Abstract

Characterization of Miniature DNA-binding Proteins

Loretta Yang

2005

This dissertation describes the characterization of miniature DNA-binding proteins designed to be high affinity and high specificity ligands. Avian pancreatic polypeptide was used as the scaffold for protein grafting to preorganize binding epitopes and optimize their function using in vitro selections.

Chapter 1 describes an attempt to generate miniature proteins that bind DNA containing the nonnatural self pair, PICS:PICS. PICS:PICS presents a uniquely hydrophobic surface in the major groove of DNA, distinct from native base pairs. In addition to enhancing our knowledge of protein•DNA interactions, PICS:PICS-binding proteins have utility as tools in research and biotechnology, as well as being potential therapeutics. Towards this end, phage display was used to select for miniature proteins based on p007 that recognize DNA containing PICS:PICS pair to produce a complex that is functionally orthogonal to protein•DNA complexes found in Nature. The selections succeeded in generating miniature proteins that bound DNA with nanomolar affinity, but little specificity, meriting a detailed study of miniature proteins that bind natural DNA.

Chapter 2 describes the further characterization of the GCN4 mimic, p007, which binds DNA containing the half-site CRE (hsCRE) sequence ATGAC, with nanomolar affinity and high specificity. A complete alanine scanning mutagenesis study of the miniature protein p007 was performed to determine the contributions of individual
residues to specific DNA affinity, nonspecific DNA affinity, and protein structure. As expected, variants containing alanine in place of known DNA-binding residues possess severely diminished DNA affinity and specificity. We also find that variants containing alanine in place of residues within the hydrophobic core also possess diminished DNA affinity and specificity. In addition, the importance of most of the residues selected by phage display was confirmed by diminished affinity and specificity of the corresponding alanine containing variants. While all variants were less structured than p007 as determined by circular dichroism spectroscopy, there was little correlation between structure and DNA-binding affinity or specificity. Finally, we find that no alanine substitutions improve upon the hsCRE specificity of p007, indicating that every residue makes a direct or indirect contribution to specificity.
Table of Contents

Abstract .................................................................................................................. i
Title Page .................................................................................................................. iii
Copyright ................................................................................................................... iv
Table of Contents ...................................................................................................... v
List of Figures ........................................................................................................... ix
List of Tables ........................................................................................................... xi
Acknowledgements .................................................................................................. xii

CHAPTER 1

Introduction .............................................................................................................. 2

Molecular recognition ............................................................................................... 2

The use of protein scaffolds in molecular recognition ............................................ 2
Homologous protein scaffolds for displaying functional epitopes ....................... 2
Homologous protein secondary structure scaffolds for displaying functional epitopes ................................................................. 5
Minimized proteins scaffolds for displaying functional epitopes ........................... 5
Small, nonhomologous protein scaffolds for displaying functional epitopes .......... 8

Alpha helices in macromolecular recognition ....................................................... 11

Stabilization of isolated alpha helices .................................................................... 14
Alpha helical propensities ....................................................................................... 14
Macrodipole stabilization ...................................................................................... 14
Capping .................................................................................................................. 15
Helix initiation .................................................................................................... 18
i, i + n interactions .............................................................................................. 18

Protein scaffolds for displaying alpha helical epitopes .................................... 18

Protein grafting ................................................................................................... 23

DNA-binding proteins ......................................................................................... 30
Helix-turn-helix (HTH) proteins .................................................................30
Zinc-containing proteins ........................................................................31
Basic region leucine zippers (bZIps) and related proteins .........................31

Designed DNA-binding proteins ................................................................32
Proteins that bind nonnatural DNA ..........................................................36
Nonnatural base pairs ..............................................................................37
Applications of miniature proteins that recognize nonnatural base pairs ....39

Results ........................................................................................................43
Experimental goals ..................................................................................43
Library 1 ....................................................................................................46
Library 2 ....................................................................................................47
Phage display selections with Library 1 ....................................................52
Phage display selections with Library 2 ....................................................57
EMSA screening of selected sequences ....................................................62
$K_d$ determination for selected peptides at 25 °C .....................................64

Discussion ................................................................................................67
Sequences of selected peptides ...............................................................67
Affinity of selected peptides ....................................................................67
Specificity of selected peptides ...............................................................68

Experimental ............................................................................................82
General ......................................................................................................82
Materials ....................................................................................................83
Media and buffers .....................................................................................84
Biomolecule synthesis and characterization ..............................................84
DNA for panning ......................................................................................85
Library construction ................................................................................85
Insert construction ....................................................................................85
Vector construction ..................................................................................87
Ligations ....................................................................................................88
Library quality ..........................................................................................89
Phage display selections .........................................................................90
Bead preparation and DNA loading .......................................................90
Panning .......................................................................................................91
Peptides .....................................................................................................91
CHAPTER 2

Introduction ............................................................................................................. 109
    Miniature proteins based on avian pancreatic polypeptide .......................... 109
    Alanine scanning mutagenesis ..................................................................... 113

Results ...................................................................................................................... 118

    Experimental goals ........................................................................................ 119

    Quantitative analysis of the contributions of individual side chains to DNA affinity. 119
        Contributions of residues identified by functional selection ...................... 125
        Contributions of putative DNA-binding residues .................................... 125
        Contributions of putative folding residues ................................................. 129
        Contributions of other residues within the type II polyproline helix .......... 130
        Contributions of turn residues ................................................................. 130
        Contributions of other alpha helix residues ............................................. 131

    Quantitative analysis of the contributions of individual side chains to DNA specificity
                                                                            ................................................. 132
        Contributions of residues identified by functional selection ...................... 133
        Contributions of putative DNA-binding residues .................................... 137
        Contributions of putative folding residues ................................................. 138
        Contributions of other residues within the type II polyproline helix .......... 138
        Contributions of turn residues ................................................................. 138
        Contributions of other alpha helix residues ............................................. 139

    Contributions of individual side chains to protein structure as measured by circular
dichroism .................................................................................................................. 139

Discussion ............................................................................................................. 144

    P007 binds DNA much like GCN4 .................................................................... 145
    Residues throughout p007 contribute to specificity .................................... 152
    The relationship between affinity and specificity ............................................ 154
    The relationship between inherent alpha helicity and DNA binding ............ 155

Experimental ......................................................................................................... 167
General ........................................................................................................................................... 167
Peptide synthesis .................................................................................................................................. 168
Peptide purification .............................................................................................................................. 169
DNA ...................................................................................................................................................... 175
Radiolabeling DNA .............................................................................................................................. 176
Electrophoretic mobility shift assay .................................................................................................. 176
Competition electrophoretic mobility shift assay ............................................................................. 178
Calculation of specificity ................................................................................................................... 179
Far UV circular dichroism ................................................................................................................ 180

References ........................................................................................................................................... 181
List of Figures

CHAPTER 1

**Figure 1.1** Homologous protein scaffolds for displaying functional epitopes .......................... 3

**Figure 1.3** Minimized protein scaffolds for displaying functional epitopes .............................. 9

**Figure 1.4** Small, nonhomologous protein scaffolds for displaying functional epitopes
.................................................................................................................................................. 12

**Figure 1.5** Methods of alpha helix stabilization ........................................................................... 16

**Figure 1.6** Protein scaffolds for alpha helical epitope stabilization ............................................. 21

**Figure 1.7** Avian pancreatic polypeptide ..................................................................................... 24

**Figure 1.8** Protein grafting in the context of DNA recognition .................................................. 27

**Figure 1.9** The crystal structures of major classes of transcription factors ................................. 33

**Figure 1.10** Nonnatural base pairs ............................................................................................... 40

**Figure 1.11** The design of Libraries 1 and 2 ............................................................................... 44

**Figure 1.12** Plasmid map of the Library 1 phagemid vector ....................................................... 48

**Figure 1.13** Plasmid map of the Library 2 phagemid vector ....................................................... 50

**Figure 1.14** Panning of Library 1 ............................................................................................... 54

**Figure 1.15** Panning of Library 2 ............................................................................................... 59

**Figure 1.16** The affinity of TAMR for DNA at 25 °C measured by electrophoretic
mobility shift assay ......................................................................................................................... 70

**Figure 1.17** The affinity of HRPVQ for DNA at 25 °C measured by electrophoretic
mobility shift assay ......................................................................................................................... 72

**Figure 1.18** The affinity of RPNST for DNA at 25 °C measured by electrophoretic
mobility shift assay ......................................................................................................................... 74

**Figure 1.19** The affinity of SRTMS for DNA at 25 °C measured by electrophoretic
mobility shift assay ......................................................................................................................... 76
CHAPTER 2

Figure 2.1  Miniature aPP-based proteins that target DNA ......................... 111
Figure 2.2  Miniature aPP-based proteins that target alpha helix-binding proteins .... 114
Figure 2.3  Miniature aPP-based proteins that target proteins which bind proline-rich sequences .............................................................. 116
Figure 2.4  Variants of p007 used in this study. ........................................ 120
Figure 2.5  Plots illustrating the affinity of each p007 variant for hsCRE DNA .... 122
Figure 2.6  The free energy of binding and specificity of p007 variants ............ 126
Figure 2.7  Plots illustrating the relative affinity of p007 variants for hsCRE DNA vs. CT DNA .............................................................................................................. 134
Figure 2.8  CD spectra of p007 variants at 10 μM in PBS at 4 °C .................... 141
Figure 2.9  Basic regions of representative bZIPS ........................................ 146
Figure 2.10 Plot of ΔG_{hsCRE} vs. ΔG_{spec} of p007 and all variants .......... 156
Figure 2.11 Plot of ΔG_{hsCRE} vs. MRE_{222} of p007 and all variants .......... 160
Figure 2.12 Plot of ΔΔG_{spec} vs. MRE_{222} of p007 and all variants .......... 163
Figure 2.13 Mapping the effect of individual alanine substitutions on the affinity, specificity, and structure of p007 ............................................................... 165
List of Tables

CHAPTER 1

Table 1.1 Binding and washing conditions used in Library 1 phage display selections. 53
Table 1.2 Selected sequences from Library 1 ................................................................. 56
Table 1.3 Binding and washing conditions used in Library 2 phage display selections. 58
Table 1.4 Selected sequences from Library 2 ................................................................. 61
Table 1.5 Initial binding affinity analysis of selected peptides by EMSA at 4 ºC ........ 64
Table 1.6 Initial specificity analysis of selected peptides by EMSA .............................. 65
Table 1.7 Binding affinity of selected peptides for DNA targets at 25 ºC ................. 66
Table 1.8 DNA sequences used as targets in EMSA experiments ............................... 69

CHAPTER 2

Table 2.1 Affinities of peptides for hsCRE DNA at 25 ºC ................................. 124
Table 2.2 Affinities of peptides for CT DNA at 25 ºC ............................................. 136
Table 2.3 Mean residue ellipticities at 222 nm of peptides at 4 ºC ............................ 143
Acknowledgements
CHAPTER 1

Selection of miniature DNA-binding proteins that recognize a target sequence containing

a nonnatural base pair
Introduction

Molecular recognition

Molecular recognition is the formation of noncovalent complexes between macromolecules. These seemingly straightforward events are crucial to all processes throughout a cell’s life cycle, yet they are not well understood. Numerous factors contribute to specific recognition including shape complementarity, charge interactions, flexibility, and burial of hydrophobic surfaces. Proteins that mimic or disrupt molecular recognition events would be useful tools to characterize these complexes or as potential therapeutics. However, the creation of such molecules is nontrivial. Small molecules that are effectively used in binding small enzyme active sites may not be effective in molecular recognition events that involve burial of a large and/or shallow surface area. Thus, protein scaffolds are a useful starting point in protein design.

The use of protein scaffolds in molecular recognition

Homologous protein scaffolds for displaying functional epitopes

The most straightforward application of protein scaffolds is the transfer of functional epitopes between homologous proteins. Only the functional epitope differs, and the overall tertiary structure is maintained, as demonstrated by the humanization of antibodies (Figure 1.1). These therapeutic antibodies have reduced immunogenicity in human patients. In general, only those parts of a mouse antibody that are directly involved in binding to its specific target are transplanted onto a human antibody. These antigen binding sites are variable domains consisting of a beta sheet framework with
Figure 1.1 Homologous protein scaffolds for displaying functional epitopes. Humanization of foreign antibodies reduces immunogenicity when used as therapeutics in human patients. To humanize a mouse antibody, the complementarity determining regions (CDRs) of the mouse antibody (red) are transferred to the human antibody (white) (1).
Complementarity determining regions

**Glycosylation**

**Mouse**

**Human**

**Humanized**

- Complementarity determining regions
- Glycosylation
hypervariable regions, or complementarity determining regions (CDRs). The first example reported by Winter and coworkers used the CDRs of the mouse antibody B1-8 that binds the hapten 4-hydroxy-3-nitrophenacetyl caproic acid (NP-cap) to replace the corresponding CDRs of a human myeloma protein (1). The new chimeric antibody bound NP-cap with similar affinity to the mouse antibody. This proof of concept experiment demonstrated that function could be maintained when transferring binding epitopes between homologous protein scaffolds.

**Homologous protein secondary structure scaffolds for displaying functional epitopes**

The above example has proven utility, but it is a very limited approach since it requires grafting onto homologous tertiary structures. A more flexible approach grafts binding epitopes onto homologous secondary structures. For example, interleukin-4 (IL-4) is a four-helix bundle protein that binds the IL-4 receptor (IL-4Rα), and the GCN4 leucine zipper is a similarly structured coiled coil. The functional epitope of IL-4 resides on two adjacent antiparallel helices, which can be recapitulated on the coiled coil of the GCN4 scaffold (Figure 1.2) (2). This strategy was not completely successful; the rationally designed IL-4 mimic bound IL-4Rα with a K_d of only 5 µM, while the parent IL-4 protein bound the target with a K_d of 1.4 nM. This weak binding likely results from the absence of a number of IL-4Rα-contacting residues. These residues could not be included in the designed protein without significant perturbation of the GCN4 coiled coil scaffold.

**Minimized proteins scaffolds for displaying functional epitopes**
Figure 1.2 Homologous protein secondary structure scaffolds for displaying functional epitopes. The functional epitope of IL-4 was transferred to the structurally homologous region of the GCN4 coiled coil (2). Energy-minimized model of the rationally designed molecule (red) superimposed on helix A and C of IL-4 (grey), showing the side chains involved in binding to IL-4Rα. Basic residues Lys7, Arg11, Lys14, Arg15, and Arg18 are displayed in blue. Acidic residue Glu20 is displayed in red. The two hydrophobic residues Trp21 and Ile24 and the disulfide bridge are shown in yellow. The rationally designed protein bound the IL-4α receptor with poor affinity relative to the IL-4 protein.
Rationally designed IL-4 mimic based on GCN4 coiled coil ($K_d$ for IL-4Rα = 5 µM)

IL-4
($K_d$ for IL-4Rα = 1.4 nM)
Protein scaffolds with homologous secondary or tertiary structures are useful, but often a protein scaffold of minimal size is desired for ease of study and manipulation. Wells and coworkers describe protein minimization as the paring down of a protein through rational design and/or combinatorial optimization, without losing the ability to interact with its binding partner (3,4). This strategy was used to minimize the three-helix bundle Z-domain of protein A, which binds IgG with high affinity ($K_d = 10 \text{ nM}$) (Figure 1.3). The functional epitope of the Z-domain is displayed on surface of two helices, stabilized by a third helix. Removal of this stabilizing helix decreased the affinity for IgG, but a functional selection generated a two-helix variant that bound IgG with an affinity only 4-fold lower than that of the parent Z-domain. Using further rational design based on structural information, a disulfide linkage and other mutations were introduced to improve the structure and function of the minimized Z-domain (5). The minimization of the Z-domain of protein A reduced the size of the protein ligand from 59 to 34 amino acids and resulted in a stable, functional, and more tractable model for studying protein•protein interactions. However, this strategy may not prove to be generally applicable for other proteins of interest because minimization of a tertiary structure can result in a loss of a well-folded structure, and consequently loss of function.

Small, nonhomologous protein scaffolds for displaying functional epitopes

A more general approach to generating functional miniature proteins is to combine epitope transfer and protein minimization strategies by substituting recognition residues from a protein of interest to a region of secondary structural homology in a smaller scaffold protein. In one case, Vita and colleagues transferred the functional
**Figure 1.3** Minimized protein scaffolds for displaying functional epitopes. The tertiary structure stabilizing the functional epitope of the Z-domain of protein A was reduced using rational design, phage display, and the introduction of a disulfide constraint (4, 5). The minimized mean NMR structure shows all side chains colored by amino acid type (hydrophobic residues are yellow; positively charged residues are blue; negatively charged residues are red; and polar residues are green). The resulting 34 residue minimized Z-domain bound IgG-Fc with affinity comparable to that of the 59 residue parent Z-domain.
Minimized protein A

Rational design
Phage display
Disulfide constraint

Minimized protein A
epitope of the CDR2-like loop of CD4 onto the β-hairpin of scyllatoxin, a scorpion toxin protein (Figure 1.4) (6). The rationally designed 27 residue chimeric miniature protein, CD4M3, inhibited the interaction between CD4 and the HIV-1 envelope glycoprotein gp120 with an IC$_{50}$ of 40 µM. NMR studies and alanine scanning mutagenesis provided further insights for protein design, which led to an 100-fold enhancement in gp120-binding affinity relative to CD4M3. However, the second generation miniature protein, CD4M9, was still 100-fold less active than the parent CD4 protein. Recently, NMR studies of CD4M9 led to five additional mutations, including two nonnatural amino acid substitutions, which improved gp120 binding affinity an additional 200- to 600-fold (7). The optimized miniature protein, CD4M33, inhibited the gp120•CD4 complex with an IC$_{50}$ of 7.5 nM, similar to that of the parent CD4 protein.

**Alpha helices in macromolecular recognition**

The alpha helix was first described by Pauling and Corey more than 50 years ago (8-10) and experimentally validated soon after by Kendrew’s X-ray structure of myoglobin (11,12). The alpha helix is a ubiquitous motif in macromolecular recognition, mediating diverse protein•protein and protein•nucleic acid interactions. Molecules that could mimic the function of recognition helices, and therefore recapitulate the activity of the larger protein in which the helix resides, would be extremely versatile scaffolds. The ability to produce such molecules is challenging, even though the alpha helix is structurally quite simple. This is due to the fact that a functional alpha helix is often stabilized by other elements in the protein tertiary structure. With few exceptions, an isolated alpha helix will not maintain its alpha helical conformation taken out of the
**Figure 1.4** Small, nonhomologous protein scaffolds for displaying functional epitopes. The binding epitope of CD4 (orange and red) was transferred to a structurally homologous β-hairpin of the scyllatoxin scorpion toxin protein (red) (6). CD4 side chains transferred to scyllatoxin, appear as red sticks. Two rounds of structure-based design yielded the miniature protein CD4M33, which binds gp120 with affinity comparable to that of the parent CD4 protein (7). Shown at right is a model of the unoptimized miniature protein CD4M9 (yellow) bound to gp120 (green) based on the known NMR structure of CD4M9 and the known X-ray structure of gp120. Amino acids side chains critical in the interactions are in blue stick for CD4M9 and red stick for gp120. Ac is the miniprotein N-terminal acetyl group.
CD4

Scyllatoxin

Binding epitope

CD4M9

gp120

Asp368

Argβ

Phe23

Ac
context of its native protein (13). Unstructured peptides suffer from loss of function, likely due to the entropic penalty paid for folding into a suitable binding conformation.

**Stabilization of isolated alpha helices**

The alpha helix has been the focus of much interest, particularly the factors that allow formation of a stable, isolated helix. Many studies have provided insight into the naturally occurring interactions that promote helix stability (Figure 1.5).

**Alpha helical propensities**

For example, host-guest studies (14-17), statistical analysis (18,19), and mutagenesis (20) of residues found in alpha helices in proteins have been used to determine the helical propensity of each amino acid. For the most part, the rankings of helical propensity from these studies agree well in general with each other (21). Nevertheless, it is generally agreed upon that helix propensity is context-dependent. While helix propensity can be used as a general indicator, it is not absolutely predictive of alpha helicity.

**Macrodipole stabilization**

The alpha helix backbone is formed from hydrogen bonds between the carbonyl oxygen of residue $i$ and the amide proton of residue $i + 4$. These hydrogen bonds are oriented in the same direction, parallel to the helical axis, causing all the main chain carbonyl oxygens to point toward the C-terminus and all the main chain amide nitrogens to point toward the N-terminus. Consequently, the individual dipole moments of the
amide bonds are aligned to give a net dipole moment, or macrodipole, for the alpha helix. The N-terminus of the alpha helix carries a partial positive charge, and the C-terminus carries a partial negative charge (22). This charge separation is destabilizing; the macrodipole can be neutralized by incorporation of negatively charged residues near the N-terminus and positively charged residues near the C-terminus of the alpha helix. For instance, the movement of a positively charged histidine residue from near the N-terminus to the C-terminus of an alanine rich peptide increased alpha helix character from 24% to 76% (23). Furthermore, helicity was significantly decreased in 1 M NaCl, indicating that the electrostatic interactions between the histidine and the helix macrodipole was the stabilizing factor, and not hydrogen bonding between histidine and the helix backbone.

*Capping*

Another way to stabilize the macrodipole is by capping the alpha helix termini. Unlike the central amino acids of the helix, the four amide protons at the N-terminus do not have hydrogen bond acceptors; the four carbonyl oxygens at the C-terminus do not have hydrogen bond donors. This results in uncompensated partial charges on these atoms (23). N- and C-caps (composed of natural or nonnatural amino acids) can satisfy the hydrogen bonding of these atoms and, consequently, stabilize the alpha helix. Studies of model peptides (24) and statistical analysis (25,26) of residues found at N-capping positions in proteins show that amino acids such as Asn, Asp, Glu, Gln, Ser, and Thr stabilize the N-terminus. At C-capping positions, residues such as Asn, Arg, Gln, Lys, and Ser stabilize the C-terminus.
Figure 1.5 Methods of alpha helix stabilization. A. Alpha helix promoting residues. B. Helix macrodipole stabilization. C. Capping. D. Helix initiation. E. $i, i + 3$ and $i, i + 4$ interactions. From: http://nobelprize.org/chemistry/articles/malmstrom/bild1.html.
A. $R = \text{helix promoting residues}$

B. Negatively charged residue to stabilize macrodipole

C. N-capping residues Asn, Asp, Glu, Gln, Ser, and Thr

D. Helix initiating residues

E. Hydrophobic, π-stacking, cation-π, electrostatic and hydrogen bonding between $i, i + 3$ or $i, i + 4$ residues

C. C-capping residues Asn, Arg, Gln, Lys, and Ser

B. Positively charged residue to stabilize macrodipole
**Helix initiation**

The formation of conformationally rigid peptides is entropically disfavored compared to an unstructured peptide strand. Therefore, the ordering of the first four amino acids of an alpha helix, or helix initiation, is entropically disfavored. Amino acids that conformationally restrict the orientation of the first four amide bonds may be incorporated at the N-terminus to decrease this entropic penalty. For example, proline is often an N-cap residue or found immediately adjacent to N-cap residues in alpha helices (26). As the only N-substituted amino acid, proline is restricted in conformation, improving alpha helix nucleation.

