of 11. PPBR<sub>4</sub><sup>SR</sup> more readily distinguished the two base pair difference between hsCRE<sub>24</sub> and hsCEBP<sub>24</sub> than G<sub>56</sub> distinguished CRE<sub>24</sub> from hsCEBP<sub>24</sub>, two sequences that differed by six of ten base pairs. These comparisons emphasize that PPBR<sub>4</sub><sup>SR</sup> was considerably more selective than was GCN4, the protein on which its design was based.

Recognition of a unique site in the human genome requires discrimination of a specific sequence of approximately 16 consecutive base pairs from all other possible sequences.<sup>18</sup> While ligands with extremely high DNA affinities have been created from linear arrays of proteins<sup>19–22</sup> or amino acids<sup>23–25</sup> that individually possess defined sequence preferences, the ability to differentiate between closely related sequences remains a considerable challenge. Molecules whose design is based on multimerization typically tolerate, with minimal losses in affinity, target sites in which as many as half of the base pairs are changed from those in the intended target site.<sup>19,20,22,26</sup> In addition, the long kinetic lifetimes of these complexes result in trapping of the molecules on nonintended targets.<sup>22,27</sup> Indeed, the disulfide-linked dimer of PPBR<sub>4</sub>, PPBR<sub>4</sub><sup>SS</sup>, confirmed these limitations: PPBR<sub>4</sub><sup>SS</sup> bound CRE<sub>24</sub> tightly ( $K_{\text{d}} = 165$  pM) but differentiated CRE<sub>24</sub> from CEBP<sub>24</sub> with a moderate specificity ratio of 39. Monomeric PPBR<sub>4</sub><sup>SR</sup>, by virtue of rapid equilibration with its preferred target site and high specificity, overcame the limitations of multimeric recognition.

**Acknowledgment.** N.J.Z. was supported by an NSF predoctoral fellowship and by the National Foundation for Cancer Research. This work was supported by the NIH (GM 53829).

**Supporting Information Available:** Figure showing CD analysis and experimental details (PDF). This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA990968Z

- (16) Suckow, M.; Von Wilcken-Bergmann, B.; Müller-Hill, B. *EMBO J.* **1993**, *12*, 1193.  
 (17) Johnson, P. F. *Mol. Cell. Biol.* **1993**, *13*, 6919.  
 (18) Dervan, P. B. *Science* **1986**, *232*, 464.  
 (19) Liu, Q.; Segal, D. J.; Ghiara, J. B.; Barbas, C. F., III *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 5525.  
 (20) Park, C.; Campbell, J. L.; Goddard, W. A. *J. Am. Chem. Soc.* **1995**, *117*, 6287.  
 (21) Kim, J.-S.; Kim, J.; Cepek, K. L.; Sharp, P. A.; Pabo, C. O. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 36.  
 (22) Kim, J.-S.; Pabo, C. O. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2812.  
 (23) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature (London)* **1998**, *391*, 468.  
 (24) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1113.  
 (25) White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 260.  
 (26) Pomerantz, J. L.; Sharp, P. A.; Pabo, C. O. *Science* **1995**, *267*, 93.  
 (27) Pomerantz, J. L.; Wolfe, S. A.; Pabo, C. O. *Biochemistry* **1998**, *37*, 965.  
 (28) Metallo, S. J.; Schepartz, A. *Chem. Biol.* **1994**, *1*, 143.  
 (29) O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Science* **1989**, *243*, 538.  
 (30) Garner, M. M.; Revzin, A. *Nucl. Acids Res.* **1981**, *9*, 3047.  
 (31) Fried, M.; Crothers, D. M. *Nucl. Acids Res.* **1981**, *9*, 6505.  
 (32) Cuenoud, B.; Schepartz, A. *Science* **1993**, *259*, 510.

(15) Agre, P.; Johnson, P.; McKnight, S. *Science* **1989**, *246*, 922.