*i, i + n interactions*

A number of interactions between residues close in proximity along the same face of the alpha helix are stabilizing. These interactions often occur between residues *i, i + 4* and *i, i + 3* positions, but they can also occur between *i* and *i + 7* positions. Examples of noncovalent interactions used to stabilize alpha helices include hydrophobic (27-31), π-stacking (32-35), cation-π (36-38), and electrostatic and hydrogen bonding interactions (39-45).

**Protein scaffolds for displaying alpha helical epitopes**

The strategies above describe the stabilization of isolated alpha helices, with little regard to function. In these studies, many of the model peptides are composed of alanines and stabilized by salt bridges. A large fraction of the residues in the peptides are required for stabilization, and substitution of these residues may disrupt the alpha helical
conformation. Therefore, these peptides are ill-suited for producing stable and functional molecules. One strategy for the design of functional mimics of isolated alpha helices is to use natural or nonnatural scaffolds to display functionality in a spatial arrangement similar to that of an alpha helical backbone displaying functionality.

For example, Hamilton and coworkers successfully used a 3,2′,2″-trisubstituted terphenyl scaffold to present the functionality of alpha helices (46-49). Also, a 14-helical β3-peptide mimicking the alpha helical p53 activation domain has been used to disrupt the p53•hDM2 interaction (50). The advantage of small molecule scaffolds over protein scaffolds is that proteins are less likely to get into cells than small organic molecules, with few exceptions (51,52). In addition, small molecules can present greater chemical diversity than protein scaffolds.

However, protein scaffolds possess other favorable characteristics. They can present a greater number of amino acid residues than a small molecule scaffold. Proteins can be genetically encoded, enabling identification and optimization of scaffold-based ligands through a wide variety of in vitro and in vivo functional selection technologies (53) such as mRNA display, phage display, and protein complementation. For these reasons, alpha helical proteins with stable tertiary folds may be more advantageous than organic molecules as molecular scaffolds for presenting alpha helical functional epitopes. The only prerequisite for such a protein scaffold is that it must contain an alpha helix amenable to multiple substitutions along one solvent accessible face that do not disrupt the overall stability of the scaffold.

A number of examples exist in which functional epitopes have been transferred between alpha helical proteins of structural homology (54). More intriguing, though, are
the infrequent examples where the functional epitopes have been transferred from the alpha helix of one protein to the alpha helix of a nonhomologous protein where the stabilization is due to significantly different tertiary contacts. For example, the alpha helical RRE recognition epitope from HIV-1 Rev protein was substituted on the alpha helix of the Zif268 zinc finger to generate an RNA binding zinc finger (Figure 1.6a) (55). The chimeric zinc finger bound RRE RNA with the same affinity as the isolated Rev alpha helix flanked by alanines ($K_d = 330$ nM) (56). The RRE binding zinc finger displayed a circular dichroism spectrum similar to that of a DNA-binding zinc finger, indicating that the helix presented by the zinc finger scaffold is preorganized in a conformation capable of binding RNA.

Small protein scaffolds have also been used to recognize protein targets. The E6 binding motif of the E6-associated protein (E6AP) of human papillomavirus was incorporated into an exposed helix of two small, stable scaffolds: the zinc finger of the SP1 protein and the Trp cage (Figure 1.6b) (57). E6apc2, based on the zinc finger scaffold, inhibited the interaction of E6 with the E6AP protein with an IC$_{50}$ of 19.3 µM; E6apn1, based on the Trp cage scaffold, inhibited the interaction with an IC$_{50}$ of 36.8 µM. While the designed E6 binding proteins were structurally similar to the scaffold proteins by NMR, inhibition was not improved compared to the unstructured parent E6ap18 peptide (IC$_{50}$ of 10.5 µM). In these cases, further optimization may be required.

Small, stable protein scaffolds can also be used to create DNA-binding proteins. For example, the DNA-binding epitope of the bZIP GCN4 was displayed on the small, stable villin headpiece domain (Figure 1.6c) (58). The chimeric bVIL protein was well-folded, with a $T_m$ of 66 °C, similar to that of the parent villin headpiece domain.
**Figure 1.6** Protein scaffolds for alpha helical epitope stabilization.  

A. Model of an RNA-binding miniature protein. The functional epitope of Rev (yellow) was transferred to the Zif268 zinc finger (red). (55).  

B. NMR structures of protein-binding miniature proteins. The E6-binding motif of E6-associated protein (E6AP), LQELLGE, was transferred to the third zinc finger of the SP1 protein and to the Trp cage scaffold.  

**LEFT:** The side chains of the E6apc2 peptide are labeled for residues important for forming a stable core (Cys5 and Cys8 are yellow; His21 and His25 are cyan; Phe12 is gray; Leu18 is pale green) or for binding to E6 (Leu24, 27, and 28 are magenta, red, and green, respectively).  

**RIGHT:** The side chains of the E6apn1 peptide are labeled for residues important for forming a stable core (Trp9 is dark orange; Pro15 is red; and Pro22 is violet) or for binding to E6 (Leu2 is magenta; Leu5 is red; and Leu6 is green) (57).  

C. Model of a DNA-binding miniature protein. The functional epitope of GCN4 (red) was transferred to the villin headpiece domain (green) (58).
A

Rev mimic based on zinc finger scaffold

B

E6AP mimic based on zinc finger scaffold

E6AP mimic based on trp cage scaffold

C

GCN4 mimic based on villin headpiece domain
However, bVIL bound DNA weakly and with little specificity. Inducing dimerization by adding the leucine zipper motif improved the DNA-binding properties, but negates the advantages of a miniature protein over full-length GCN4 bZIP dimer. In this case, the villin headpiece domain does not appear to be a useful alpha helical scaffold protein. Preorganization of a functional epitope is only beneficial when the stabilized epitope is presented in a binding-competent conformation. This is illustrated by a more successful approach to mimic the GCN4 basic region in our laboratory using avian pancreatic polypeptide (aPP) (59,60) as the scaffold.

**Protein grafting**

Our laboratory has pioneered the use of aPP as a scaffold for protein grafting, with much success in the generation of miniature proteins that bind macromolecular targets with high affinity and selectivity (61-68). Schematically, the residues on an alpha helical ligand required for recognition are substituted on the alpha helix of the small, stable protein scaffold aPP. Though it is only 36 amino acids long, aPP has a $T_m = 60 \degree C$. This stability is derived from the packing of an N-terminal type II polyproline helix (PPII) against a C-terminal alpha helix, burying 700 $\AA^2$ of surface area and leaving the solvent-exposed residues on the alpha helix available for substitution (Figure 1.7). The aPP-based protein grafting strategy was first applied to DNA recognition by grafting the alpha helical functional epitope of the bZIP protein GCN4 onto aPP to generate the rationally designed peptide PPBR4 (61). Though PPBR4 bound DNA with low nanomolar affinity at 4 $\degree$C, it failed to bind at 25 $\degree$C and was not folded in the absence of DNA. In an attempt to reestablish folding, the PPII helix of PPBR4 was evolved using a
Figure 1.7  Avian pancreatic polypeptide (59). Residues from the polyproline II helix which contribute to the hydrophobic core are in yellow and residues from the alpha helix which form part of the hydrophobic core are in blue.
Phage display functional selection resulting in the well-folded miniature protein, p007, that binds target DNA with nanomolar affinity and enhanced specificity (Figure 1.8) (62). However, optimization of grafted miniature proteins is not always necessary, as with the case of the miniature Q50K homeodomain mimic PPeng4, which was able to bind DNA even without the contacts provided by the parent protein’s N-terminal arm (65).

PPBR4 is an unstructured miniature protein because residues in the aPP hydrophobic core that stabilizes the native structure were replaced with residues from GCN4 to facilitate DNA binding (Dr. Brian Linton, unpublished results). In an ideal protein graft, the resulting miniature proteins are well-folded like aPP because the structural residues are retained, and highly functional because the binding residues are preorganized into an alpha helical conformation.

Protein grafting and functional optimization was first successfully applied to the problem of designing ligands for protein recognition in the context of the Bcl-2 protein family (64). The binding epitope of the Bak BH3 domain was grafted onto the aPP scaffold, and phage display selection generated the miniature protein PPBH3-1, which bound Bcl-2 and Bcl-X\textsubscript{L} with K\textsubscript{d}s of 52 nM and 7 nM, respectively. These binding affinities represent an improvement of three orders of magnitude compared to the Bak peptide. In addition, PPBH3-1 is as stable as aPP, with a T\textsubscript{m} of 65 °C. The increased binding affinity could be due to preorganization of the Bak functional epitope, which is unstructured in the absence of its binding partner.

The design of protein ligands that bind Bcl-2 and Bcl-X\textsubscript{L} is a relatively conservative test of the grafting strategy because the alpha helical regions of the ligands bind in deep, hydrophobic grooves. A more challenging protein grafting target is the
Figure 1.8  Protein grafting in the context of DNA recognition. The functional epitope of GCN4 was transferred to the solvent exposed face of the aPP alpha helix to yield the miniature protein PPBR4, which binds DNA specifically in the nanomolar concentration range at 4 °C, but does not bind DNA at room temperature (61), due to lack of a well defined structure (Dr. Brian Linton, unpublished results). The hydrophobic core of PPBR4 was repacked by phage display, yielding the miniature protein p007 which binds DNA specifically in the nanomolar concentration range at room temperature and is well folded (62).
protein grafting

introduce diversity

miniature protein library based on PPBR4

functional selection

dissect residues required for hsCRE DNA recognition

optimized miniature protein p007 (well folded, functional)

aPP scaffold (well folded, nonfunctional)
KIX domain of the transcriptional coactivator CPB. In contrast to the Bcl family proteins, the binding surface of CBP KIX is large and shallow. Furthermore, its ligand, the alpha helical KID domain of CREB, contains a phosphorylated serine that contributes a large portion of the binding energy. Protein grafting and phage display selections yielded phosphorylated PPKID peptides that bound the CBP KIX with high nanomolar to low micromolar affinity and also yielded PPKID peptides that bypassed the need for phosphorylation (66).

The aPP scaffold has shown its versatility by the successful grafting of functional epitopes onto the PPII helix rather than the alpha helix. The proline-rich functional residues of the ActA protein of Listeria monocytogenes was grafted onto the PPII helix of aPP to create a well-folded miniature protein that binds the Mena EVH1 domain with a $K_d$ of 700 nM (67). This affinity is 10-fold greater than that of a single functional repeat of the ActA protein, ActA$_{11}$. This miniature protein pGolemi, unlike ActA$_{11}$, can discriminate between related EVH1 domains.

Other laboratories have also had success using the aPP scaffold to create functional miniature proteins. Grafting the alpha helical epitope of the Kaposi’s sarcoma-associated herpesvirus protease (KSHV Pr) onto the aPP scaffold generated a miniature protein that disrupted formation of the obligate dimer (69). The reported $K_d$ of the KSHV Pr dimer is approximately 1 µM, and the grafted protein binds with an estimated $K_d$ of 300 µM. The PPII helix of aPP was also used as a scaffold to generate a miniature protein that could bind another proline rich motif. Residues from a proline rich peptide RP1 were grafted on the PPII helix of aPP to generate miniature proteins that could bind the Abl-SH3 domain (70). Though one, APP-RP1, was quite stable ($T_m =$
72.1 °C), most bound no better than the RP1 peptide (K_d = 30 µM).

**DNA-binding proteins**

One subset of macromolecular recognition that is of particular interest in our laboratory is that of DNA•protein interactions. DNA-binding proteins are critical in many aspects of biology, especially transcriptional regulation. The spatial and temporal regulation of transcription is crucial for the development and survival of all organisms. A key mechanism of transcriptional regulation is the action of activators or repressors on the basal transcription machinery (71). Transcriptional activators or repressors usually contain a DNA-binding domain and an activation or repression domain which are functionally separable (72,73). Therefore, the specificity of activation is simply controlled by the sequence specific binding of the DNA-binding domain (71,74). The DNA-binding domains of many transcription factors often use alpha helices to bind to the major groove of DNA (75). Notable exceptions include the prokaryotic repressors met (76), mnt (77), arc (78), the novel DNA-binding protein AbrB (79), and the eukaryotic TATA binding protein (80). Because the aPP scaffold has proven so useful and versatile, we are most interested in the design of DNA-binding proteins that use an alpha helix in the major groove to bind DNA. Such proteins could be used to regulate gene expression. The major classes of transcription factors that contact DNA in the major groove include the helix-turn-helix (HTH) proteins, zinc-containing proteins, and basic region leucine zipper (bZIP) proteins.

**Helix-turn-helix (HTH) proteins**
The CAP protein (81) from *E. coli* and the cro protein (82) from lambda phage were the first structures of sequence-specific DNA-binding proteins solved. Both proteins bind DNA using the helix-turn-helix (HTH) motif, the most common DNA-binding motif in prokaryotes (it is also found in eukaryotic homeodomain proteins and Myb) (Figure 1.9a). Several structures of homeodomain–DNA complexes reveal that a recognition alpha helix is bound in the major groove of DNA. Furthermore, homeodomain proteins can have additional DNA contacts mediated by an N-terminal arm in the adjacent minor groove (83).

**Zinc-containing proteins**

Three subclasses of eukaryotic transcription factors that use zinc to stabilize their tertiary structures are the zinc finger proteins (Cys$_2$His$_2$) (Figure 1.9b), the nuclear receptor proteins (Cys$_4$), and the binuclear zinc clusters. A representative protein from each of these classes has been solved by X-ray crystallography: Zif268 (84), murine glucocorticoid receptor (85), and yeast GAL4 (86), respectively. The crystal structures show that these disparate proteins share a common mode of binding DNA, placing a recognition helix in the major groove. The alpha helices in these proteins are stabilized by other secondary structure elements in which zinc ions are coordinated by cysteine and/or histidine residues on the alpha helix and other regions of the protein.

**Basic region leucine zippers (bZIPs) and related proteins**

The basic region leucine zipper protein is a third class of eukaryotic transcription factors (Figure 1.9c). These include the yeast GCN4, C/EBP, and the oncoproteins Fos
and Jun. Crystal structures of the GCN4 bZIP region bound to its target DNA site reveal that it binds DNA as an alpha helical homodimer, with the N-terminal basic region of each monomer contacting adjacent major grooves of DNA. Dimerization occurs via contacts within a C-terminal coiled coil (87,88). The bZIP element of bZIP proteins is solely responsible for DNA recognition and is localized to 60 contiguous amino acids. The bZIP element consists of three domains: an N-terminal basic region that binds DNA, a spacer region that influences relative orientation and half site spacing of the monomers, and a C-terminal leucine zipper that induces dimerization through formation of a coiled coil. The basic helix-loop-helix zipper protein (bHLHZIP) Max is closely related to bZIPS, but contains a helix-loop-helix dimerization domain, rather than the leucine zipper domain (89).

Although isolated basic regions from bZIPS can bind to DNA as monomers, their affinities are in the high micromolar range, compared to full-length bZIPS which bind in the nanomolar range (90-92). Intriguingly, the basic region-containing protein Skn-1 binds DNA as a monomer with a nanomolar dissociation constant, in contrast to other dimeric basic region proteins (93,94). This is possible because the basic region of Skn-1 is stabilized by other elements of tertiary structure. Furthermore, in addition to placing the basic region alpha helix in the major groove, Skn-1 uses an N-terminal arm to contact DNA in the minor groove (93,94).

**Designed DNA-binding proteins**

A number of attempts have been made to design DNA-binding proteins based on existing motifs. The first reports of engineering protein-DNA interactions were from the
Figure 1.9  The crystal structures of major classes of transcription factors.  
A. The HTH cro protein from bacteriophage 434 (80).  
B. The zinc finger Zif268 (82).  
C. The bZIP GCN4 (85).  

Ptashne lab who modified the recognition helix of the HTH binding domain of the bacteriophage 434 repressor to mimic the recognition helix of the Cro protein and the P22 repressor (95,96). While the HTH proteins were the first engineered DNA-binding proteins, more design strategies have been based on zinc finger proteins. In 1994, the laboratories of Pabo (97), Wells (98), and Klug (99) independently demonstrated that it is possible to generate phage libraries of zif268, a three-finger zinc finger that binds the nine bp site CGCACC, and select for mutants that bind to altered DNA sequences. In most examples, the selected zinc finger proteins had dissociation constants comparable to that of wild type zinc finger proteins ($K_d \sim 1$-10 nM). Their specificities were also as good as zif268 (97). However, efforts to sequence the human genome revealed that 16-18 bp of DNA would be required to specify a unique site in the genome. Strings of zinc fingers were designed to bind 18 bp of DNA, with some achieving femtomolar binding affinities (100-102). Nevertheless, several shortcomings of these proteins prevent their practical use in gene regulation. First, this library did not produce proteins that could specify all 18 bp of the target DNA sequence. At best, these zinc finger proteins can only specify 10 bp of the targeted 18 bp site (100). These proteins form stable DNA•protein complexes that have half-lives on the order of days, making it impossible for transcriptional regulation \textit{in vivo} to occur within the biological timescale (100). Furthermore, because of the high affinity and long lifetime of the proteins on nonspecific DNA sites, there is little chance of these proteins finding their specific DNA sites within a reasonable amount of time \textit{in vivo}.

Other DNA-binding proteins besides the HTH and zinc finger proteins have been used as scaffolds to create DNA-binding proteins with altered DNA-binding properties.
Mutants of the progesterone receptor that prefer to bind to the DNA target site of the estrogen receptor were generated using in vivo selection methods (103). Minimal bZIP constructs (104) and bZIPs with altered specificities (105-112) were designed rationally. In vivo selection methods also afforded variants of the bZIP protein C/EBP that bind target DNA sequences with two of five bases mutated in each half site (113).

**Proteins that bind nonnatural DNA**

To avoid the complications arising from identifying an 18 bp target site, proteins could be designed to bind a target site containing nonnatural DNA that is completely orthogonal to all natural DNA sequences. Furthermore, high affinity DNA-binding proteins that can recognize nonnatural DNA can be used to regulate transcriptional control by effectively decreasing the number of base pairs necessary to define a unique site in the genome. These proteins can also be used to study principles governing protein-DNA interactions. Using rational design and phage display selection, Shokat and colleagues engineered a mutant of the engrailed homeodomain, a developmental transcription factor, to bind a modified target site containing a nonnatural base (thymine with the C5 methyl substituted with a propynyl-oxazolidinone) (114). The wild type homeodomain bound the modified target site 60-fold worse than its own target site, and the engineered homeodomain bound 3-fold better to the modified target site over the natural target site.

In addition, there has been increasing interest in expansion of the genetic alphabet. Using an increasingly diverse number of amino acids coded by an expanded genetic alphabet could increase structural and functional diversity of proteins (115).
necessitates amber suppression, tRNAs that identify four bases, or nonnatural base pairs that can be replicated and transcribed for stable incorporation into a genome.

**Nonnatural base pairs**

Many nonnatural base pairs have been synthesized. The first nonnatural base pairs possessed structures based on natural base pairs, but contained alternative hydrogen bonding patterns. Since the limited possible arrangements of hydrogen bonding patterns between base pairs constrains the number of possible nucleosides, more recent efforts focus on synthesizing hydrophobic base pairs that are sterically complementary but cannot hydrogen bond. Originally, these nonnatural base pairs were also used to study the requirement for hydrogen bonding in replication. A third variety of base pairs, in addition to being hydrophobic, do not have the canonical purine or pyrimidine shape. Importantly, the presence of these base pairs does not radically alter the local structure within B-form DNA.

A classic example of a nonnatural base pair with a nonnatural hydrogen bonding pattern is a G:C mimetic synthesized in the Switzer group (Figure 1.10a) (116). Exchange of the carbonyl and amino groups of G and C results in iso-guanine (iso-G) and iso-cytosine (iso-C). The base pair formed by iso-G and iso-C is as stable as a G:C base pair in an oligonucleotide duplex (117). One disadvantage of the iso-G:iso-C base pair is that it is not orthogonal to natural base pairs; the hybrid base pair iso-G:C has stability comparable to an A:U base pair. Although the Klenow fragment, RNA polymerase, AMV reverse transcriptase, and HIV-1 reverse transcriptase can incorporate iso-G opposite iso-C in a template, iso-G can also adopt a tautomeric enol form and be
incorporated opposite T (118-120). Moreover, iso-C is unstable, making this base pair ill-suited for genetic storage (121). Still, this base pair remains a useful addition to the genetic alphabet. In fact, iso-G:iso-C has been used to create a 65th codon (iso-C)AG that is translated by the ribosome as the nonnatural amino acid iodotyrosine, which is charged to the corresponding tRNA containing the anticodon CU(iso-G) (122). This base pair has also been used to probe the role of triplex formation in RecA mediated strand exchange. The hydrogen bonding pattern of iso-G and iso-C bases would not allow the modeled triplexes to form, and this pair was used to confirm that strand exchange does not involve triple helix formation. Rather, the breathing and base flipping of double stranded DNA allows the invading strand to sample the DNA to find sites of homology (123).

An example of a well-studied hydrophobic base pair is the 4-methylbenzimidazole: difluorotoluene (Z:F) pair reported by Kool and colleagues (Figure 1.10b) (124). These bases are nonpolar isosteres of A and T, respectively. Polymerases can read through Z:F base pairs with low fidelity, as with iso-G and iso-C. In this case, A is inserted (as well as Z) opposite F in a DNA template. Presumably, this is because Z:F is about 3 kcal\(\cdot\)mol\(^{-1}\) less stable than an A:T base pair in a duplex, and only slightly more stable than A:F. If hydrogen bonding is the primary mechanism of base pairing, there appears to be no reason why A should pair with F, as F cannot hydrogen bond. This emphasizes the fact that hydrogen bonding alone does not dictate base pairing, and geometry plays an important role. These nonnatural base pairs have been essential as probes of protein\(\cdot\)DNA interactions, specifically the ability of various polymerases to recognize and extend DNA templates containing hydrophobic, non-hydrogen bonding
bases. In addition, the Kool laboratory has designed a nonpolar pyrene nucleotide that is efficiently incorporated opposite an abasic site (125). The pyrene nucleotide essentially occupies the space of an entire base pair. Fortuitously, it is also fluorescent, making it a useful tool for detecting DNA lesions in genetic diseases.

To investigate a broader range of nonpolar base pairs, Schultz and Romesberg have synthesized numerous hydrophobic nucleotides. Of special interest is a hydrophobic self-pair called 7-propynyl isocarbostyril nucleoside, or PICS (Figure 1.10c) (126). This nucleotide is efficiently inserted opposite itself in vitro by the Klenow fragment. Unlike iso-G:iso-C and Z:F, it discriminates quite well against the natural bases because its hydrophobicity and lack of a purine or pyrimidine shape prevent pairing with natural bases. The PICS:PICS pair is more stable than a G:C base pair in a duplex, and mispairs between PICS and a native base are as destabilizing as mispairs between any two native bases. Thus, the PICS:PICS pair is the first truly orthogonal base pair reported in the literature (127).

**Applications of miniature proteins that recognize nonnatural base pairs**

The PICS self pair can be used to gain insight into the evolution of genetic information storage. Why has nature chosen the two native base pairs? Could other base pairs have performed the necessary functions if nature had happened upon them? Examination of the successes or failures of nonnatural base pairs in recapitulating the functions of the natural base pairs could provide clues as to what properties of the natural base pairs led to their selection.

Studying the PICS:PICS base pair will also increase our knowledge about the
**Figure 1.10** Nonnatural base pairs.  

A. Iso-guanine:iso-cytosine (iso-G:iso-C) (114).

B. Nonpolar isosteres of adenine and thymine, 4-methylbenzimidazole:difluorotoluene (Z:F) (122).

C. 7-Propynlisocarbostyril nucleoside self pair, PICS:PICS (124).
noncovalent interactions that guide all macromolecular recognition events, specifically between DNA and proteins. The relative importance of polar and charged interactions vs. hydrophobic interactions in protein-DNA recognition can be investigated by replacing natural hydrogen bonding base pairs with hydrophobic base pairs. PICS-containing DNA would be a unique tool for learning about general principles governing molecular recognition, similar to the way in which the nucleotide analogues of nucleotide analogue interference mapping are used to study the importance of RNA chemical functionalities required for folding and activity (128). PICS is ideally suited for this purpose because it is truly orthogonal to natural base pairs both structurally and functionally. This gives rise to the possibility of discovering new DNA-protein interactions that do not yet exist and exploiting their potential applications.

Designing a protein that specifically recognizes PICS-containing DNA with high affinity will enhance our understanding of how to create macromolecules that bind specific targets and allow us to take advantage of the specificity of their interaction. For example, a high affinity DNA-binding protein tailored to recognize a promoter region containing the nonnatural base pair could be used to regulate transcriptional control. The binding site should not be recognized by other DNA-binding proteins due to the presence of the PICS pair. This on/off switch could be used to create conditional genetic knockouts. Of course, transcriptional control is not only useful for investigating cellular processes, but in gene therapy, as well. This unique DNA-protein interaction would be a stringent control mechanism for gene therapy. The gene cannot be replicated, transcribed, or expressed without an outside source of PICS and the PICS-binding proteins.
The design of the orthogonal PICS:PICS pair is the first step towards a synthetic biology. The stable incorporation of the PICS:PICS pair into DNA increases the information storage capacity of DNA. If it can be reliably transcribed into messenger RNA, there would be 125 possible codons instead of 64. The additional nonnatural codons could code for an increased variety of amino acids. If tRNAs could be designed and charged accordingly, any nonnatural amino acid could be specifically incorporated into a polypeptide chain. Thus, this system would be an additional method to synthesize novel proteins with new functionalities presented by nonnatural amino acids. These novel proteins would represent a breakthrough in biotechnology, and could be used as designer enzymes, drugs, bioindicators, or tools in the research lab.

Results

Experimental goals

One way to further enhance the specificity of protein-DNA interactions is by using nonnatural base pairs to dramatically decrease the number of base pairs necessary to define a unique site in the genome. Towards this end, we are interested in designing a miniature protein based on p007 that recognizes DNA containing a nonnatural base pair to produce a complex that is functionally orthogonal to protein•DNA complexes found in Nature. We introduced the nonnatural base pair PICS into the DNA sequence recognized by p007 and selected for p007 variants that bind this nonnatural sequence with high affinity and specificity.

Residues on the DNA-binding surface of p007 were randomized in our first
Figure 1.11 The design of Libraries 1 and 2. **TOP:** A model of the alpha helix of p007 targeting the modified hsCRE sequence ATGAP and a model of the alpha helix of p007 targeting the modified hsCRE sequence ATPAC. **BOTTOM:** Alignment of the sequences of the p007 alpha helix and Libraries 1 and 2 based on p007. Residues that form part of the hydrophobic core of aPP are shown blue. DNA binding residues of GCN4 are shown in red. Randomized residues are indicated by a green X. Adapted from reference (85).
Library 1

Library 2

p007  GGSRAPATMPGDDAPVEDLRKFRNTLAARRSRAARKAARAAA
Library 1  GGSRAPATMPGDDAPVEDLRKFRNXAXXRSXARXAARAAAA
Library 2  GGSRAPATMPGDDAPVEDLRKFRXTLXXRXXRXXAARAAAA
generation libraries (Figure 1.11). As much as possible, the residues required for packing of the hydrophobic core were maintained. Judiciously chosen residues on the DNA-binding surface of the alpha helix were be randomized, particularly in the region predicted to contact the PICS pair in the nonnatural hsCRE. Cloning will be performed using standard procedures. Phage display will be used to select for well-folded proteins that bind the unnatural hsCRE over the natural hsCRE. Ideally, a consensus sequence, or a few consensus sequences will be reached. Each of the consensus sequences will be chemically synthesized and HPLC purified. The relative $K_d$s of the complexes between PICS-containing hsCRE and the selected proteins will be determined by electrophoretic mobility shift assays. Side by side experiments using hsCRE itself as well as sequences containing PICS at different positions will be performed as controls.

**Library 1**

Guided by the GCN4•CRE crystal structures, for Library 1, we chose to substitute PICS at the fifth position of hsCRE, ATGAC$_{87,88}$. This position was chosen based on several considerations. The first three base pairs of the hsCRE sequence are recognized by residue Asn235, which is conserved across the bZIP family and retained in p007. Changing this asparagine to almost all other amino acids diminishes the ability of GCN4 to discriminate between this DNA sequence and other similar sites (107,129). We felt that substitution of any of the first three base pairs could disrupt the collective interactions that allow this specificity. The next A:T pair forms hydrophobic interactions with the conserved Ala238 in GCN4 (also retained in p007) (106). The hydrophobic PICS:PICS pair at this position may not differ enough from the natural hsCRE A:T pair.
for selective binding. Therefore, we chose to substitute the PICS:PICS pair for the C:G pair at the fifth position. We felt that substitution at this position was not likely to significantly destabilize global binding, while the functionality presented in the major groove by the PICS:PICS pair should be substantially different than that presented by C:G, and therefore, will allow for distinction between base pairs.

Library 1 was created to identify p007-based miniature proteins that could bind DNA containing a conservatively mutated hsCRE sequence, PICS5 (ATGAP) (Figure 1.12). The members of Library 1 differ from p007 at four positions (23, 26, 27, and 30) on the solvent-exposed face of p007’s alpha helix. Library 1 was constructed at the DNA level as described in the experimental section using the NNS codon scheme and contained 1.3 x 10^7 independent transformants. No parental clones were observed upon sequencing 26 clones from pool zero of this library, suggesting that it contains very little genetic background (130). However, 5 of 26 sequences clones contained single insertion and deletion mutants, so correcting for the number of transformants that are aberrant, the actual number of viable clones would be roughly 1.1 x 10^7. The adjusted number of transformants still exceeds the theoretical diversity of the library, which is 32^4 = 1.0 x 10^6. Statistically, we calculate that the library is approximately 100% complete (130).

**Library 2**

Library 2 was panned against a less conservative target. The second target was hsCRE substituted at the third position (ATGAC), completely disrupting the hsCRE and creating a new and unique recognition site for novel peptides that can bind the hydrophobic PICS pair.
**Figure 1.12** Plasmid map of the Library 1 phagemid vector, which codes for the p007 based miniature protein library as an N-terminal fusion to the M13 bacteriophage coat protein pIII. The DNA and amino acid (bold) sequences of the library-coding region are shown. The vector Library 1 was created by ligation of double-digested alpha helix library insert into the triple-digested phagemid vector pJC21-p007. Randomized residues are indicated by an X. N indicates an equimolar mixture of A, C, G, and T. S indicates an equimolar mixture of G and C.
Library 1
Figure 1.13  Plasmid map of the Library 2 phagemid vector, which codes for the p007 based miniature protein library as an N-terminal fusion to the M13 bacteriophage coat protein pIII. The DNA and amino acid (bold) sequences of the library-coding region are shown. The vector Library 2 was created by ligation of double-digested alpha helix library insert into the triple-digested phagemid vector pJC21-p007. Randomized residues are indicated by an X. N indicates an equimolar mixture of A, C, G, and T. S indicates an equimolar mixture of G and C.
A Q P A G G S R
GCCCAGCCGCGCAGGTGGGTCCCGG

A T M P G D D A
GCTACCATGCCAGGTGACGACGCA

P V E D L K R F
CCGTTTGAAAGATCTGAAGCGCTTTT

R X T L X X R R
CGTNNSACCCTGNNSSNNSGCACGT

X R A R X A A R
NNSSCGTGCACGTTNNSGCTGCACGT

A A A G G G G A
GCTGCAGCTGGTGTTGGTGCCCGG

Library 2
Library 2 was created to identify p007 based miniature proteins that could bind DNA containing a nonconservatively mutated hsCRE sequence, PICS3 (ATPAC) (Figure 1.13). The members of Library 2 differ from p007 at five positions (22, 25, 26, 29, and 33) on the solvent-exposed face of p007’s alpha helix. Library 2 was constructed at the DNA level as described in the experimental section using the NNS codon scheme and contained $9.8 \times 10^7$ independent transformants. No parental clones were observed upon sequencing 26 clones from pool zero of this library, suggesting that it contains very little genetic background (130). However, 5 of 26 sequenced clones contained deletion mutants, so correcting for the number of transformants that are aberrant, the actual number of viable clones would be roughly $7.9 \times 10^7$. The adjusted number of transformants still exceeds the theoretical diversity of the library, which is $32^5 = 3.3 \times 10^7$. Statistically, we calculate that the library is approximately 91% complete (130).

**Phage display selections with Library 1**

Eight rounds of selection were performed with Library 1. All but the first round included nonbiotinylated hsCRE$_{25}$ as a competitor. Phage were sorted on the basis of their ability to bind biotinylated PICS5 duplex attached to streptavidin-coated magnetic beads. Stringency was increased over the course of the selection by increasing the binding and washing temperature from 4 °C in round 1 to 25 °C by round 3, and by increasing the length and number of washes from 5 x 2 min washes in round 1 to 10 x 8 min washes in round 8 (Table 1.1).

The progress of the selection was monitored by measuring the retention of Library 1 phage in comparison to the retention of p007 phage (Figure 1.14). P007 phage was not
**Table 1.1** Binding and washing conditions used in Library 1 phage display selections.

<table>
<thead>
<tr>
<th>Round</th>
<th>Temperature (°C)</th>
<th>Number of washes</th>
<th>Wash length (min)</th>
<th>Competitor DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>5</td>
<td>2</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>5</td>
<td>2</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>5</td>
<td>8</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>5</td>
<td>8</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
<tr>
<td>7</td>
<td>25</td>
<td>10</td>
<td>8</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
<td>10</td>
<td>8</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

Phage binding buffer was 1X PBS (pH 7.4) supplemented with 1 mM EDTA, 0.1% NP-40, 0.4 mg/ml BSA, and 8 µg/mL poly(dI-dC). When competitor DNA was required in Library panning, 8 µg/mL hsCRE<sub>25</sub> was added instead of poly(dI-dC).
Figure 1.14  Panning of Library 1. In each round of selection, aPP phage and Library 1 phage were sorted based on their ability to bind immobilized PICS5 DNA. Binding and washing conditions are shown in Table 1.1. In each round, input and output titers were used to calculate percent retention, defined as (output titer / input titer) x 100, of both p007 and Library 1 phage. The progress of the selection was monitored by calculation of the fold retention of Library 1 phage over p007 phage in each round.
Round Fold retention over p007 phage
Table 1.2  Selected sequences from Library 1, chemically synthesized for detailed analysis.

<table>
<thead>
<tr>
<th>Sequence</th>
<th># of occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Round 4</td>
</tr>
<tr>
<td>APPS</td>
<td></td>
</tr>
<tr>
<td>APSA</td>
<td></td>
</tr>
<tr>
<td>DAVA</td>
<td></td>
</tr>
<tr>
<td>SPLR</td>
<td></td>
</tr>
<tr>
<td>SQAP</td>
<td></td>
</tr>
<tr>
<td>TAML</td>
<td></td>
</tr>
<tr>
<td>TAMR</td>
<td></td>
</tr>
</tbody>
</table>

Residues selected at randomized positions are in bold.
expected to bind PICS5, and thus served as a negative control. In round 1, relatively permissive conditions were used for binding and washing to allow retention of even weakly PICS5-binding phage. As the stringency of the selection increased, the fold retention of Library 1 phage over p007 phage was variable. The fold retention of Library 1 phage over p007 phage was 86 at round 7, but increasing numbers of mutations in the clones were being generated as detected by restriction digest screening and sequencing of plasmids. By the final round of selection, the phage from Library 1 were retained only 0.4-fold over p007 phage, and panning could not be continued due to the deterioration of the library.

An additional method to monitor the progress of the selection was sequencing of individual clones after rounds of selection. No consensus was observed, but one clone was selected for twice in round 5, TAMR. Six clones were chosen to be chemically synthesized based on the reoccurrence of amino acids at two or three positions and their appearance in later rounds of selection (Table 1.2).

**Phage display selections with Library 2**

Eight rounds of selection were performed with Library 2. All but the first two rounds included nonbiotinylated hsCRE25 as a competitor. Library quality deteriorated rapidly on the first attempt at panning beyond a third round, so the library was repanned from round 3. Phage were sorted on the basis of their ability to bind biotinylated PICS3 duplex attached to streptavidin-coated magnetic beads. Stringency was increased over the course of the selection by increasing the binding and washing temperature from 4 °C in round 1 to 25 °C by round 3, and by increasing the number of washes from 5 x 2 min
**Table 1.3** Binding and washing conditions used in Library 2 phage display selections.

<table>
<thead>
<tr>
<th>Round</th>
<th>Temperature (°C)</th>
<th>Number of washes</th>
<th>Wash length (min)</th>
<th>Competitor DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>5</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3'</td>
<td>5</td>
<td>10</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>5</td>
<td>2</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>5</td>
<td>8</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>5</td>
<td>8</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
<tr>
<td>7</td>
<td>25</td>
<td>10</td>
<td>8</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
<td>10</td>
<td>2</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

Phage binding buffer was 1X PBS (pH 7.4) supplemented with 1 mM EDTA, 0.1% NP-40, 0.4 mg/ml BSA, and 8 µg/mL poly(dI-dC). When competitor DNA was required in Library panning, 8 µg/mL hsCRE<sub>25</sub> was added instead of poly(dI-dC). Due to increasing loss in retention over negative control phage, Library 2 phage were repanned from round 3.
Figure 1.15  Panning of Library 2. In each round of selection, aPP phage and Library 2 phage were sorted based on their ability to bind immobilized PICS3 DNA. Binding and washing conditions are shown in Table 1.2. In each round, input and output titers were used to calculate percent retention, defined as (output titer / input titer) x 100, of both p007 and Library 2 phage. The progress of the selection was monitored by calculation of the fold retention of Library 2 phage over p007 phage in each round.
Table 1.4  Selected sequences from Library 2, chemically synthesized for detailed analysis.

<table>
<thead>
<tr>
<th>Sequence</th>
<th># of occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>APLMS</td>
<td>GGSRATMPGDDAPVEDLKRFRATLPPLRRMRARSA 1</td>
</tr>
<tr>
<td>HRPVQ</td>
<td>GGSRATMPGDDAPVEDLKRFRHTLRRPRVRARQAA 1</td>
</tr>
<tr>
<td>LRLPY</td>
<td>GGSRATMPGDDAPVEDLKRFRRLTRLRRPARYAA 2</td>
</tr>
<tr>
<td>NPHYT</td>
<td>GGSRATMPGDDAPVEDLKRFRNTLPHRRYRARTAA 1</td>
</tr>
<tr>
<td>PLWTP</td>
<td>GGSRATMPGDDAPVEDLKRFRPTLLWRRTRARPAA 1</td>
</tr>
<tr>
<td>PRPQQ</td>
<td>GGSRATMPGDDAPVEDLKRFRPTLRPRQRARQAA 1</td>
</tr>
<tr>
<td>RPNST</td>
<td>GGSRATMPGDDAPVEDLKRFRTRLPNRRSARATAA 2</td>
</tr>
<tr>
<td>SHQRA</td>
<td>GGSRATMPGDDAPVEDLKRFRSTLHQRRRRAA 2</td>
</tr>
<tr>
<td>SRTMS</td>
<td>GGSRATMPGDDAPVEDLKRFRSTLRTRMRARSA 1</td>
</tr>
<tr>
<td>STHMA</td>
<td>GGSRATMPGDDAPVEDLKRFRSTLTHRRMRARA 1</td>
</tr>
<tr>
<td>TAPRH</td>
<td>GGSRATMPGDDAPVEDLKRFRTLAPRRRARHAA 1</td>
</tr>
</tbody>
</table>

Residues selected at randomized positions are in bold.
washes in round 1 to 10 x 2 min washes in round 8’ (Table 1.3).

The progress of the selection was monitored by measuring the retention of Library 2 phage in comparison to the retention p007 phage (Figure 1.15). P007 phage is not expected to bind PICS3, and thus served as a negative control. In round 1, relatively permissive conditions were used for binding and washing to allow retention of even weakly PICS3-binding phage. As the stringency of the selection increased, the fold retention of phage from Library 2 over p007 phage was variable. The fold retention of Library 2 phage over p007 phage was 11 at round 8’, and panning was not continued due concerns about library quality.

An additional method to monitor the progress of the selection was sequencing of individual clones after rounds of selection. No consensus was observed, but four clones were selected for twice: in round 7, PRPQQ, and in round 7’, LRLPY, RPNST, and SHQRA. Seven clones were chosen to be chemically synthesized based on the reoccurrence of amino acids at three of five positions and their appearance in later rounds of selection (Table 1.4).

**EMSA screening of selected sequences**

Electrophoretic mobility shift assays were performed at 4 °C to provide preliminary characterization of the in vitro affinity of the selected peptides for PICS-containing DNA. The peptide was titrated in a seven point serial dilution, from 32 μM to 7.8 nM. The apparent dissociation constants were estimated at the concentration at which free and unbound DNA are equal in proportion. The results are shown in Table 1.5. From these initial studies, peptide APPS was not considered further due to poor binding
Table 1.5 Initial binding affinity analysis of selected peptides by EMSA at 4 °C.

<table>
<thead>
<tr>
<th>Library 1</th>
<th>Sequence</th>
<th>K_d (µM)</th>
<th>PICS5</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPS</td>
<td>GGSRATMPGDDAPVEDLKRFRNALAPRSMASARKAA</td>
<td>&gt;35</td>
<td></td>
</tr>
<tr>
<td>APSA</td>
<td>GGSRATMPGDDAPVEDLKRFRNALAPSRASARKAA</td>
<td>&gt;180</td>
<td></td>
</tr>
<tr>
<td>DAVA</td>
<td>GGSRATMPGDDAPVEDLKRFRNDLAAVRSARKAA</td>
<td>&gt;48</td>
<td></td>
</tr>
<tr>
<td>SPLR</td>
<td>GGSRATMPGDDAPVEDLKRFRNPSLAPRSARKAA</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>SQAP</td>
<td>GGSRATMPGDDAPVEDLKRFRNSLAQARSPARKAA</td>
<td>&gt;94</td>
<td></td>
</tr>
<tr>
<td>TAML</td>
<td>GGSRATMPGDDAPVEDLKRFRNTLAMRSLARKAA</td>
<td>&gt;16</td>
<td></td>
</tr>
<tr>
<td>TAMR</td>
<td>GGSRATMPGDDAPVEDLKRFRNTLAMRSRARKAA</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>Library 2</td>
<td>Sequence</td>
<td>PICS3</td>
<td></td>
</tr>
<tr>
<td>APLMS</td>
<td>GGSRATMPGDDAPVEDLKRFRATLPLRMARSA</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>HRPVQ</td>
<td>GGSRATMPGDDAPVEDLKRFRHTLPVRVRARQAA</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>LRLPY</td>
<td>GGSRATMPGDDAPVEDLKRFRRLRLRRPRARYAA</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>NPHYT</td>
<td>GGSRATMPGDDAPVEDLKRFRNTLPhRRYRARTAA</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>PLWTP</td>
<td>GGSRATMPGDDAPVEDLKRFRPLLWRTRARPAA</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>PRPQQ</td>
<td>GGSRATMPGDDAPVEDLKRFRPTLPRRQRARQAA</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>RPNST</td>
<td>GGSRATMPGDDAPVEDLKRFRTRLPNRSSRATTA</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>SHQRA</td>
<td>GGSRATMPGDDAPVEDLKRFRSTLHQRRARAAN</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>SRTMS</td>
<td>GGSRATMPGDDAPVEDLKRFRSTLRTRMRASAA</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>STHMA</td>
<td>GGSRATMPGDDAPVEDLKRFRSTLTHRRMRARAA</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>TAPRH</td>
<td>GGSRATMPGDDAPVEDLKRFRTTLAPRRRARHAA</td>
<td>3.9</td>
<td></td>
</tr>
</tbody>
</table>
affinity and purification difficulties. Peptides APSA, DAVA, SQAP, and TAML displayed high micromolar affinities and were not considered further. Two peptides from Library 1 (SPLR and TAMR) and 11 peptides from Library 2 (APLMS, HRPVQ, LRLPY, NPHYT, PLWTP, PRPQQ, RPNST, SHQRA, SRTMS, STHMA, and TAPRH) were selected for a specificity screen at 4 °C. Each peptide was tested with PICS5 and PICS3 DNA, and it was expected that those derived from Library 1 would bind PICS5, and those derived from Library 2 would bind PICS3. The other PICS-containing DNA would serve as a negative control. Three other DNA sequences served as additional negative controls: hsCRE25, containing the natural CRE half site; NON24, containing a scrambled CRE site; and QREP, a nonnatural negative control, which contained the PICS base pair in a different sequence context. Electrophoretic mobility shift experiments were performed to assess the affinities of peptides for the five different DNA targets. LRLPY had low affinity for all targets at 4 °C, and was omitted from the specificity screen at 25 °C (Table 1.6).

*Kₐ determination for selected peptides at 25 °C*

From the initial specificity screenings, one peptide from Library 1 (TAMR) and four peptides from Library 2 (HRPVQ, RPNST, SRTMS, and TAPRH) were selected for more accurate *Kₐ* determinations at 25 °C based on their high affinity for target DNA and apparent preference against distantly related DNA sequences. While the peptides bound with nanomolar affinity to DNA, each peptide bound with approximately equal affinity for all DNA targets tested (Table 1.7).
<table>
<thead>
<tr>
<th>Library 1</th>
<th>PICS5</th>
<th>hsCRE&lt;sub&gt;25&lt;/sub&gt;</th>
<th>PICS3</th>
<th>QREP</th>
<th>NON&lt;sub&gt;24&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPLR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>1790</td>
<td>263</td>
<td>3250</td>
<td>1560</td>
<td>1910</td>
</tr>
<tr>
<td>25 °C</td>
<td>&gt;12,800</td>
<td>&gt;12,800</td>
<td>&gt;12,800</td>
<td>&gt;12,800</td>
<td>&gt;12,800</td>
</tr>
<tr>
<td>TAMR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>222</td>
<td>131</td>
<td>265</td>
<td>522</td>
<td>232</td>
</tr>
<tr>
<td>25 °C</td>
<td>283</td>
<td>252</td>
<td>&gt;12,800</td>
<td>&gt;12,800</td>
<td>&gt;12,800</td>
</tr>
<tr>
<td>Library 2</td>
<td>PICS3</td>
<td>hsCRE&lt;sub&gt;25&lt;/sub&gt;</td>
<td>PICS5</td>
<td>QREP</td>
<td>NON&lt;sub&gt;24&lt;/sub&gt;</td>
</tr>
<tr>
<td>APLMS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>116</td>
<td>198</td>
<td>46.5</td>
<td>179</td>
<td>191</td>
</tr>
<tr>
<td>25 °C</td>
<td>182</td>
<td>113</td>
<td>112</td>
<td>418</td>
<td>65.4</td>
</tr>
<tr>
<td>HRPVQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>15.0</td>
<td>23.5</td>
<td>19.2</td>
<td>257</td>
<td>76.5</td>
</tr>
<tr>
<td>25 °C</td>
<td>20.9</td>
<td>14.3</td>
<td>18.3</td>
<td>55.6</td>
<td>56.3</td>
</tr>
<tr>
<td>LRLPY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>2480</td>
<td>2120</td>
<td>2660</td>
<td>297</td>
<td>3150</td>
</tr>
<tr>
<td>NPHYT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>1120</td>
<td>737</td>
<td>577</td>
<td>861</td>
<td>557</td>
</tr>
<tr>
<td>25 °C</td>
<td>974</td>
<td>818</td>
<td>927</td>
<td>1360</td>
<td>373</td>
</tr>
<tr>
<td>PLWTP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>246</td>
<td>377</td>
<td>334</td>
<td>1310</td>
<td>1020</td>
</tr>
<tr>
<td>25 °C</td>
<td>1800</td>
<td>1230</td>
<td>787</td>
<td>2130</td>
<td>904</td>
</tr>
<tr>
<td>PRPQQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>121</td>
<td>124</td>
<td>107</td>
<td>159</td>
<td>163</td>
</tr>
<tr>
<td>25 °C</td>
<td>66.1</td>
<td>85.2</td>
<td>181</td>
<td>137</td>
<td>94.2</td>
</tr>
<tr>
<td>RPNST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>91.1</td>
<td>85.0</td>
<td>92.2</td>
<td>129</td>
<td>139</td>
</tr>
<tr>
<td>25 °C</td>
<td>55.9</td>
<td>105</td>
<td>84.3</td>
<td>303</td>
<td>191</td>
</tr>
<tr>
<td>SHQRA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>179</td>
<td>163</td>
<td>170</td>
<td>238</td>
<td>160</td>
</tr>
<tr>
<td>25 °C</td>
<td>142</td>
<td>137</td>
<td>257</td>
<td>329</td>
<td>177</td>
</tr>
<tr>
<td>SRTMS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>23.9</td>
<td>23.0</td>
<td>28.0</td>
<td>55.8</td>
<td>49.6</td>
</tr>
<tr>
<td>25 °C</td>
<td>17.1</td>
<td>17.5</td>
<td>208</td>
<td>106</td>
<td>148</td>
</tr>
<tr>
<td>STHMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>332</td>
<td>443</td>
<td>466</td>
<td>1250</td>
<td>355</td>
</tr>
<tr>
<td>25 °C</td>
<td>1470</td>
<td>1550</td>
<td>1340</td>
<td>1620</td>
<td>1090</td>
</tr>
<tr>
<td>TAPRH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td>190</td>
<td>193</td>
<td>164</td>
<td>241</td>
<td>311</td>
</tr>
<tr>
<td>25 °C</td>
<td>895</td>
<td>1020</td>
<td>763</td>
<td>3090</td>
<td>1410</td>
</tr>
</tbody>
</table>
Table 1.7  Binding affinity of selected peptides for DNA targets at 25 °C.

<table>
<thead>
<tr>
<th>Library</th>
<th>PICS5</th>
<th>hsCRE_{25}</th>
<th>PICS3</th>
<th>QREP</th>
<th>NON_{24}</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAMR</td>
<td>767 ± 303</td>
<td>361 ± 153</td>
<td>657 ± 344</td>
<td>491 ± 416</td>
<td>311 ± 125</td>
</tr>
<tr>
<td>Library 2</td>
<td>PICS3</td>
<td>hsCRE_{25}</td>
<td>PICS5</td>
<td>QREP</td>
<td>NON_{24}</td>
</tr>
<tr>
<td>HRPVQ</td>
<td>127 ± 33</td>
<td>184 ± 45</td>
<td>135 ± 36</td>
<td>123 ± 37</td>
<td>127 ± 31</td>
</tr>
<tr>
<td>RPNST</td>
<td>76.9 ± 24.9</td>
<td>41.9 ± 3.6</td>
<td>27.2 ± 5.6</td>
<td>74.8 ± 8.9</td>
<td>35.2 ± 6.9</td>
</tr>
<tr>
<td>SRTMS</td>
<td>22.5 ± 3.0</td>
<td>24.2 ± 4.3</td>
<td>44.4 ± 10.8</td>
<td>37.5 ± 9.8</td>
<td>31.6 ± 2.2</td>
</tr>
<tr>
<td>TAPRH</td>
<td>253 ± 150</td>
<td>178 ± 93</td>
<td>61.6 ± 20.0</td>
<td>199 ± 70</td>
<td>305 ± 105</td>
</tr>
</tbody>
</table>
Discussion

**Sequences of selected peptides**

We would predict that amino acids with hydrophobic side chains would be selected to complement the hydrophobic surface of the PICS:PICS pair. However, there was no clear consensus observed. There was some preference for small, polar amino acids. The basic amino acid arginine was also observed, which may be expected due to the negatively charged phosphate backbone of the DNA.

**Affinity of selected peptides**

Of the seven synthesized peptides from Library 1, five had high micromolar dissociation constants and were not considered further. After more detailed analysis of the remaining two peptides, only one peptide from Library 1, TAMR, bound DNA well, with 222 nM $K_d$ at 4 °C. Interestingly, TAMR differs from p007 by only one residue, Met27, which is arginine in p007. However, preliminary studies showed that p007 bound PICS5 DNA with > 2.6 µM $K_d$. This suggests that what was considered a conservative mutation of hsCRE from ATGAC to ATGAP was perhaps more deleterious than expected, since only one p007 based peptide in Library 1 could bind this sequence well.

In contrast to the Library 1 peptides synthesized, all 11 of the peptides synthesized from Library 2 displayed low micromolar dissociation constants at 4 °C. P007 also bound PICS3 with > 2.6 µM $K_d$. Though we designed this PICS-containing hsCRE sequence to require more new contacts than PICS5 DNA, Library 2 contained more high affinity binders than Library 1. However, it is possible that peptides from both
libraries can bind some other site along the DNA without contacting the PICS:PICS self pair.

**Specificity of selected peptides**

The selected peptides were analyzed for specificity for their putative PICS-containing binding site. For each peptide, five DNA sequences were analyzed for binding affinity: PICS5, the target sequence of Library 1 containing PICS at position 5 of the hsCRE (ATGAP); hsCRE25, the DNA containing the hsCRE (ATGAC); PICS3, the target sequence of Library 2 containing PICS at position 3 of the hsCRE (ATPAC); QREP, an unrelated DNA sequence containing PICS; and NON24, a nonspecific DNA sequence (Table 1.8). In the plots, the target PICS-containing DNA is green for each Library.

TAMR was analyzed in detail for binding specificity. It bound PICS5 DNA with a $K_d$ of 767 ± 303. However, it bound all other DNA sequences with approximately the same affinity (Figure 1.16). HRPVQ, RPNST, SRTMS, and TAPRH were studied similarly. The $K_d$s of these peptides for the target PICS3 sequence ranged from 22.5 to 253 nM. However, the affinity of each peptide for the target sequence was approximately equal to the affinity for all the other sequences tested (Figure 1.17-20). These peptides likely bind in a nonspecific manner, albeit with low nanomolar affinity.

The selected peptides bound DNA with nanomolar affinity but no specificity for DNA containing a nonnatural base pair. There are several possibilities why the selections were not successful. In our laboratory and others, most successful design and selection strategies were guided by structural and biochemical information of the protein to be
Table 1.8  DNA sequences used as targets in EMSA experiments.

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PICS5</td>
<td>AGTGGGA\textbf{GATGAPGAGCTACTCGTGC–biotin}</td>
</tr>
<tr>
<td>PICS3</td>
<td>AGTGGGA\textbf{ATPACGAGCTACTCGTGC–biotin}</td>
</tr>
<tr>
<td>hsCRE\textsubscript{25}</td>
<td>AGTGGGA\textbf{ATGACGAGCTACTCGTGC}</td>
</tr>
<tr>
<td>QREP</td>
<td>CGCA\textbf{GTAAPCCCTCGAC–biotin}</td>
</tr>
<tr>
<td>NON\textsubscript{24}</td>
<td>AGTGGGA\textbf{TAAAGCCCTATCTCGTGC}</td>
</tr>
</tbody>
</table>
**Figure 1.16** The affinity of TAMR for DNA at 25 °C measured by electrophoretic mobility shift assay. The fraction of radiolabelled DNA bound was monitored as a function of peptide concentration after incubation of the binding reaction for 1 hr in 1X PBS (pH 7.4) supplemented with 1 mM EDTA, 0.1% NP-40, 0.4 mg/ml BSA, and 5% glycerol. Each point represents the average of three independent experiments, and the error bars denote the standard error. The curves shown are the best fit of the fraction of DNA bound values to the Langmuir equation (1.1).
<table>
<thead>
<tr>
<th>DNA</th>
<th>$K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PICS5</td>
<td>$767 \pm 303$</td>
</tr>
<tr>
<td>hsCRE$_{25}$</td>
<td>$361 \pm 153$</td>
</tr>
<tr>
<td>PICS3</td>
<td>$657 \pm 344$</td>
</tr>
<tr>
<td>QREP</td>
<td>$491 \pm 416$</td>
</tr>
<tr>
<td>NON$_{24}$</td>
<td>$311 \pm 125$</td>
</tr>
</tbody>
</table>
Figure 1.17  The affinity of HRPVQ for DNA at 25 °C measured by electrophoretic mobility shift assay. The fraction of radiolabelled DNA bound was monitored as a function of peptide concentration after incubation of the binding reaction for 1 hr in 1X PBS (pH 7.4) supplemented with 1 mM EDTA, 0.1% NP-40, 0.4 mg/ml BSA, and 5% glycerol. Each point represents the average of three independent experiments, and the error bars denote the standard error. The curves shown are the best fit of the fraction of DNA bound values to the Langmuir equation (1.1).
DNA | $K_d$ (nM)  
---|---  
PICS3 | 127 ± 33  
hsCRE$_{25}$ | 184 ± 45  
PICS5 | 135 ± 36  
QREP | 123 ± 37  
NON$_{24}$ | 127 ± 31
**Figure 1.18** The affinity of RPNST for DNA at 25 °C measured by electrophoretic mobility shift assay. The fraction of radiolabelled DNA bound was monitored as a function of peptide concentration after incubation of the binding reaction for 1 hr in 1X PBS (pH 7.4) supplemented with 1 mM EDTA, 0.1% NP-40, 0.4 mg/ml BSA, and 5% glycerol. Each point represents the average of three independent experiments, and the error bars denote the standard error. The curves shown are the best fit of the fraction of DNA bound values to the Langmuir equation (1.1).
<table>
<thead>
<tr>
<th>DNA</th>
<th>$K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PICS3</td>
<td>76.9 ± 24.9</td>
</tr>
<tr>
<td>hsCRE$_{25}$</td>
<td>41.9 ± 3.6</td>
</tr>
<tr>
<td>PICS5</td>
<td>27.2 ± 5.6</td>
</tr>
<tr>
<td>QREP</td>
<td>74.8 ± 8.9</td>
</tr>
<tr>
<td>NON$_{24}$</td>
<td>35.2 ± 6.9</td>
</tr>
</tbody>
</table>
Figure 1.19  The affinity of SRTMS for DNA at 25 °C measured by electrophoretic mobility shift assay. The fraction of radiolabelled DNA bound was monitored as a function of peptide concentration after incubation of the binding reaction for 1 hr in 1X PBS (pH 7.4) supplemented with 1 mM EDTA, 0.1% NP-40, 0.4 mg/ml BSA, and 5% glycerol. Each point represents the average of three independent experiments, and the error bars denote the standard error. The curves shown are the best fit of the fraction of DNA bound values to the Langmuir equation (1.1).
<table>
<thead>
<tr>
<th>DNA</th>
<th>$K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PICS3</td>
<td>22.5 ± 3.0</td>
</tr>
<tr>
<td>hsCRE$_{25}$</td>
<td>24.2 ± 4.3</td>
</tr>
<tr>
<td>PICS5</td>
<td>44.4 ± 10.8</td>
</tr>
<tr>
<td>QREP</td>
<td>37.5 ± 9.8</td>
</tr>
<tr>
<td>NON$_{24}$</td>
<td>31.6 ± 2.2</td>
</tr>
</tbody>
</table>
Figure 1.20  The affinity of TAPRH for DNA at 25 °C measured by electrophoretic mobility shift assay. The fraction of radiolabelled DNA bound was monitored as a function of peptide concentration after incubation of the binding reaction for 1 hr in 1X PBS (pH 7.4) supplemented with 1 mM EDTA, 0.1% NP-40, 0.4 mg/ml BSA, and 5% glycerol. Each point represents the average of three independent experiments, and the error bars denote the standard error. The curves shown are the best fit of the fraction of DNA bound values to the Langmuir equation (1.1).
<table>
<thead>
<tr>
<th>DNA</th>
<th>$K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PICS3</td>
<td>253 ± 150</td>
</tr>
<tr>
<td>hsCRE$_{25}$</td>
<td>178 ± 93</td>
</tr>
<tr>
<td>PICS5</td>
<td>61.6 ± 20.0</td>
</tr>
<tr>
<td>QREP</td>
<td>199 ± 70</td>
</tr>
<tr>
<td>NON$_{24}$</td>
<td>305 ± 105</td>
</tr>
</tbody>
</table>

Fraction DNA bound vs. [TAPRH] (M)
**Figure 1.21** Molecular model of PICS:PICS self pair in 12-mer duplex DNA: 5’-GAGAPGAAACG and 5’-CGTTTCPTTCTC. Picture from reference (129).
grafted onto aPP and its macromolecular target. At the time, there was no structure of the PICS:PICS base pair in DNA. NMR evidence showed that the B-form helix was not significantly distorted. Subsequent reports suggest that the PICS self-pair is stacked (Figure 1.21) (131). This presents a much different major groove than in our initial model. Perhaps the libraries were not optimally designed. Thus, there may be no solution to the problem from the limited members of our strategically randomized libraries. Nevertheless, it is also possible that the selections were not stringent enough to exclude all nonspecific DNA-binding peptides.

If this study was to be revisited, new design elements could enhance the possibility of success. New libraries could be designed based on the current model of the PICS:PICS pair in DNA. Alternatively, p007 does not have to be the scaffold, as the hsCRE is considerably altered such that there is no longer a bias toward binding the hsCRE DNA. A more stable scaffold, perhaps aPP, could be used instead. In addition, all five peptides studied in-depth have an arginine and many polar residues. High salt panning conditions may promote selection of residues that can form hydrophobic interactions with the PICS:PICS pair.

Experimental

General

Materials

All restriction enzymes and restriction enzyme buffers, dNTPs, DNA modifying enzymes (Vent® (exo-) DNA polymerase, T4 DNA ligase) and associated buffers, and
BSA were purchased from New England Biolabs (Beverly, MA). Sequenase Version 2.0 DNA polymerase and buffer were purchased from USB Corporation (Cleveland, OH). Calf thymus DNA was purchased from Gibco BRL at 10 mg/mL. Poly(dI-dC) was purchased from Amersham Pharmacia Biotech. Nonfat milk was purchased as a powder from Carnation. NAP-10 columns were purchased from Pharmacia Biotech (Piscataway, NJ). Chromaspin-1000 columns were purchased from Clontech (Palo Alto, CA). Sequel NE Concentrate (Part A) and Diluent (Part B) for preparation of denaturing polyacrylamide gels were purchased from American Bioanalytical. The QIAGEN Plasmid Maxi Kit, QIAprep Spin Miniprep Kit, QIAquick Nucleotide Removal Kit, and QIAquick Gel Extraction Kit were purchased from Qiagen, Inc (Valencia, CA). Electrocompetent XL1 Blue E. coli cells were purchased from Stratagene (La Jolla, CA).

**Media and buffers**

2X YT media was prepared by addition of 17 g Bacto-tryptone, 10 g Bacto-yeast extract and 5 g NaCl to 1 L ddH₂O. Following autoclaving, the solution was cooled to less than 50 °C and antibiotics and/or glucose were added as needed. 2X YT agar was prepared as 2X YT media, with 15 g Bacto-agar added before autoclaving. LB media was prepared by addition of 10 g Bacto-tryptone, 5 g Bacto-yeast extract, 10 g NaCl and 1 mL 1 N NaOH to 1 L ddH₂O. Following autoclaving, the solution was cooled to less than 50 °C and antibiotics and/or glucose were added as needed. SOB agar was prepared by addition of 20 g Bacto-tryptone, 5 g Bacto-yeast extract, 15 g Bacto-agar and 0.5 g NaCl to 1L ddH₂O. Following autoclaving, the solution was cooled to less than 50 °C, and antibiotics, 5 mL 2 M MgCl₂, and glucose were added as needed. “A” indicates that
ampicillin was added to a final concentration of 100 µg/mL. “K” indicates that kanamycin was added to a final concentration of 30 µg/mL. “G” indicates that glucose was added to a final concentration of 2%.

2X formamide loading buffer contained 80% deionized formamide, 10 mM EDTA, 0.001% xylene cyanole, 0.001% bromophenol blue. 1X TBE contained 89 mM Tris, 89 mM borate, 2 mM EDTA, pH 8.0. 1X TE contained 10 mM Tris, 1 mM EDTA, pH 8.0. Preparative and analytical agarose gels were prepared in 1X TAE. 1X TAE contained 40 mM Tris, 20 mM acetate, 1 mM EDTA, pH 8.0. 1X PBS contained 1.4 mM KH₂PO₄, 4.3 mM Na₂HPO₄, pH 7.5, 137 mM NaCl, 2.7 mM KCl. Phage binding buffer contained 1X PBS supplemented with 1 mM EDTA, 0.1% NP-40, 0.4 mg/ml BSA, and 8 µg/mL poly(dI-dC). When competitor DNA was required in Library panning, 8 µg/mL hsCRE₂₅ was added instead of poly(dI-dC). Phage wash buffer was identical to binding buffer except that it lacked poly(dI-dC). Phage elution buffer was identical to phage wash buffer except that it contained 4 M NaCl. 2X B+W buffer contained 10 mM Tris (pH 7.5), 1 mM EDTA, 2.0 M NaCl in ddH₂O. PEG/NaCl contained 20% (w/v) PEG-8000, 2.5 M NaCl.

Biomolecule synthesis and characterization

Peptide synthesis, oligonucleotide synthesis, amino acid analysis and automated DNA sequencing were performed at the W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University School of Medicine (New Haven, CT). Mass spectra were acquired on an Applied Biosystems (Foster City, CA) Voyager-DE-Pro MALDI-TOF mass spectrometer.
**DNA for panning**

For phage panning, PICS5A had the sequence 5’-ATGGAGATGAPGAGCTACTCGTGC-biotin, and PICS3A had the sequence 5’-ATGGAGATGAPGAGCTACTCGTGC-biotin, where P = PICS and biotin was attached to the 3’ end of the DNA through a TEG linker (Glen Research). PICS was synthesized by Dr. Elke Pierron and Dr. Scott Hart, and 1 mL of a 100 mM solution in acetonitrile was used to synthesize four 1 µmol scale oligonucleotides at the W. M. Keck Foundation Biotechnology Resource Laboratory. Duplex DNA was produced by annealing the “A” strands with their non-biotinylated complement. The sequencing primer S1 had the sequence 5’-CAACGTGAAAAAATTATTATTATTCGC-3’. DNA sequences used as targets in EMSA experiments are listed in Table 1.8.

**Library construction**

**Insert construction**

The oligonucleotides Library 1, Library 2, and Primer were ordered on a 0.2 µmol scale from the W. M. Keck Foundation Biotechnology Resource Laboratory. Long library oligonucleotides were most efficiently synthesized on 1000Å 0.2 µmol LV polystyrene columns from Applied Biosystems. Library 1, Library 2, and Primer had the following sequences: Library 1: 5’-CCACCAGCTGCAGCACGTGCAGCTTTACGTGCSNNGGAACGSNNSNNAGCCAGSNNGTTACGAAACGGCTTCAGATCTTCAACCGG-3’; Library 2: 5’-CCACCAGCTGCAGCACGTGCAGCSNNACGTGCACGSNNNNSNNACGTGCACGSNNNNSNNACGGGTSSNNACGAAAGCGCTTCAGATCTTCAACCGG-3’; Primer: 5’-
CCGGTTGAAGATCTGAAGC-3’. Crude Library 1 and Library 2 were resuspended in 100 µL ddH₂O, followed by addition of 100 µL 2X loading dye. The solutions were heated to 90 °C for 2 min and cooled on dry ice before loading onto a 13.3% (19:1 acrylamide:bisacrylamide) denaturing polyacrylamide gel pre-electrophoresed for 30 min at 65 W in 1X TBE. After sample loading, the gel was electrophoresed in 1X TBE for 5.5 h at 65 W. The appropriate 86 nucleotide bands were identified by UV shadowing and excised. The excised gel slices were crushed through a 10 mL disposable syringe, and the oligonucleotides were each eluted into 15 mL 1X TE overnight. Library 1, Library 2, and Primer (with gel purification) were desalted by passage through a NAP-10 column equilibrated in ddH₂O, and elution with 1.5 mL ddH₂O. The oligonucleotides were lyophilized, resuspended in 200 µL ddH₂O, and ethanol precipitated. The oligonucleotides were then resuspended in 1 mL ddH₂O.

Double-stranded Library 1 and Library 2 inserts were generated by primer extension of Primer. Equimolar amounts (400 pmol scale) of Library 1 or Library 2 single-stranded DNA and Primer were annealed in 1X Sequenase buffer (USB Corporation) in a total volume of 200 µL by heating to 95 °C for 10 min and subsequent slow cooling to room temperature. Primer extension was initiated by addition of the following components to each reaction: 2 µL 25 mM dNTPs, 2 µL 10 µg/µL BSA, 2 µL 100 mM DTT, and 52 U Sequenase DNA polymerase (USB Corporation). The extension reactions were incubated at 37 °C for 30 min, followed by 1 h incubation at 65 °C to heat inactivate the Sequenase enzyme.

The double-stranded Library 1 and Library 2 inserts were double-digested with BglII and PstI in the following reaction: 2.5 µg Library 1 or Library 2 insert, 30 µL 10X
NEBuffer 3, 30 U BglII, 60 U PstI in a total volume of 300 µL. The reaction was incubated for 16 h at 37 °C. The double-digested Library 1 and Library 2 inserts were separated from uncut or partial digestion products on a 15% native polyacrylamide gel. DNA was visualized by ethidium bromide staining. The appropriate bands were excised and crushed through a 5 mL disposable syringe, and the inserts were each eluted into 6 mL 1X TE overnight. The inserts were desalted by passage through a NAP-10 column, equilibrated in ddH2O, and elution with 1.5 mL ddH2O. The inserts were lyophilized, resuspended in 360 µL ddH2O, further purified by ethanol precipitation, and resuspended in a final volume of 100 µL ddH2O for Library 1 and 170 µL ddH2O for Library 2.

Vector construction

The phagemid vector pJC21 containing p007 (a gift from Dr. Jason Chin) was prepared from 200 mL of XL1 Blue E. coli using the Qiagen Maxi-prep Kit (Qiagen). This procedure yielded 250 µL of plasmid at 1.5 mg/mL from 200 mL of cells. pJC21-p007 was triple-digested with BglIII, PstI, and BssHIII in the following reaction: 60 µg plasmid, 75 µL 10X NEBuffer 3, 500 U BglIII, 1000 U PstI, 200 U BssHIII in a total volume of 750 µL. The reaction was incubated for 5 h at 37 °C. In order to digest any residual singly digested or undigested pJC21-p007 remaining with BssHIII, the reaction was incubated for 2 additional h at 50 °C. The triple-digested pJC21-p007 vector was purified from the excised DNA using Chromaspin-1000 columns, following the manufacturer’s protocol.

Ligations
The Library 1 and Library 2 inserts were ligated into pJC21-p007 using T4 DNA ligase, following the manufacturer’s protocol with the following DNA components: 1135 ng digested pJC21-p007 and digested Library 1 and Library 2 to vector:insert molar ratios of 1:1, 1:5, 1:10 or 1:20. The ligated DNAs were purified from protein contaminants by ethanol precipitation, and each reaction was resuspended in 5 µL ddH₂O. The ligated library vectors were transformed by electroporation into XL1 Blue E. coli cells; each microliter of DNA was transformed into 40 µL cells. Cells were recovered by the addition of 960 µL of 2X YT-G and incubation with shaking for 1 h at 37 °C. 20 µL of ten-fold serial dilutions was plated on SOB-AG plates and incubated at 37 °C overnight. For Library 1, the recovered cells from three transformations were combined, and added to 75 mL 2X YT-AG. For Library 2, the recovered cells from 20 transformations were combined, and added to 500 mL 2X YT-AG. The cultures were incubated at 37 °C with shaking at 250 rpm for 15 h. These cultures were made into glycerol stocks by addition of 0.4 mL 50% glycerol to each 0.8 mL of cells and freezing in a dry ice/ethanol bath.

Library quality

To assess the purity of the library from contaminating parental clones, plasmid DNA was extracted from 26 clones selected at random from each library and subjected to digestion with a restriction endonuclease for which the DNA target site in pJC21-p007 is destroyed upon library construction. Single library clones were picked from SOB-AG plates and used to inoculate 4 mL of 2X YT-AG. Plasmid DNA was extracted from 2 mL of an overnight culture using the QIAprep Spin Miniprep Kit. Library 1 clones were
double-digested with EcoRI and ApaLI in the following reaction: 7 µL plasmid, 1.5 µL 10X NEBuffer 4, 1.5 µL 10X BSA, 10 U EcoRI, 5 U ApaLI in a total volume of 15 µL. Library 2 clones were double-digested with EcoRI and BssHII in the following reaction: 7 µL plasmid, 1.5 µL 10X NEBuffer 3, 10 U EcoRI, 2 U BssHII in a total volume of 15 µL. pJC21-p007 was digested as a positive control. The fragments were resolved by electrophoresis on a 1% analytical agarose gel. None of the library clones were background. However, DNA sequencing confirms that 19% of clones in each library contain single deletion or insertions, which is factored into the theoretical diversity of the libraries.

The number of independent transformants recovered for each library was determined by selection on SOB-AG agar. Library 1 contained approximately 1.3 x 10^7 independent transformants, of which 1.1 x 10^7 unmutated transformants were sufficient to cover the theoretical diversity of the library (32^4 = 1.0 x 10^6) with 100% confidence. Library 2 contained approximately 9.8 x 10^7 independent transformants, of which 7.9 x 10^7 unmutated transformants were sufficient to cover the theoretical diversity of the library (32^5 = 3.4 x 10^7) with 91% confidence.

**Phage display selections**

A glycerol stock of the initial pool (round 1), output from the previous round (rounds 2-8), or a starter culture of pJC21-p007-harboring XL1 Blue E. coli cells was used to inoculate 10 mL 2X YT-AG media. The cultures were incubated at 37 °C until they reached an optical density of 0.8 absorbance units at 600 nm. The culture was then infected with 4 x 10^{10} pfu M13K07 helper phage (a gift from Dr. Jason Chin) and
incubated at 37 °C for 1 h. Cells were pelleted by centrifugation, resuspended in 10 mL 2X YT-AK media and incubated for 12 h at 37 °C. Cells were then pelleted by centrifugation and the retained supernatant was filtered through a 0.45 µm syringe filter. Phage were precipitated with 1/5 volume PEG/NaCl on ice for 45 min, and then pelleted by centrifugation at 20,000 g at 4 °C for 30 min, and resuspended in binding buffer.

**Bead preparation and DNA loading**

0.5 mg of streptavidin-coated M280 magnetic beads (Dynal) were washed six times with 50 µL of 2X B+W buffer. Each wash step was performed for two min and mixing was performed by rotation of the tubes on a LABQUAKE shaker rotisserie. The beads were blocked by incubation in 1X B+W buffer containing 6% nonfat milk (Carnation) for 14 h, mixing as described above. The beads were then washed 5 times with 50 µL of 1X B+W buffer and resuspended in 50 µL of 1X B+W buffer containing 0.5 µM biotinylated duplex PICS5 or PICS3 and incubated for 12 min. The loading onto the beads was quantified by measuring the A<sub>260</sub> of the DNA-containing solution before and after exposure to the beads. The difference between these values was taken as the OD of the DNA loaded on the beads. This analysis suggested that the procedure loaded approximately 100-200 pmol DNA per mg bead. The beads were then washed 5 times with 50 µL of phage binding buffer.

**Panning**

10<sup>10</sup> phage in a volume of 0.4 mL were added to the beads prepared as described above and incubated with rotation on a LABQUAKE shaker rotisserie for 2 h at 4 °C.
Beads were washed 5 times for 5 min at 4 °C. Bound phage were eluted by the addition of 200 μL of elution buffer and an increase in temperature to 25 °C for 2 h. 200 μL of the elution and 200 μL of phage not subjected to panning were used to infect 6 mL of log phase XL1 Blue *E. coli* cells. After 1 h, serial dilutions of infected cells were plated on SOB-AG and grown for 12 h at 37 °C. Values of percent retention were calculated from the equation percent retention = (output titer / input titer) x 100. The log phase XL1 Blue *E. coli* used in these experiments were grown from a single colony picked from an LB-tet plate. Cells infected with output phage were used to make glycerol stocks and stored at -70 °C.

**Peptides**

Peptides were synthesized on a 25 μmol scale using Fmoc chemistry at the W. M. Keck Foundation Biotechnology Resource Laboratory. Peptides were dissolved in acetonitrile and water, then filtered through a 0.2 μm filter. Peptides were purified by reverse phase HPLC on a Varian ProStar 210 Solvent Delivery System with a ProStar 330 PDA Detector. Peptide purifications were performed on a Vydac semi-preparative C18 column (300Å, 5 μm, 10 mm x 150 mm) at a flow rate of 5 mL/min. Solvent A contained 98% HPLC grade water, 2% HPLC grade acetonitrile, 0.06% trifluoroacetic acid, and solvent B contained 20% HPLC grade water, 80% HPLC grade acetonitrile, and 0.05% trifluoroacetic acid. Peptide purity was confirmed by reinjection on a Vydac analytical C18 column (300Å, 5 μm, 4.6 mm x 150 mm). Concentrated peptide stocks in ddH₂O were made from lyophilized HPLC purified peptides. Peptide concentrations were determined by amino acid analysis, using homoserine and norleucine as internal
standards, on a Beckman 7300 instrument. Mass spectrometry on an Applied Biosystems (Foster City, CA) Voyager-DE-Pro MALDI-TOF instrument using an \( \alpha \)-cyano-4-hydroxy-cinnamic acid matrix confirmed the identity of all peptides. Peptide name is in bold and W.M. Keck Foundation Biotechnology Resource Laboratory numbering is denoted in parentheses.

**APPS (5179):** Mass expected: 3708.3; observed: major peak of 3710.3 and minor peak of 3613.3 (contaminant which could not be removed by HPLC).

**APSA (5175):** Amino acid analysis expected: Asx4 Thr1 Ser3 Glx1 Pro3 Gly3 Ala8 Val1 Met1 Leu2 Phe1 Lys2 Arg5; observed: Asx4.4 Thr1.0 Ser2.8 Glx1.1 Pro3.3 Gly3.2 Ala8.3 Val1.1 Met0.9 Leu2.1 Phe1.1 Lys2.2 Arg4.5. Mass expected: 3682.1; observed: 3682.9.

**DAVA (5174):** Amino acid analysis expected: Asx5 Thr1 Ser2 Glx1 Pro2 Gly3 Ala8 Val2 Met1 Leu2 Phe1 Lys2 Arg5; observed: Asx5.5 Thr0.9 Ser1.9 Glx1.1 Pro2.2 Gly3.4 Ala8.4 Val1.9 Met0.8 Leu2.2 Phe1.0 Lys2.2 Arg4.5. Mass expected: 3712.2; observed: 3713.5.

**SPLR (5178):** Amino acid analysis expected: Asx4 Thr1 Ser3 Glx1 Pro3 Gly3 Ala6 Val1 Met1 Leu3 Phe1 Lys2 Arg6; observed: Asx4.4 Thr1.0 Ser2.8 Glx1.1 Pro3.3 Gly3.3 Ala6.3 Val1.0 Met0.9 Leu3.3 Phe1.1 Lys2.2 Arg5.5. Mass expected: 3809.3; observed: 3812.3.

**SQAP (5177):** Amino acid analysis expected: Asx4 Thr1 Ser3 Glx2 Pro3 Gly3 Ala7 Val1 Met1 Leu2 Phe1 Lys2 Arg5; observed: Asx4.4 Thr1.0 Ser2.7 Glx2.1 Pro3.2 Gly3.2 Ala7.3 Val1.0 Met0.9 Leu2.1 Phe1.1 Lys2.2 Arg4.5. Mass expected: 3739.2; observed: 3741.9.
**TAML** (5176): Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala7 Val1 Met2 Leu3 Phe1 Lys2 Arg5; observed: Asx4.4 Thr1.9 Ser1.7 Glx1.0 Pro2.1 Gly3.4 Ala7.4 Val11.0 Met2.0 Leu3.2 Phe1.0 Lys2.2 Arg4.5. Mass expected: 3772.3; observed: 3774.9.

**TAMR** (5347): Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala7 Val1 Met2 Leu2 Phe1 Lys2 Arg6; observed: Asx4.3 Thr1.9 Ser1.8 Glx1.1 Pro2.4 Gly3.3 Ala7.4 Val11.0 Met1.7 Leu2.2 Phe1.1 Lys2.2 Arg5.5. Mass expected: 3815.4; observed: 3816.9.

**APLMS** (5180): Amino acid analysis expected: Asx3 Thr2 Ser2 Glx1 Pro3 Gly3 Ala6 Val1 Met2 Leu3 Phe1 Lys1 Arg7; observed: Asx3.3 Thr2.0 Ser2.0 Glx1.1 Pro3.2 Gly3.3 Ala6.3 Val11.0 Met2.0 Leu3.2 Phe1.1 Lys1.1 Arg6.4. Mass expected: 3868.5; observed: 3870.9.

**HRPVQ** (5181): Amino acid analysis expected: Asx3 Thr2 Ser1 Glx2 Pro3 Gly3 Ala5 Val2 Met1 Leu2 Phe1 His1 Lys1 Arg8; observed: Asx3.3 Thr2.0 Ser1.1 Glx2.2 Pro3.4 Gly3.3 Ala5.2 Val11.9 Met0.8 Leu2.2 Phe1.0 His1.1 Lys1.1 Arg7.3. Mass expected: 3986.5; observed: 3989.2.

**LRLPY** (5348): Amino acid analysis expected: Asx3 Thr2 Ser1 Glx1 Pro3 Gly3 Ala5 Val1 Met1 Leu4 Tyr1 Phe1 Lys1 Arg8; observed: Asx3.3 Thr2.0 Ser0.9 Glx1.1 Pro3.1 Gly3.3 Ala5.3 Val11.1 Met0.9 Leu4.4 Tyr1.1 Phe1.1 Lys1.1 Arg7.4. Mass expected: 4011.6; observed: 4009.7.

**NPHYT** (5182): Amino acid analysis expected: Asx4 Thr3 Ser1 Glx1 Pro3 Gly3 Ala5 Val1 Met1 Leu2 Tyr1 Phe1 His1 Lys1 Arg7; observed: Asx4.5 Thr3.1 Ser1.1 Glx1.1 Pro3.2 Gly3.2 Ala5.1 Val0.9 Met0.7 Leu2.2 Tyr1.2 Phe1.1 His1.1 Lys1.1 Arg6.3.
Mass expected: 3981.5; observed: 3983.2.

**PLWTP** (5183): Amino acid analysis expected: Asx3 Thr3 Ser1 Glx1 Pro4 Gly3 Ala5 Val1 Met1 Leu3 Phe1 Lys1 Arg7; observed: Asx3.6 Thr3.3 Ser1.2 Glx1.2 Pro4.3 Gly3.5 Ala5.5 Val1.0 Met0.8 Leu3.4 Phe1.1 Lys1.2 Arg6.7. Mass expected: 3963.6; observed: 3966.5.

**PRPQQ** (5184): Amino acid analysis expected: Asx3 Thr2 Ser1 Glx3 Pro4 Gly3 Ala5 Val1 Met1 Leu2 Phe1 Lys1 Arg8; observed: Asx3.3 Thr2.1 Ser1.0 Glx3.2 Pro4.3 Gly3.3 Ala5.2 Val1.0 Met0.9 Leu2.2 Phe1.1 Lys1.1 Arg7.3. Mass expected: 3975.5; observed: 3977.4.

**RPNST** (5349): Amino acid analysis expected: Asx4 Thr3 Ser2 Glx1 Pro3 Gly3 Ala5 Val1 Met1 Leu2 Phe1 Lys1 Arg8; observed: Asx3.3 Thr2.9 Ser1.9 Glx1.1 Pro3.2 Gly3.3 Ala5.4 Val1.1 Met0.9 Leu2.2 Phe1.1 Lys1.2 Arg7.4. Mass expected: 3924.4; observed: 3923.2.

**SHQRA** (5350): Amino acid analysis expected: Asx3 Thr2 Ser2 Glx3 Pro2 Gly3 Ala6 Val1 Met1 Leu2 Phe1 His1 Lys1 Arg8; observed: Asx3.2 Thr2.0 Ser1.8 Glx2.2 Pro2.3 Gly3.3 Ala6.4 Val1.1 Met0.9 Leu2.2 Phe1.1 Lys1.21 Arg7.3. Mass expected: 3948.5; observed: 3948.3.

**SRTMS** (5185): Amino acid analysis expected: Asx3 Thr3 Ser3 Glx1 Pro2 Gly3 Ala5 Val1 Met2 Leu2 Phe1 Lys1 Arg8; observed: Asx3.4 Thr3.1 Ser2.9 Glx1.1 Pro2.3 Gly3.3 Ala5.3 Val1.0 Met1.7 Leu2.2 Phe1.1 Lys1.1 Arg7.4. Mass expected: 3931.5; observed: 3931.9.

**STHMA** (5186): Amino acid analysis expected: Asx3 Thr3 Ser2 Glx1 Pro2 Gly3 Ala6 Val1 Met2 Leu2 Phe1 His1 Lys1 Arg7; observed: Asx3.2 Thr2.8 Ser1.8 Glx1.1
Pro2.3 Gly3.3 Ala6.4 Val1.1 Met1.9 Leu2.2 Phe1.1 His1.1 Lys1.2 Arg6.4. Mass expected: 3896.4; observed: 3899.2.

**TAPRH (5187):** Amino acid analysis expected: Asx3 Thr3 Ser1 Glx1 Pro3 Gly3 Ala6 Val1 Met1 Leu2 Phe1 His1 Lys1 Arg8; observed: Asx3.2 Thr2.9 Ser0.9 Glx1.1 Pro3.2 Gly3.2 Ala6.5 Val1.1 Met0.9 Leu2.2 Phe1.1 His1.1 Lys1.2 Arg7.3. Mass expected: 3931.5; observed: 3934.9.

**DNA purification**

Oligonucleotides were purified by denaturing gel electrophoresis on a 20% acrylamide (19:1 acrylamide:bisacrylamide crosslinking) gel containing 1X TBE and 8 M urea. The oligonucleotide was visualized by UV shadowing, excised from the gel, and eluted in 1X TE. Gel filtration on a NAP-10 column (Pharmacia) removed urea and salt from the oligonucleotide. The concentration and purity of the oligonucleotide was determined by measuring its UV absorbance at 260 nm.

**Radiolabeling DNA**

The B strands of hsCRE$_{25}$, NON$_{24}$, PICS5, PICS3, or QREP were radiolabeled at the 5' end with [$\gamma$-32P]-ATP and T4 polynucleotide kinase. Labeling reactions were carried out in a reaction volume of 10 µL containing 100 pmol of DNA, 1X kinase buffer (70 mM Tris-HCl, 10 mM MgCl$_2$, 5 mM DTT, pH 7.6), 10 U of T4 polynucleotide kinase (New England Biolabs), and 10 µCi of [$\gamma$-32P]-ATP (New England Nuclear). The reactions were incubated for at least 1 h at 37 °C and the reaction volume brought up to a volume of 50 µL with dH$_2$O. Reactions were extracted with an equal volume of 25:24:1
phenol:chloroform:isoamyl alcohol, followed by an equal volume of 24:1 chloroform:isoamyl alcohol. Unincorporated label was removed by passage through a G-25 spin column (Pharmacia). 1 µL of a 1 µM solution of the complementary strand was added and the mixture heated to 95 °C for 2 min in a hot block. The block was then removed from heat and the DNA strands allowed to anneal by slow cooling to room temperature.

**Electrophoretic mobility shift assay**

To quantify the affinity of the selected peptides for DNA, each peptide was titrated with a fixed concentration of radiolabeled DNA and the free and bound DNA separated by gel electrophoresis. In a typical binding procedure, a peptide was serially diluted from a stock of known concentration into binding buffer [1.4 mM KH$_2$PO$_4$, 4.3 mM Na$_2$HPO$_4$, 2.7 mM KCl, 137 mM NaCl (pH 7.4), 1 mM EDTA, 0.1% NP-40, 0.4 mg/ml BSA, and 5% glycerol]. To 8 µL of this serial dilution was added 2 µL of $^{32}$P-labeled double stranded DNA. The final concentration of DNA was < 200 pM, ensuring that the concentration of labeled DNA was at least 10-fold lower than the K$_d$ of the peptide being tested. The binding reactions were equilibrated for 1 h at 4 °C or 25 °C. 8 µL of each reaction was loaded in a single well of the pre-equilibrated gel.

Polyacrylamide gels were 8% w/v with 79:1 acrylamide (ICN ultrapure):bisacrylamide (American Bioanalytical) crosslinking in 10 mM Tris, pH 8.1. Gels were by deoxygenation of the acrylamide solution by bubbling nitrogen followed by polymerization with 0.15% of a 20% ammonium persulfate solution and 0.075% tetramethylenediamine (TEMED) (American Bioanalytical). Gels were pre-equilibrated
to 4 °C or 25 °C and pre-electrophoresed at 100 V for 30 min. After samples were loaded the gels were run at 500 V for 30 min. After electrophoresis, gels were transferred to Whatman filter paper and dried with vacuum at 80 °C on a Model 583 Gel Dryer (Bio-Rad). The dry gels were exposed to Phosphor screens (Molecular Dynamics) and the intensity of bands imprinted on the Phosphor screen imaged using a Storm 840 Phosphoimager (Molecular Dynamics). The amounts of free and bound DNA were quantified using ImageQuant software (Molecular Dynamics). Dissociation constants (K_d) were calculated by fitting the data to the Langmuir equation (1.1), using non-linear least squares fitting in the program Kaleidagraph 3.6 (Synergy Software), where Θ is the fraction DNA bound = counts bound DNA/(counts bound DNA + counts free DNA), [peptide]_T = total peptide concentration, and c is an adjustable parameter representing the maximum fraction bound (132,133).

1.1

\[ \Theta = c \left( 1 + \frac{K_d}{[\text{peptide}]_T} \right)^{-1} \]
References


32, 9668-76.


USA 96, 9521-6.


102


CHAPTER 2

Relationship between folding and function in a sequence-specific miniature DNA-binding protein
Introduction

Chemical biology has blossomed in the post-genome era; modulating protein function through selective recognition of macromolecular targets has emerged as a major challenge for chemical biologists. Molecules capable of tight and (especially) selective recognition have widespread utility in elucidating protein function and, in certain cases, have potential therapeutic value (1-3). Natural products are a logical starting point in the search of biologically active compounds because they are often highly potent. However, generating large stocks of these compounds and their derivative for research objectives or medicinal purposes can be challenging. These limitations have directed chemical biologists towards combinatorial libraries as a source for large pools of structurally diverse compounds that may possess a desired biological activity. Another alternative to natural products are the various selection methods, such as SELEX and phage display, which can evolve ligands for desired physical properties, such as high affinity and stability. Ideally, a combination of these strategies will produce molecules that are small, well-structured, and possess high affinity and specificity for their target.

Miniature proteins based on avian pancreatic polypeptide

Miniature protein design is a technology developed in our laboratory that has the potential to address a number of the challenges to chemical biology described above. This technology currently represents a general approach for the design of small, well-folded proteins that bind macromolecular targets with high affinity and selectivity (4-11). This approach is referred to as protein grafting, and the molecules that result are
miniature proteins – miniature because they have molecular weight less than 4200 and proteins because they often exhibit cooperative melting transitions. Using the protein grafting strategy, the residues that comprise a functional epitope from a natural ligand are grafted onto the small yet stable protein avian pancreatic polypeptide (aPP). The remarkable stability of aPP derives from the packing of proline side chains along an N-terminal type II polyproline (PPII) helix and hydrophobic residues aligned on one face of a C-terminal alpha helix (12,13). The aPP scaffold presents both a solvent-exposed alpha helical face and a type II polyproline helix, which can both be substituted depending on which face is more akin to the native protein ligand. This procedure, often in combination with molecular evolution, can identify miniature protein ligands with high affinity and specificity for macromolecular targets.

The protein grafting strategy was first applied to DNA recognition by grafting the alpha helical functional epitope of the bZIP protein GCN4 onto aPP, yielding the rationally designed peptide PPBR4, which binds the half-site CRE (hsCRE) sequence (Figure 2.1a) (4). Though PPBR4 bound DNA with low nanomolar affinity at 4 °C, it was not folded in the absence of DNA. In some cases, such as PPBR4, protein grafting can destabilize the aPP scaffold, incurring an entropic penalty on the ligand and diminishing binding affinity. The PPII helix of PPBR4 was then optimized using a phage display functional selection, resulting in the well-folded miniature protein, p007, that binds target DNA with nanomolar affinity at 25 °C and enhanced specificity over PPBR4 (Figure 2.1b) (5). Optimization of grafted miniature proteins is not always necessary, as with the case of the miniature Q50K engrailed homeodomain mimic PPeng4, which was able to bind DNA without the contacts provided by the parent protein’s N-terminal arm.
Figure 2.1 Miniature aPP-based proteins that target DNA. **A.** Protein grafting strategy for PPBR4, a GCN4 mimic that binds hsCRE (ATGAC) (4). **B.** Evolution of PPBR4 into p007 (7). **C.** Protein grafting strategy for PPeng4, a Q50K *engrailed* homeodomain mimic that binds the QRE site (TAATCC) (9).
A

Functional selection of GCN4-DNA complex residues required for DNA recognition. Dissect residues required for DNA recognition, then graft residues required for DNA recognition on aPP scaffold.

B

Functional selection of miniature protein library based on PPBR4 to find p007.

C

Dissect DNA contact residues from Q50K-QRE complex, ignore N-terminal arm, and graft residues from helix 3 to aPP to form PPeng4.
Our protein grafting strategy was also applied to the design of ligands that recognize protein surfaces, including the relatively deep clefts on the anti-apoptotic proteins Bcl-2 and Bcl-X\textsubscript{l} (Figure 2.2a) (7,11), as well as the shallow cleft of the KIX domain of the transcriptional coactivator CPB (Figure 2.2b) (9). Furthermore, we have utilized the PPII helix of aPP to target the binding groove on EVH1 proteins (Figure 2.3a) (10). Other laboratories have also met with success using the aPP scaffold to create functional miniature proteins. Grafting the alpha helical epitope of the Kaposi’s sarcoma-associated herpesvirus protease (KSHV Pr) onto the aPP scaffold generates a miniature protein that disrupts formation of the obligate dimer, which is required for activity (Figure 2.2c). The dissociation constant of the KSHV Pr dimer has been reported as approximately 1 µM, and the grafted protein binds with an estimated dissociation constant of 300 µM (14). Another laboratory has grafted residues from a proline rich peptide, RP1, onto the PPII helix of aPP to generate miniature proteins that bind the Abl-SH3 domain with dissociation constants ranging from 20-90 µM (Figure 2.3b) (15).

**Alanine scanning mutagenesis**

Protein grafting is obviously a developing technology; it is not yet clear which residues within the aPP scaffold may be mutated without losing structure and affecting potential binding affinity. Furthermore, it would be useful to score each position for its contributions to structure and utility for grafting. Alanine scanning mutagenesis is a well-established method for mapping of ligand binding sites or “hot spots” via the systematic substitution of individual amino acids with alanine (16-18). Substitution of an
Figure 2.2 Miniature aPP-based proteins that target alpha helix-binding proteins. A. Protein grafting strategy and evolution of PPBH3 proteins, mimics of the BH3 domain of Bak that bind Bcl-X\textsubscript{L} and Bcl-2 (5,13). B. Protein grafting strategy and evolution of PPKID proteins, mimics of the KID domain of CREB that bind the KIX domain of CBP (8). C. Protein grafting strategy for KSaPP. LEFT: Dimer interface in the crystal structure of KSHV Pr, showing residues in the dimer interface (red and purple) and residues in the catalytic triad (orange). RIGHT: A model structure of the KSHV Pr-KSaPP peptide complex constructed by replacing the α5 helix in one monomer of KSHV Pr with aPP (14).
A. Introduce diversity at select residues required for Bcl-2 discrimination.

PPBH3-5/PPBH3-6: Bcl-2-selective

PPBH3-1: Bcl-XL-selective

functional selection

introduce diversity

protein grafting

...for Bcl-2 recognition

B. KID^5•KIX complex

PPKID peptides

- PKA

+ PKA

functional selection

PPKID Library 1

C. Catalytic triad

α5 helix

α2 helix

α1 helix

KSHV Pr
Figure 2.3  Miniature aPP-based proteins that target proteins which bind proline-rich sequences.  A. Protein grafting strategy for pGolemi, an FP4 peptide mimic that binds the Mena EVH1 domain (10). B. The Abl-SH3 domain (yellow) complexed with a proline-rich peptide (green) and its three-dimensional superimposition with aPP (orange) (15).
Abl-SH3 domain

Proline-rich peptide

Mena-binding miniature protein

FP4 functional epitope

aPP scaffold

Mena•FP4

Proline-rich peptide

Abl-SH3 domain
amino acid with alanine truncates the side chain at the beta carbon, generally without disruption of the peptide backbone. Since prolines are N-substituted amino acids, they are scanned using both alanine and sarcosine (N-methylglycine) in independent experiments. Analyzing both alanine isomers provides information that can distinguish independent effects of the constrained ring, N-substitution, and side chain truncation.

Herein I describe alanine scanning mutagenesis of the functional miniature protein, p007, for the purpose of better delineating sources of structural stability and positions that can accept substitution for other functions. This incremental substitution method will reveal the difference in free energy of binding between a substituted protein and the wild type protein and relate the difference to mutation of a single side chain. In the present study, alanine scanning mutagenesis is not applied to the more general problem of defining an epitope. Rather, it is applied to a detailed thermodynamic study of a miniature protein whose function is already known and for the validation of its design strategy and evolution. In contrast to a large globular protein, local and global effects of mutations in a miniature protein could be one and the same. Therefore, a greater wealth of information can be obtained per mutation from an alanine scanning mutagenesis study of a miniature protein than of a large globular protein. Understanding of contributions to folding and binding energies on the residue level should provide insight into the contribution of structural preorganization to DNA-binding affinity and specificity, and in addition, will guide the design of future functional miniature proteins.

Results
**Experimental goals**

To enhance our understanding of miniature protein design, we characterized one of our best miniature proteins, which mimics GCN4, p007. A complete alanine scanning mutagenesis study of p007 was performed to determine the individual contribution of each side chain to structure, DNA-binding affinity, and sequence specificity.

**Quantitative analysis of the contributions of individual side chains to DNA affinity**

Twenty-nine single alanine p007 variants were synthesized to study the energetic contribution of side chains of p007 to DNA affinity. Additionally, two proline variants were designed in which the proline residue was replaced with sarcosine (Figure 2.4). The affinity of each p007 variant for duplex oligonucleotide hsCRE\textsubscript{25} (5’-AGTGGAGATGACGAGCTACTCGTGC-3’) was assayed using a quantitative electrophoretic mobility shift assay (Figure 2.5). In brief, a constant concentration of radiolabeled DNA (< 40 pM) was incubated for 1 h with a serial dilution of each p007 variant ranging from 0.244 nM to 64 µM. The free DNA and peptide-bound DNA species were separated by nondenaturing polyacrylamide gel electrophoresis. The population of each DNA species was quantified through the radioactivity emitted, and the data was fit to the Langmuir equation yielding the equilibrium dissociation constant (K\textsubscript{d}). Dissociation constants for the p007 variants and hsCRE\textsubscript{25} DNA ranged from 1.5 to 691 nM, which corresponds to binding free energies between −12.0 to −8.4 kcal•mol\textsuperscript{−1} (Table 2.1, Figure 2.6b). The affinity of each variant peptide was then compared to that of wild type p007 (\(\Delta G_{p007\cdot hsCRE} = -12.0\) kcal•mol\textsuperscript{−1}). The change in binding energy of each peptide \(\Delta\Delta G_{hsCRE}\) is defined by the free energy relationship \(\Delta\Delta G_{hsCRE} = \Delta G_{variant\cdot hsCRE} - \Delta G_{p007\cdot hsCRE}\).
Figure 2.4 Variants of p007 used in this study.
A = alanine

P = proline

Z = sarcosine
Figure 2.5 Semilogarithmic plots illustrating the affinity of each p007 variant for a 25 bp oligonucleotide containing the target sequence ATGAC. Binding reactions were performed at 25 °C in PBS buffer supplemented with 0.1% NP-40, 0.4 mg/ml BSA, and 5% glycerol. Curves represent the best fit of the data to the equation \( \theta = C \cdot \frac{1}{1 + K_d \cdot [\text{peptide}]_t} \), where \( K_d \) and \( C \) are adjustable parameters. Each point represents an average of three independent determinations. Error bars denote the standard error.
Table 2.1  Affinities of peptides for hsCRE$_{25}$ DNA at 25 °C. The value for p007 was determined previously.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>$K_d$ (nM)</th>
<th>$\Delta G_{\text{hsCRE}}$ (kcal•mol$^{-1}$)</th>
<th>$\Delta \Delta G_{\text{hsCRE}}$ (kcal•mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p007</td>
<td>1.6 ± 0.1</td>
<td>-12.0</td>
<td>0.0</td>
</tr>
<tr>
<td>p007$^{G1A}$</td>
<td>18.5 ± 5.5</td>
<td>-10.5</td>
<td>1.4</td>
</tr>
<tr>
<td>p007$^{G2A}$</td>
<td>91.0 ± 28.0</td>
<td>-9.6</td>
<td>2.4</td>
</tr>
<tr>
<td>p007$^{S3A}$</td>
<td>16.5 ± 1.4</td>
<td>-10.6</td>
<td>1.4</td>
</tr>
<tr>
<td>p007$^{R4A}$</td>
<td>427 ± 66</td>
<td>-8.7</td>
<td>3.3</td>
</tr>
<tr>
<td>p007$^{T6A}$</td>
<td>39.3 ± 9.1</td>
<td>-10.1</td>
<td>1.9</td>
</tr>
<tr>
<td>p007$^{M7A}$</td>
<td>9.5 ± 1.7</td>
<td>-10.9</td>
<td>1.0</td>
</tr>
<tr>
<td>p007$^{P8A}$</td>
<td>90.7 ± 19.0</td>
<td>-9.6</td>
<td>2.4</td>
</tr>
<tr>
<td>p007$^{P8Z}$</td>
<td>31.0 ± 6.0</td>
<td>-10.2</td>
<td>1.8</td>
</tr>
<tr>
<td>p007$^{G9A}$</td>
<td>14.8 ± 1.3</td>
<td>-10.6</td>
<td>1.3</td>
</tr>
<tr>
<td>p007$^{H10A}$</td>
<td>2.3 ± 0.5</td>
<td>-11.7</td>
<td>0.2</td>
</tr>
<tr>
<td>p007$^{D11A}$</td>
<td>17.1 ± 5.1</td>
<td>-10.6</td>
<td>1.4</td>
</tr>
<tr>
<td>p007$^{P13A}$</td>
<td>20.0 ± 2.6</td>
<td>-10.5</td>
<td>1.5</td>
</tr>
<tr>
<td>p007$^{P13Z}$</td>
<td>12.0 ± 1.0</td>
<td>-10.8</td>
<td>1.2</td>
</tr>
<tr>
<td>p007$^{V14A}$</td>
<td>219 ± 32</td>
<td>-9.0</td>
<td>2.9</td>
</tr>
<tr>
<td>p007$^{E15A}$</td>
<td>1.8 ± 0.2</td>
<td>-11.9</td>
<td>0.1</td>
</tr>
<tr>
<td>p007$^{D16A}$</td>
<td>1.5 ± 0.1</td>
<td>-12.0</td>
<td>0.0</td>
</tr>
<tr>
<td>p007$^{L17A}$</td>
<td>168 ± 21</td>
<td>-9.2</td>
<td>2.7</td>
</tr>
<tr>
<td>p007$^{K18A}$</td>
<td>239 ± 39</td>
<td>-9.0</td>
<td>3.0</td>
</tr>
<tr>
<td>p007$^{R19A}$</td>
<td>633 ± 81</td>
<td>-8.4</td>
<td>3.5</td>
</tr>
<tr>
<td>p007$^{F20A}$</td>
<td>93.3 ± 12.8</td>
<td>-9.6</td>
<td>2.4</td>
</tr>
<tr>
<td>p007$^{R21A}$</td>
<td>245 ± 25</td>
<td>-9.0</td>
<td>3.0</td>
</tr>
<tr>
<td>p007$^{N22A}$</td>
<td>608 ± 121</td>
<td>-8.4</td>
<td>3.5</td>
</tr>
<tr>
<td>p007$^{T23A}$</td>
<td>9.0 ± 1.0</td>
<td>-10.9</td>
<td>1.0</td>
</tr>
<tr>
<td>p007$^{S24A}$</td>
<td>125 ± 21</td>
<td>-9.4</td>
<td>2.6</td>
</tr>
<tr>
<td>p007$^{R27A}$</td>
<td>551 ± 146</td>
<td>-8.5</td>
<td>3.4</td>
</tr>
<tr>
<td>p007$^{R28A}$</td>
<td>503 ± 108</td>
<td>-8.6</td>
<td>3.4</td>
</tr>
<tr>
<td>p007$^{S29A}$</td>
<td>86.5 ± 10.9</td>
<td>-9.6</td>
<td>2.4</td>
</tr>
<tr>
<td>p007$^{R30A}$</td>
<td>691 ± 173</td>
<td>-8.4</td>
<td>3.6</td>
</tr>
<tr>
<td>p007$^{R32A}$</td>
<td>422 ± 83</td>
<td>-8.7</td>
<td>3.3</td>
</tr>
<tr>
<td>p007$^{R33A}$</td>
<td>593 ± 111</td>
<td>-8.5</td>
<td>3.5</td>
</tr>
<tr>
<td>p007$^{K36A}$</td>
<td>85.0 ± 14.4</td>
<td>-9.6</td>
<td>2.3</td>
</tr>
</tbody>
</table>

$\Delta G_{\text{hsCRE}} = -RT\ln K_d$  
$\Delta \Delta G_{\text{hsCRE}} = \Delta G_{\text{variant•hsCRE}} - \Delta G_{\text{p007•hsCRE}}$
$\Delta G_{p007 \text{hsCRE}}$ (Table 2.1). In the sections below, we consider the roles of each class of residues found within p007.

**Contributions of residues identified by functional selection**

Four residues located on the N-terminal PPII helix of p007 draw particular interest in the alanine scan because they were selected during the molecular evolution of p007: Gly2, Arg4, Ala5, and Met7 (5). The three p007 variants containing alanine replacing these residues bound hsCRE significantly worse than p007. Dissociation constants for these peptides in complex with hsCRE range between 9.5 and 427 nM, which correspond to $\Delta \Delta G_{\text{hsCRE}}$ of 1.0 to 3.3 kcal$\cdot$mol$^{-1}$ relative to p007. Of the three variants, p007$^{R4A}$ bound hsCRE with the lowest affinity ($K_d = 427 \pm 66$ nM; $\Delta \Delta G_{\text{hsCRE}} = 3.3$ kcal$\cdot$mol$^{-1}$); p007$^{G2A}$ also bound hsCRE poorly ($K_d = 91 \pm 28$ nM; $\Delta \Delta G_{\text{hsCRE}} = 2.4$ kcal$\cdot$mol$^{-1}$); p007$^{M7A}$ had a more modest effect ($K_d = 9.5 \pm 1.7$ nM and $\Delta \Delta G_{\text{hsCRE}} = 1.0$ kcal$\cdot$mol$^{-1}$).

**Contributions of putative DNA-binding residues**

Next we consider those residues within p007 that are expected to directly contact DNA in the context of the well-characterized structures of GCN4 bound to DNA: Arg19, Arg21, Asn22, Thr23, Ala25, Ala 26, Arg27, Arg28, Ser29, Arg30, Arg32, and Lys33 (19,20). The majority of these residues, highlighted in red in Figure 2.6a, are crucial to DNA-binding affinity; alanine substitutions at these positions reduce the free energy of binding by $\geq 3.0$ kcal$\cdot$mol$^{-1}$ relative to p007. This is consistent with the hypothesis that if the contacts made by p007 accurately mimic those of GCN4, these residues should also
Figure 2.6 The free energy of binding and specificity of p007 variants. A. The primary sequences of aPP, GCN4(226-252), PPBR4, and p007. B. Graph illustrating the stability of each variant complex [$\Delta G_{\text{hsCRE}} = -RT\ln K_d$]. C. Graph illustrating the specificity of each variant for hsCRE and CT DNA [$\Delta\Delta G_{\text{spec}} = \Delta G_{\text{hsCRE}} - \Delta G_{\text{CT DNA}}$]. Variants that contain alanine in place of selected residues from phage display evolution of PPBR4 into p007 are green. Variants that contain alanine in place of structural residues from aPP are blue. Variants that contain alanine in place of DNA-binding residues from GCN4 are red. All other variants are white. Variants that contain sarcosine in place of proline are hatched.
A

PPII helix  turn  alpha helix

aPP   GCN4
GFSQPTPGDDAPVEDLIRFYDNLQQYLNVVRTHRY
...DPAALKRANTEARAARRSKLQRMKQ...

PPBR4   p007
GFSQPTPGDDAPVEDLRFRNTLAAARRSRKAAARAAGGC
GGRATMPGDDAPVEDLRFRNTLAAARRSRKAAA

B

\[ \Delta G_{\text{hsCRE}} \text{ (kcal/mol)} \]

\[ \begin{array}{cccccccccccccccccccc}
\end{array} \]

C

\[ \Delta G_{\text{spec}} \text{ (kcal/mol)} \]

\[ \begin{array}{cccccccccccccccccccc}
\end{array} \]
contribute to p007’s DNA affinity by analogy.

Five of the DNA-binding residues of GCN4 make direct base contacts (Asn235, Ala238, Ala 239, Ser242, and Arg243). These include the two residues that are invariant in all bZIP proteins, Asn235 and Arg243. Asn235 contacts the thymine of the second T:A base pair and the cytosine of the third G:C base pair. Arg243 contacts the guanine of the central C:G base pair. Previous gel shift studies with the bZIP motif of GCN4 have shown that when Asn235 and Arg243 are each varied to all other amino acids, the majority of variants bind poorly to CRE or not at all (21-23). Other test systems also show that mutation of Asn235 decreases the activity of GCN4 in vivo (21,22). P007 variants with alanine at corresponding positions Asn22 and Arg30 bind hsCRE with the most reduced affinity ($K_d = 608 \pm 121$ nM and $691 \pm 173$ nM, respectively). These dissociation constants correspond to free energy losses of 3.5 and 3.6 kcal•mol$^{-1}$ relative to the wild type complex. The conserved basic residue Lys246 makes a water-mediated base contact when GCN4 is bound to CRE. As would be expected, the corresponding variant, p007$^{K33A}$, binds poorly ($K_d = 593 \pm 111$ nM and $\Delta \Delta G_{hsCRE} = 3.5$ kcal•mol$^{-1}$).

The arginines that make direct and water-mediated phosphate contacts in GCN4 are preserved in p007 (Arg19, Arg21, Arg27, Arg28, and Arg32). Mutation of these residues to alanine results in dissociation constants ranging from 245 to 503 nM ($\Delta \Delta G_{hsCRE} = 3.0$ to 3.4 kcal•mol$^{-1}$). Another residue which makes phosphate contacts, the base-contacting residue Ser242 in GCN4, interacts with the methyl group of thymine in the fourth A:T base pair. The corresponding p007$^{S29A}$ has a binding affinity reduced by only 2.4 kcal•mol$^{-1}$, in contrast to the other variants that have altered DNA-binding residues, which have reduced binding affinities of 3.0 kcal•mol$^{-1}$ or greater. The final
DNA-binding residue of GCN4, Thr236, makes no direct or water-mediated base contacts, but instead influences the conformation of nearby DNA-binding residues Arg232, Glu237, Arg243 so that these side chains are properly positioned for binding DNA. However, mutation of Thr23 to alanine in p007 resulted in the smallest reduction in DNA-binding affinity of all the DNA-binding variants \( (K_d = 9.0 \pm 1.0 \text{ nM} \) and \( \Delta \Delta G_{\text{hsCRE}} = 1.0 \text{ kcal} \cdot \text{mol}^{-1} \)). As predicted, most of the DNA-binding residues originally derived from GCN4 contributed significantly to p007’s ability to bind to hsCRE.

**Contributions of putative folding residues**

The final class of functional residues we examine includes residues of p007 that comprise the hydrophobic core of p007 derived from the aPP scaffold, Pro8, Leu17, Phe20, and Leu24. These residues are important for structure and preorganization of our ligand, and this is reflected in their contribution to binding affinity, \( \Delta \Delta G_{\text{hsCRE}} = 1.8 \) to 2.7 kcal•mol\(^{-1}\). NMR spectroscopy reveals long range NOEs between Pro8 and Leu17, and Pro8 and Phe20, in the PPII helix of p007, recapitulating a key folding element of aPP. Substitution of Leu17 and Phe20 to alanine results in a decrease in binding affinity of 2.7 kcal•mol\(^{-1}\) \( (K_d = 168 \pm 21 \text{ nM}) \) and 2.4 kcal•mol\(^{-1}\) \( (K_d = 93.3 \pm 12.8 \text{ nM}) \) respectively. Substitution of Pro8 to alanine also had a significant effect of 2.4 kcal•mol\(^{-1}\) \( (K_d = 90.7 \pm 19.0 \text{ nM}) \) on binding. However, substitution of Pro8 to sarcosine has a lesser effect, with \( \Delta \Delta G_{\text{hsCRE}} = 1.8 \text{ kcal} \cdot \text{mol}^{-1} \) and \( K_d = 31.0 \pm 6.0 \text{ nM} \). The third folding residue in the alpha helix of p007, Leu24, makes van der Waals contacts with Pro5, Phe20, Asn23, and Tyr27 in the crystal structure of aPP. In the context of p007, NOEs were seen between Leu24 and residues at the same positions (Phe20, Thr23, and
Arg27). P007^{T24A} had a similar effect on binding affinity as the other two folding residues in the alpha helix of p007 ($K_d = 125 \pm 21$ nM and $\Delta \Delta G_{hsCRE} = 2.6$ kcal•mol\(^{-1}\)). The putative folding residues contributed significantly to DNA-binding affinity, but slightly less than that of the DNA-binding residues.

**Contributions of other residues within the type II polyproline helix**

To complete this thorough analysis, the contribution of the remaining residues of p007, beginning with the N-terminal PPII helix were also quantified. Residues Gly1, Ser3, and Thr6 are located along the N-terminal PPII helix of p007/aPP. All show modestly reduced affinity, with dissociation constants ranging from 16.5 to 39.3 nM, corresponding to changes in free energy ranging from 1.4 to 1.9 kcal•mol\(^{-1}\). Of these three PPII variants, P007^{T6A} had a slightly greater reduction in DNA-binding affinity, ($K_d = 39.3 \pm 9.1$ nM and $\Delta \Delta G_{hsCRE} = 1.9$ kcal•mol\(^{-1}\)). P007^{G1A}, the start of the polyproline helix, had a smaller effect on DNA-binding affinity ($K_d = 18.5 \pm 5.5$ nM and $\Delta \Delta G_{hsCRE} = 1.4$ kcal•mol\(^{-1}\)). P007^{S3A} displayed a similar reduction in binding affinity ($K_d = 16.5 \pm 1.4$ nM and $\Delta \Delta G_{hsCRE} = 1.4$ kcal•mol\(^{-1}\)). In general, mutation of the residues within the PPII helix that were not selected for through the molecular evolution of p007 were less significant to DNA binding than mutation of selected residues, as would be expected.

**Contributions of turn residues**

Residues 9 to 13, Gly-Asp-Asp-Ala-Pro, all preserved from the aPP scaffold, compose the link between the PPII and alpha helices. These residues all contribute very modestly to DNA-binding affinity, as might be expected since this portion of the peptide
does not significantly influence the structure of the miniature protein nor does it contact the DNA. The carbonyl oxygen of the final turn residue, Pro13, hydrogen bonds to the amide hydrogens of Asp16 and Leu17 in aPP. Mutation of Pro13 to alanine in p007 results in a modest change in dissociation constant of $20.0 \pm 2.6$ nM and $\Delta \Delta G_{bsCRE}$ of $1.5$ kcal•mol$^{-1}$. However, mutation of Pro13 to sarcosine had a smaller effect ($K_d = 12.0 \pm 1.0$ nM and $\Delta \Delta G_{bsCRE} = 1.2$ kcal•mol$^{-1}$). At the start of the turn, the carbonyl oxygen of Gly9 forms a hydrogen bond to the amide hydrogen of Ala12 within the turn of aPP; an NOE is observed between Gly9 and Ala12 in p007. Mutation of Gly9 in p007 decreased binding by $1.3$ kcal•mol$^{-1}$ ($K_d = 14.8 \pm 1.3$ nM). Substitution of Asp11 with alanine also had a modest effect of $1.4$ kcal•mol$^{-1}$ ($K_d = 17.1 \pm 5.1$ nM). In aPP, Asp10 makes an ionic contact with the Tyr7 hydroxyl. However, this position is the selected residue Met7 in p007, and no NOE is observed between Asp10 and Met7. Mutation of Asp10 to alanine had virtually no effect, with $K_d = 2.3 \pm 0.5$ nM and $\Delta \Delta G_{bsCRE} = 0.2$ kcal•mol$^{-1}$. Overall, the turn residues contribute little to DNA-binding affinity.

**Contributions of other alpha helix residues**

Five residues in the alpha helix of p007 were not previously classified as folding or DNA-binding residues. These include residues at the beginning of the alpha helix retained from aPP (Val14, Glu15, and Asp16) and residues incorporated from GCN4 (Lys18 and Arg36). These residues had widely variable contributions to DNA-binding affinity, from 0.0 to 3.0 kcal•mol$^{-1}$. In the crystal structures of GCN4 bound to DNA, Lys231 does not directly interact with DNA. However, this position is conserved in all bZIP proteins. Substitution of the corresponding Lys18 in p007 with alanine results in a
large decrease in binding affinity ($K_d = 239 \pm 39$ nM and $\Delta \Delta G_{\text{hsCRE}} = 3.0$ kcal•mol$^{-1}$). Arg36 is similarly derived from GCN4. In the crystal structure of the GCN4•AP1 complex, though not the GCN4•CRE complex, the corresponding residue (Arg249) contacts phosphates flanking the binding site. Substitution of Arg36 with alanine results in a significant change in dissociation constant of $85.0 \pm 14.4$ nM and a $\Delta \Delta G_{\text{hsCRE}}$ of 2.3 kcal•mol$^{-1}$. Mutation of Val14, the start of the alpha helix in aPP and p007, also had a large effect ($K_d = 219 \pm 32$ nM and $\Delta \Delta G_{\text{hsCRE}} = 2.9$ kcal•mol$^{-1}$). In contrast, mutation of Glu15 to alanine had little effect on binding ($K_d = 1.8 \pm 0.2$ nM and $\Delta \Delta G_{\text{hsCRE}} = 0.1$ kcal•mol$^{-1}$). Mutation of Asp16 to alanine also had deleterious effect on binding, and it is the only mutant with perhaps a slight enhancement in binding ($K_d = 1.5 \pm 0.1$ nM and $\Delta \Delta G_{\text{hsCRE}} = 0.0$ kcal•mol$^{-1}$). Three of the five alpha helix residues predicted not to be involved in structure or function, Val14, Lys18, and Arg36, contributed to binding.

Quantitative analysis of the contributions of individual side chains to DNA specificity

In addition to possessing high affinity for hsCRE, p007 is distinguished by its exceptionally high specificity for DNA containing the sequence ATGAC. P007 binds the wild type DNA sequence 430- to 1380-fold stronger than analogous DNA sequences containing only two mismatches (5). Furthermore, p007 displayed exceptional preference for hsCRE over nonspecific calf thymus DNA (CT DNA) (> 4000-fold higher affinity) (5). In addition to the general binding affinity conferred by each residue within p007, it is important to score their contributions to specificity. To determine the extent to which each residue within p007 contributes to specificity, the relative affinities of each p007 variant for CT DNA over hsCRE was quantified in a competitive binding assay.
Radiolabeled hsCRE$_{25}$ (< 50 pM) was present in subsaturating concentrations and the concentration of p007 analogue was varied between different alanine mutants such that 50-80% of hsCRE$_{25}$ was bound. To this pre-equilibrated mixture, CT DNA was added such that the concentration of 5 base pair competing sites ranged from 9.77 nM to 2.56 mM. After equilibration, free and peptide-bound hsCRE$_{25}$ were separated by nondenaturing polyacrylamide gel electrophoresis, the proportion of hsCRE$_{25}$ in complex with a p007 variant was monitored by phosphorimaging, and the dissociation constant for nonspecific DNA ($K_{d,ns}$) was calculated (Figure 2.7). Using this data, $K_{rel}$ values, defined as $K_{d,ns}/K_d$, were obtained (Table 2.2). $\Delta G_{spec}$ values could then be calculated (Figure 2.6c).

**Contributions of residues identified by functional selection**

Once again, the residues that were incorporated into p007 through molecular evolution are of particular interest in this study: Gly2, Arg4, and Met7. Each of these positions contributes to binding affinity and these experiments can correlate these binding energetics to specificity of binding. In contrast to the consistent deleterious effect each of alanine mutations have on binding affinity, the specificity of these p007 analogues for hsCRE over CT DNA varies depending on the position of the mutation. Two of the variants, p007$^{G2A}$ and p007$^{R4A}$, were promiscuous with binding to specific and nonspecific DNA, while p007$^{M7A}$ was equally as selective as p007. P007$^{R4A}$ was completely nonspecific, with $K_{rel} = 0.1$; p007$^{G2A}$ was slightly more specific, with $K_{rel} = 129$. In contrast, Met7 was the only residue that did not contribute to specificity. Of all the variants, the selectivity of p007$^{M7A}$ is most similar to that of p007 ($K_{rel} = 2429$ and 4169,
Figure 2.7 Plots illustrating the relative affinity of p007 variants for γ-[32P]-hsCRE25 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 assuming that every base on either DNA strand represents the start of a competitor site. C0 is multiplied by the factor indicated in each plot.
Table 2.2 Affinities of peptides for CT DNA at 25 °C. The value for p007 was determined previously.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>$K_{d,ns}$ (nM)</th>
<th>$K_{rel}$ ($K_{d,ns}/K_d$)</th>
<th>$\Delta G_{CT DNA}$ (kcal•mol$^{-1}$)</th>
<th>$\Delta\Delta G_{spec}$ (kcal•mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p007</td>
<td>6670 ± 1200</td>
<td>4169</td>
<td>-7.0</td>
<td>-4.9</td>
</tr>
<tr>
<td>p007$^{G1A}$</td>
<td>13,400 ± 2900</td>
<td>724</td>
<td>-6.6</td>
<td>-3.9</td>
</tr>
<tr>
<td>p007$^{G2A}$</td>
<td>11,700 ± 2300</td>
<td>129</td>
<td>-6.7</td>
<td>-2.9</td>
</tr>
<tr>
<td>p007$^{S3A}$</td>
<td>10,400 ± 1400</td>
<td>630</td>
<td>-6.8</td>
<td>-3.8</td>
</tr>
<tr>
<td>p007$^{R4A}$</td>
<td>44.0 ± 2.3</td>
<td>0.1</td>
<td>-10.0</td>
<td>1.3</td>
</tr>
<tr>
<td>p007$^{T6A}$</td>
<td>13,700 ± 3100</td>
<td>349</td>
<td>-6.6</td>
<td>-3.5</td>
</tr>
<tr>
<td>p007$^{M7A}$</td>
<td>23,000 ± 3100</td>
<td>2429</td>
<td>-6.3</td>
<td>-4.6</td>
</tr>
<tr>
<td>p007$^{P8A}$</td>
<td>14,900 ± 2300</td>
<td>164</td>
<td>-6.6</td>
<td>-3.0</td>
</tr>
<tr>
<td>p007$^{P8Z}$</td>
<td>15,000 ± 2600</td>
<td>478</td>
<td>-6.6</td>
<td>-3.6</td>
</tr>
<tr>
<td>p007$^{G9A}$</td>
<td>12,100 ± 1100</td>
<td>818</td>
<td>-6.7</td>
<td>-4.0</td>
</tr>
<tr>
<td>p007$^{D10A}$</td>
<td>403 ± 54</td>
<td>174</td>
<td>-8.7</td>
<td>-3.0</td>
</tr>
<tr>
<td>p007$^{D11A}$</td>
<td>4160 ± 330</td>
<td>243</td>
<td>-7.3</td>
<td>-3.2</td>
</tr>
<tr>
<td>p007$^{P13A}$</td>
<td>10,900 ± 1100</td>
<td>545</td>
<td>-6.7</td>
<td>-3.7</td>
</tr>
<tr>
<td>p007$^{P13Z}$</td>
<td>6000 ± 690</td>
<td>500</td>
<td>-7.1</td>
<td>-3.7</td>
</tr>
<tr>
<td>p007$^{V14A}$</td>
<td>22,500 ± 3500</td>
<td>103</td>
<td>-6.3</td>
<td>-2.7</td>
</tr>
<tr>
<td>p007$^{E15A}$</td>
<td>744 ± 1112</td>
<td>418</td>
<td>-8.3</td>
<td>-3.6</td>
</tr>
<tr>
<td>p007$^{D16A}$</td>
<td>647 ± 43</td>
<td>428</td>
<td>-8.4</td>
<td>-3.6</td>
</tr>
<tr>
<td>p007$^{L17A}$</td>
<td>18,200 ± 3000</td>
<td>108</td>
<td>-6.4</td>
<td>-2.8</td>
</tr>
<tr>
<td>p007$^{K18A}$</td>
<td>33.7 ± 7.7</td>
<td>0.1</td>
<td>-10.2</td>
<td>1.3</td>
</tr>
<tr>
<td>p007$^{R19A}$</td>
<td>72.7 ± 16.8</td>
<td>0.1</td>
<td>-9.7</td>
<td>1.3</td>
</tr>
<tr>
<td>p007$^{F20A}$</td>
<td>15,300 ± 4800</td>
<td>164</td>
<td>-6.5</td>
<td>-3.0</td>
</tr>
<tr>
<td>p007$^{R21A}$</td>
<td>40.3 ± 6.8</td>
<td>0.2</td>
<td>-10.0</td>
<td>1.1</td>
</tr>
<tr>
<td>p007$^{N22A}$</td>
<td>91.2 ± 17.4</td>
<td>0.2</td>
<td>-9.6</td>
<td>1.1</td>
</tr>
<tr>
<td>p007$^{T23A}$</td>
<td>7280 ± 340</td>
<td>806</td>
<td>-7.0</td>
<td>-3.9</td>
</tr>
<tr>
<td>p007$^{L24A}$</td>
<td>6540 ± 900</td>
<td>52</td>
<td>-7.0</td>
<td>-2.3</td>
</tr>
<tr>
<td>p007$^{R27A}$</td>
<td>160 ± 11</td>
<td>0.3</td>
<td>-9.2</td>
<td>0.7</td>
</tr>
<tr>
<td>p007$^{R28A}$</td>
<td>81.5 ± 21.3</td>
<td>0.1</td>
<td>-9.6</td>
<td>1.6</td>
</tr>
<tr>
<td>p007$^{S29A}$</td>
<td>27,800 ± 2500</td>
<td>321</td>
<td>-6.2</td>
<td>-3.4</td>
</tr>
<tr>
<td>p007$^{R30A}$</td>
<td>184 ± 13</td>
<td>0.3</td>
<td>-9.2</td>
<td>0.8</td>
</tr>
<tr>
<td>p007$^{R32A}$</td>
<td>26.8 ± 4.3</td>
<td>0.2</td>
<td>-10.3</td>
<td>0.9</td>
</tr>
<tr>
<td>p007$^{K33A}$</td>
<td>223 ± 47</td>
<td>0.4</td>
<td>-9.0</td>
<td>0.6</td>
</tr>
<tr>
<td>p007$^{R36A}$</td>
<td>26.4 ± 7.5</td>
<td>0.3</td>
<td>-10.3</td>
<td>0.7</td>
</tr>
</tbody>
</table>

$\Delta G_{CT DNA} = -RT\ln K_{d,ns}$

$K_{rel} = K_{d,ns}/K_d$

$\Delta\Delta G_{spec} = \Delta G_{hsCRE} - \Delta G_{CT DNA}$
respectively).

**Contributions of putative DNA-binding residues**

Residues that are designed to make direct DNA contacts (Arg19, Arg21, Asn22, Thr23, Arg27, Arg28, Ser29, Arg30, Arg32, and Lys33) would be expected to contribute greatly to binding specificity of p007. Indeed, with the exception of Thr23 and Ser29, each DNA-binding residue contributes a great deal to specificity. Based on analogous binding studies using GCN4, one would predict that mutation of any of the putative base-contacting residues, Asn22, Ser29, or Arg30, would diminish specificity the most. In fact, as shown through a gel shift analysis, mutating Asn235 of the bZIP element of GCN4 to any other amino acid results in weak or no binding to CRE and abolishes selectivity (23). In a parallel study, mutation of the other invariant residue, Arg243 in GCN4, to all other amino acids also results in weak or no binding (23). The corresponding variants, p007N22A and p007R30A had a very low $K_{rel}$ of 0.2 and 0.3, respectively. While mutation of the of the final base-contacting residue Ser29 to alanine only modestly affected binding hsCRE$_{25}$ relative to the other DNA-binding variants, p007S29A was also more specific than the other DNA-binding variants ($K_{rel} = 321$). Any single mutation of the basic residues Arg19, Arg21, Arg27, Arg28, Arg32, and Lys33 to alanine abolished specificity ($K_{rel} = 0.1$ to 0.4). Thr236 in GCN4 is not explicitly a DNA-binding residue, but positions the other binding residues as seen in the crystal structure of the GCN4 homodimer bound DNA (20). Mutation of the corresponding Thr23 in p007 to alanine mildly reduced specificity with $K_{rel} = 806$. 


Contributions of putative folding residues

Because p007 is a miniature protein that is very dependent on the packing of hydrophobic residues from the alpha helix onto the PPII helix, it is reasonable to assume that these packing residues could impact the presentation of binding residues and the protein’s binding selectivity. The p007 variants corresponding to Pro8, Leu17, Phe20, and Leu24 display varying degrees of specificity ($K_{rel}$ values ranging from 52 to 478). While the folding residues do not directly contact DNA, they still affect DNA-binding affinity. Similarly, we hypothesize that folding residues may indirectly contribute to DNA-binding specificity. P007L24A was only modestly specific, with $K_{rel} = 52$. P007L17A and p007F20A were somewhat more specific, with $K_{rel}$s of 108 and 164, respectively. The alanine substitution of Pro8 ($K_{rel} = 164$) was more deleterious to specificity than the sarcosine substitution ($K_{rel} = 478$).

Contributions of other residues within the type II polyproline helix

The residues with no defined function in p007 are considered next, beginning with the PPII helix residues Gly1, Ser3, and Thr6. These residues contribute modestly to specificity. P007T6A had modest specificity ($K_{rel} = 349$); p007G1A and p007S3A were slightly more specific ($K_{rel}$s of 724 and 630, respectively). In the PPII helix of p007, substitution of these residues with alanine proved to be less costly in terms of specificity than substitution of the folding residues with alanine, which is as one would expect.

Contributions of turn residues

The subset of residues that compose a turn element between the alpha and PPII
helices would not be expected to play a large role in p007 binding specificity; however, to be comprehensive, these positions were also examined. Surprisingly, the turn residues exhibited moderate levels of specificities. Although p007\textsuperscript{D10A} bound to DNA with near wild type affinity, it was not nearly as specific as p007 (K\textsubscript{rel} = 174). P007\textsuperscript{D11A} was slightly more specific, with K\textsubscript{rel} = 243. P007\textsuperscript{P13A} (K\textsubscript{rel} = 545) and p007\textsuperscript{P13Z} (K\textsubscript{rel} = 500) were both moderately specific. Though sarcosine substitution of Pro13 was less harmful to DNA-binding affinity than alanine substitution, we observe that there is no difference between p007\textsuperscript{P13A} and p007\textsuperscript{P13Z}, with regard to specificity. Gly9 contributes the least of the turn residues. P007\textsuperscript{G9A} remained specific, with K\textsubscript{rel} = 818.

**Contributions of other alpha helix residues**

The remaining residues on p007 (Val14, Glu15, Asp16, Lys18, and Arg36) reside in its alpha helix and are not predicted to make contact to the DNA target, nor contribute to the hydrophobic contacts to the PPII helix. Contrary to expectations, p007\textsuperscript{K18A} and p007\textsuperscript{R36A} were completely nonspecific (K\textsubscript{rel} = 0.1 and 0.3, respectively). However, considering the unexpected contribution these positions play in DNA-binding affinity, it can be rationalized that they may be involved in specificity as well. Mutation of Val14 had a significant impact on specificity, with K\textsubscript{rel} = 103. P007\textsuperscript{E15A} and p007\textsuperscript{D16A} had high affinity for DNA and proved to be specific as well, with K\textsubscript{rel}s = 418 and 428, respectively.

**Contributions of individual side chains to protein structure as measured by circular dichroism**

Alanine substitution of key residues in p007 may severely affect its binding
affinity or selectivity. The cause of these effects, however, could be due to misfolding of the p007 variant, and not necessarily immediate interactions between the miniature protein and the target DNA. The set of p007 variants also allows us to monitor the effect of individual side chains on protein secondary structure. To this end, each variant was characterized by circular dichroism (CD) at a concentration of 10 µM in PBS at 4 °C in the absence of DNA (Figure 2.8). The mean residue ellipticity of the wild type p007 at 222 nm (MRE$_{222}$) is $-18,000 \text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$. The relatively minor mutation of single amino acids could have a significant effect on protein folding. For example, PPBR4 is only different by four residues compared to p007 (in addition to the Gly-Gly-Cys at the C-terminus) and is unfolded in the absence of DNA. These four mutations in the PPII helix p007 result in an MRE$_{222}$ of $-8000 \text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$. All variants displayed less alpha helical content by CD than p007 with MRE$_{222}$ between $-6214$ and $-16,537 \text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$ (Table 2.3).

In the PPII helix of p007, the folding residues and selected residues contributed the most to alpha helicity. The folding variant p007$^{P8A}$ had an MRE$_{222}$ of only $-7044 \text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$, and the sarcosine variant had an MRE$_{222}$ of $-9674 \text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$. The selected residues also contributed a great deal, with the MRE$_{222}$ of the corresponding variants p007$^{G2A} = -8841$, p007$^{R4A} = -9919$, and p007$^{M7A} = -7327 \text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$. In contrast, the remaining PPII residues did not contribute as much to helicity. Variants generally had higher MRE$_{222}$ (p007$^{G1A} = -10,435$, p007$^{S3A} = -8787$, and p007$^{T6A} = -12,397 \text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$).

Though the turn residues 9 to 13 displayed modest contributions to affinity and specificity, they exhibited widely different contributions to alpha helicity. Asp11
Figure 2.8  CD spectra of p007 variants at 10 µM in PBS at 4 °C.
Table 2.3 Mean residue ellipticities at 222 nm of peptides at 4 °C.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>MRE$_{222}$ (deg cm$^2$ dmol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p007</td>
<td>−18,000</td>
</tr>
<tr>
<td>p007$^{G1A}$</td>
<td>−10,435</td>
</tr>
<tr>
<td>p007$^{G2A}$</td>
<td>−8841</td>
</tr>
<tr>
<td>p007$^{S3A}$</td>
<td>−8787</td>
</tr>
<tr>
<td>p007$^{R4A}$</td>
<td>−9919</td>
</tr>
<tr>
<td>p007$^{T6A}$</td>
<td>−12,397</td>
</tr>
<tr>
<td>p007$^{M7A}$</td>
<td>−7327</td>
</tr>
<tr>
<td>p007$^{P8A}$</td>
<td>−7044</td>
</tr>
<tr>
<td>p007$^{P8Z}$</td>
<td>−9674</td>
</tr>
<tr>
<td>p007$^{G9A}$</td>
<td>−9151</td>
</tr>
<tr>
<td>p007$^{D10A}$</td>
<td>−10,855</td>
</tr>
<tr>
<td>p007$^{D11A}$</td>
<td>−6214</td>
</tr>
<tr>
<td>p007$^{P13A}$</td>
<td>−16,537</td>
</tr>
<tr>
<td>p007$^{P13Z}$</td>
<td>−9598</td>
</tr>
<tr>
<td>p007$^{V14A}$</td>
<td>−9805</td>
</tr>
<tr>
<td>p007$^{E15A}$</td>
<td>−7149</td>
</tr>
<tr>
<td>p007$^{D16A}$</td>
<td>−9351</td>
</tr>
<tr>
<td>p007$^{L17A}$</td>
<td>−7975</td>
</tr>
<tr>
<td>p007$^{K18A}$</td>
<td>−8955</td>
</tr>
<tr>
<td>p007$^{R19A}$</td>
<td>−8402</td>
</tr>
<tr>
<td>p007$^{F20A}$</td>
<td>−8110</td>
</tr>
<tr>
<td>p007$^{R21A}$</td>
<td>−8984</td>
</tr>
<tr>
<td>p007$^{N22A}$</td>
<td>−11,329</td>
</tr>
<tr>
<td>p007$^{T23A}$</td>
<td>−10,556</td>
</tr>
<tr>
<td>p007$^{L24A}$</td>
<td>−8067</td>
</tr>
<tr>
<td>p007$^{R27A}$</td>
<td>−10,112</td>
</tr>
<tr>
<td>p007$^{R28A}$</td>
<td>−9004</td>
</tr>
<tr>
<td>p007$^{S29A}$</td>
<td>−11,361</td>
</tr>
<tr>
<td>p007$^{R30A}$</td>
<td>−10,699</td>
</tr>
<tr>
<td>p007$^{R32A}$</td>
<td>−9362</td>
</tr>
<tr>
<td>p007$^{K33A}$</td>
<td>−12,214</td>
</tr>
<tr>
<td>p007$^{E36A}$</td>
<td>−8759</td>
</tr>
</tbody>
</table>
contributed the most to alpha helicity, with the MRE$_{222}$ of p007$^{D11A}$ equal to only –6214 deg•cm$^2$•dmol$^{-1}$. Pro13 contributed the least to alpha helicity of all residues in p007, with the MRE$_{222}$ of p007$^{P13A}$ close to that of p007 (–16,537 deg•cm$^2$•dmol$^{-1}$). However, when Pro13 is substituted with sarcosine, p007$^{P13Z}$ had a lower alpha helical content (MRE$_{222}$ = –9598 deg•cm$^2$•dmol$^{-1}$). The remaining turn variants, p007$^{G9A}$ and p007$^{D10A}$, had similar helicities with MRE$_{222}$ of –9151 and –10,855 deg•cm$^2$•dmol$^{-1}$, respectively.

Substitution of any residue in the alpha helix with alanine did not increase alpha helicity. In fact, variants in which residues comprising the hydrophobic core are replaced with alanine exhibited the lowest helicity of all alpha helix variants, with the MRE$_{222}$ of p007$^{L17A}$, p007$^{F20A}$, and p007$^{L24A}$ only ranging from –7975 to –8110 deg•cm$^2$•dmol$^{-1}$. Undefined alpha helix residues (Val14, Glu15, Asp16, Lys18, and Arg36) also contributed to helicity, with the MRE$_{222}$ of the alanine variants ranging from –7149 to –8955 deg•cm$^2$•dmol$^{-1}$. Variants in which DNA-binding residues were replaced with alanine had a lesser effect, with the MRE$_{222}$ ranging from –8402 to –12,214 deg•cm$^2$•dmol$^{-1}$.

**Discussion**

bZIP proteins are perhaps the best studied eukaryotic transcription factors. Since their designation as a unique family in 1988 (25), seven high resolution X-ray crystal structures of bZIPs in complex with DNA have been published (Figure 2.9a) (19,20,26-30) and numerous papers have documented the role of conserved amino acids in mediating specific DNA interactions. This wealth of data provides an excellent context
in which to analyze the thirty-one p007 variants described in this work in terms of both binding mode and the relative contributions of side chains to overall binding affinity.

*P007 binds DNA much like GCN4*

The DNA-binding residues of p007 derived from GCN4 significantly contribute to DNA-binding affinity. In particular, the invariant residues, Asn22 and Arg30, are major contributors to binding affinity and selectivity. Substitution of Asn22 and Arg30 with alanine greatly reduce the resulting miniature protein’s affinity for hsCRE by 3.5 and 3.6 kcal•mol⁻¹, respectively. This is consistent with mutagenesis studies of the bZIP element of GCN4, in which substitution of Asn235 or Arg243 with all other amino acids diminishes or eliminates binding to the CRE site in gel shift assays (23). Mutation of the remaining base-contacting residue, Ser29 in p007, to alanine reduces binding affinity for hsCRE by 2.4 kcal•mol⁻¹. Similarly, mutation of Ser242 to alanine abolished the ability of GCN4 to bind hsCRE (31). Though this mutation does not have as large an effect as mutation of the invariant residues, Asn22 and Arg30, Ser29 is also important for DNA-binding affinity and highlights how successful p007 is at recapitulating the GCN4 binding epitope.

Each of the phosphate-contacting residues of p007, Arg19, Arg21, Arg27, Arg28, and Arg32, contributed a great deal to DNA-binding affinity, with ΔΔG₉₅CRE of the corresponding variants ≥ 3.0 kcal•mol⁻¹. Though nonspecific, the electrostatic interactions of the basic residues with the phosphate backbone are as important for binding affinity as the base-contacting residues. In mimics of the GCN4 bZIP element that replace other residues of the basic region with alanine, at least four of the five base-
Figure 2.9 Basic regions of representative bZIPs. **A.** Amino acid sequences and DNA-binding sites for proteins in the bZIP subfamilies for which crystal structures have been determined. Base-contacting residues are shaded. The invariant Asn and Arg are depicted in red and conserved basic residues in blue. **B.** Superposition of p007 and the basic region of GCN4 bound to hsCRE. The side chains of all basic and DNA-binding residues are shown.
<table>
<thead>
<tr>
<th>Family</th>
<th>Protein</th>
<th>Basic Region</th>
<th>Binding Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP-1</td>
<td>c-Fos</td>
<td>EKRRIRRRERNMAAAKCRNRRRELTDTL</td>
<td>ATGACTCAT</td>
</tr>
<tr>
<td></td>
<td>c-Jun</td>
<td>IRAEKRMRRRIAMASKCRKKLERIARLR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCN4</td>
<td>DPAALKRARNTAARRSRARKLRQREQL</td>
<td>ATGACGTAC</td>
</tr>
<tr>
<td>CREB/ATF</td>
<td>GCN4</td>
<td>DPAALKRARNTAARRSRARKLRQREQL</td>
<td>ATGACGTAC</td>
</tr>
<tr>
<td></td>
<td>CREB</td>
<td>RKREVRLMKREAAARSCRKKEYVKS</td>
<td></td>
</tr>
<tr>
<td>C/EBP</td>
<td>C/EBPα</td>
<td>NEYVRKRRERRNNIARVRKRDRKARQRNVET</td>
<td>ATTGCAGCAAT</td>
</tr>
<tr>
<td>PAP</td>
<td>PAP1</td>
<td>EPSSKRAQNRAQRAPRKRKEDHLKAL</td>
<td>GTTACGTAAC</td>
</tr>
</tbody>
</table>

B
contacting residues are sufficient for maintaining binding affinity. However, as these residues are mutated to alanines, a similar decline in binding affinity is observed as with the alanine screening described in this work (32).

Mutation of the remaining DNA-binding residue, Thr23 in p007, results in a decrease in binding affinity of only 1.0 kcal•mol\(^{-1}\). The corresponding DNA-binding residue in GCN4 is Thr236, which makes a phosphate contact and hydrophobic contact with the deoxyribose of the central guanosine, and orients residues Arg232, Glu237, and Arg243 in the crystal structure of GCN4 bound to CRE (20). Perhaps in the context of p007, which binds the half-site DNA sequence, Thr23 no longer has the same functional role. The alpha helix of p007 may present DNA-binding residues in a fashion compensating for the organizational role of Thr236 in GCN4. These results provide evidence that the tertiary structure of p007 is optimized to allow the alpha helix to adapt a similar binding mode as GCN4.

The panel of variants also allowed us to evaluate the relationship between DNA affinity and folding, both in the presence and absence of DNA. We predict that the folding residues would play a crucial, if indirect, role in DNA-binding affinity. In fact, the binding affinities of the p007 variants containing alanine in place of folding residues (Pro8, Leu17, Phe20, and Leu24) diminished binding to DNA relative to p007 with \(\Delta\Delta G_{\text{hsCRE}} = 2.4\) to 2.7 kcal•mol\(^{-1}\). This indicates disruption of the aPP-like fold of p007. In contrast to GCN4, which uses extended alpha helices to bind each half of the CRE site, p007 is not stabilized by a dimerization domain; rather it is stabilized by the folding of a PPII helix against the alpha helix. The effect of mutating a ‘folding’ residue is as significant as mutating one of the DNA-binding residues, demonstrating that the stability
provided by the folding residues contributes to DNA-binding affinity to a great extent. Also, although the folding residues do not play a direct role in DNA binding, they were still able to impact specificity. The $K_{\text{mS}}$ of the corresponding alanine variants were only about 100.

The importance of folding is further emphasized by the fact that PPBR4, the predecessor of p007 that differs only in the PPII helix, binds more poorly than p007 because of its lack of structure in the absence of DNA. This deficiency was corrected when four residues in the PPII helix were identified in a phage display selection and evolved to stabilize the miniature protein. Because p007 was now folded in the absence of DNA and its binding affinity improved 1000-fold, we assumed that the four selected residues were important for folding, and that this prestructuring is responsible for the enhanced binding affinity. Interestingly, substitution of selected residues Gly2, Arg4, and Met7 with alanine had widely different effects. Mutation of Arg4 to alanine diminished binding affinity equivalent to that of mutation of the basic residues in the alpha helix. A model of the p007 backbone superimposed on the GCN4 backbone in the GCN4•CRE complex suggests that it is possible for Arg4 to contact a phosphate in the DNA (Figure 2.9b). This suggests that in addition to, or in lieu of, stabilizing the fold of p007, a new DNA-binding contact was established. Substitution of Gly2 with alanine resulted in a decrease in binding affinity comparable to that of variants in which the folding residues were substituted with alanine. At the other end of the spectrum, Met7 contributed only 1.0 kcal•mol$^{-1}$ to binding affinity. Nevertheless, all the selected residues contributed to DNA-binding affinity through stabilizing the folding of p007, confirming that the concerted evolution of PPBR4 into a highly functional and well-
structured molecule was successful.

Residues of p007 not predicted to be involved in DNA binding, folding, or selection (Gly1, Ser3, and Thr6) were also important for DNA-binding affinity, contrary to what one might anticipate. Because many well-folded pancreatic polypeptides from other organisms have alanine at position 1, one would have predicted that variant p007$_{G1A}$ would behave similarly to p007. In fact, mutation of Gly1 to alanine results in a modest decrease in DNA-binding affinity of 1.4 kcal•mol$^{-1}$. P007$_{S3A}$ and p007$_{T6A}$ exhibit similar affinities of 1.4 and 1.9 kcal•mol$^{-1}$, respectively. Overall, these residues are not as important as the selected residues in the PPII helix, but exhibit an unexpected effect on DNA binding.

Three residues in the turn of p007 (Gly9, Asp11, and Pro13) contribute only modestly to binding affinity. Positions 9 and 12 are absolutely conserved in PP fold polypeptides, likely due to a hydrogen bonding interaction between the Gly9 carbonyl oxygen and the amide proton of Ala12, as observed in the crystal structure of aPP (12). Substitution of Gly9 with alanine diminishes the ability of p007 to bind DNA, though it is not immediately clear why this would be. Perhaps substitution with alanine disrupts the flexibility in the backbone in p007 necessary to form this bond, and consequently disrupts the ability of p007 to fold and bind DNA. Substitution of Asp11 or Pro13 with alanine diminishes DNA binding to the same extent as the substitution of the nonfolding PPII residues, highlighting the fact that these residues are not essential to binding as would be expected.

Three residues in the alpha helix of p007 yielded unexpected results suggesting that they also contributed to DNA-binding affinity despite the fact that they were not
predicted to have a role in folding or in DNA binding: Val14, Lys18, and Arg36. Val14 initiates the alpha helix of p007, and significantly contributes to the binding energy of p007 ($\Delta \Delta G_{hsCRE} = 2.9 \text{ kcal}\text{mol}^{-1}$). Lys18 and Arg36, are basic residues flanking DNA-binding residues in the alpha helix, and they too are important for binding affinity. Lys18 contributes 3.0 kcal•mol$^{-1}$ to binding. Lys18 was incorporated into p007 as an analogue to Lys231 in GCN4, which is conserved in all bZIP proteins. This implies that this residue may have an important influence on DNA binding by making nonspecific contacts to the DNA phosphate backbone. This is consistent with Arg249 in GCN4 (Arg36 in p007), which can make phosphate contacts in the pseudosymmetric AP-1 site, but not the fully symmetric CRE site. It is possible that since p007 lacks a dimerization domain, these basic residues are sufficiently flexible to make phosphate contacts with hsCRE, and thus these results may indicate that p007 forms a complex with hsCRE that is more like the GCN4•AP-1 complex than the GCN4•CRE complex.

Three residues did not contribute to DNA-binding affinity: Asp10, Glu15, and Asp16. Asp10 is a part of the turn, which in general, was tolerant of substitution, and since it neither makes contact with the DNA nor contributes to the hydrophobic packing of p007, it is no surprise that this residue plays no significant role in DNA binding. It is slightly surprising that Glu15 and Asp16 do not contribute to the p007’s affinity for DNA. These two residues have ionic interactions with Arg19 in aPP (12) and in the NMR structure of p007, Asp16 contacts Arg19. These contacts seem to imply that Glu15 and Asp16 potentially prestructure p007 and could help in positioning Arg19 for DNA binding. However, in the presence of DNA, Arg19 is thought to be a DNA-binding residue, which could be exclusive of the intramolecular binding with Glu15 and Asp16.
Because p007\textsuperscript{E15A} and p007\textsuperscript{D16A} display little or no change in binding affinity compared to p007, these intramolecular contacts may be irrelevant in the DNA-bound complex.

**Residues throughout p007 contribute to specificity**

Any number of small molecules can display high DNA affinity through nonspecific interactions with DNA’s phosphate backbone; however, what makes p007 so remarkable is its selectivity for the hsCRE sequence. Luckily, parallel studies with GCN4 variants also permit comparison of the relative contribution of residues to specificity between p007 and GCN4 – in particular the ability to differentiate the sequence ATGAC from the large number of sequences found in CT DNA. GCN4 forms a tighter complex with CRE DNA than CT DNA by 3.3 kcal\textperiodcentered mol\textsuperscript{-1} (33). This translates into a 100 to 1000 fold preference for the CRE DNA binding site over nonspecific DNA (34). Remarkably, p007 displays even greater selectivity, binding CT DNA 4.9 kcal\textperiodcentered mol\textsuperscript{-1} worse than hsCRE DNA (K\textsubscript{rel} = 4169). Furthermore, p007 appears to be a highly optimized ligand for hsCRE, since no p007 variant retained this level of specificity. We would predict that the DNA-binding residues of p007 would contribute the most to specificity, in particular, the base-contacting residues as seen with Asn22 and Arg30 for which alanine replacement abolishes specificity with K\textsubscript{rel} = 0.2 and 0.3, respectively. Not surprisingly, our findings mirror that already known in the GCN4 system where the GCN4 mutant N235A displays similarly reduced specificity (23). The other base-contacting residues in p007, Thr23 and Ser29, contributed moderately to specificity, with K\textsubscript{rel}s of the corresponding variants equal to 806 and 321, respectively.

However, it is quite unexpected that when any basic, phosphate-contacting
residue from p007 is substituted with alanine, specificity was abolished. One would certainly expect a loss of general binding affinity, but this loss of specificity is surprising. The values of $K_{rel}$ for the alanine variants of Arg19, Arg21, Arg27, Arg28, and Arg32 were all $< 1$. Interestingly, the three other variants that exhibited $K_{rel} < 1$ (p007$^{R4A}$, p007$^{K18A}$, and p007$^{R36A}$) are also basic residues. Arg4 was incorporated through a phage display selection and was assumed to be primarily important for DNA-binding affinity, but it could also be important for specificity. Lys18, a conserved basic residue in GCN4, was also important for DNA-binding affinity. Arg36 is also a basic residue that unexpectedly contributed to DNA-binding affinity. It is possible that these residues form a phosphate contact and therefore, could contribute to specificity. Electrostatic interactions played a critical role in specificity, as every variant in which a basic residue was substituted with an alanine was completely nonspecific.

The residues that are not involved in DNA binding contribute moderately to specificity, with $K_{rel} = 52$ to 478. Interestingly, the variants that bind as well as p007 (p007$^{D10A}$, p007$^{E15A}$, and p007$^{D16A}$) are all less specific than p007. These variants are an example of electrostatic “negative readout”: the variants can bind hsCRE$_{25}$ with similar affinity, but with decreased specificity.

Met7 makes the smallest contribution to specificity, with a $K_{rel}$ of 2429. It is interesting that p007$^{M7A}$ had only modestly reduced affinity, as well ($\Delta \Delta G_{hsCRE} = 1.0$ kcal•mol$^{-1}$). Perhaps there is more than one solution for the selected residue at this position. Additionally, the energy well for p007 binding to hsCRE could be so deep that numerous, alternate mutations are required to create another aPP-based ligand that binds this sequence. In general, not only the DNA-contacting residues, but residues throughout
the entire sequence contributed to specificity.

While mutagenesis studies can provide data about the roles of individual residues in specificity, the most illustrative information about how DNA-binding proteins bind specifically can be gleaned from the structures of the dimeric lac DNA-binding domain (DBD) bound to either specific DNA or nonspecific DNA (35). In the nonspecific complex, lac DBD makes only electrostatic contacts with DNA and no base specific contacts. However, in the specific complex, there are considerably fewer electrostatic interactions and numerous base specific contacts. This suggests a scanning or surveying mechanism in which the lac DBD can bind DNA nonspecifically, and upon locating target DNA, reorganize the electrostatic interactions to enable the base specific interactions of the final complex. It is likely that p007 owes its high specificity to a similar mechanism of binding. All variants bind with submicromolar affinities, yet their specificities varied widely. In p007 variants which mutate a basic residue, the loss of the phosphate contact and inability to reorganize into the specific complex results in a diminished specificity equal to the loss of a base-specific contact. This would explain why the charge-charge interactions with the phosphate backbone were just as important for specificity. A crystal structure of p007 bound to DNA would help study these observations.

The relationship between affinity and specificity

There is ongoing debate about the relationship between affinity and specificity. On one hand, it is believed that affinity and specificity are inextricably linked. In the lock and key model (36), a ligand is perfectly tailored to its target in terms of shape
complementarity and charge distribution. This results in high affinity, and specificity automatically follows suit. Therefore, ligands that are flexible and can bind a variety of targets have a decreased specificity. In addition, the cost of conformational restriction results in decreased affinity (37). The opposing induced fit model proposes that folding is coupled to binding (38), resulting in increased specificity in molecular recognition. The additional energy required for folding reduces the free energy of complex formation, so high affinity binding only occurs when complementarity is maximal (39).

P007 displays greater affinity and specificity than all of its alanine variants. In this study, we can explore the two theories about the relationship between affinity and specificity with these variants, which exhibit a broad spectrum of these characteristics. In general, there is a correlation between decreasing binding affinity and specificity by the p007 variants (Figure 2.10). This data supports the model that high specificity automatically follows high affinity. However, induced fit also plays a role in recognition. By CD, none of the variants were more structured than p007 in the absence of target DNA. Since p007 folds even further upon binding specific DNA (indicative of induced fit), the variants that bound as well as p007 necessarily must fold to form a similarly stable complex with DNA. Therefore, it is unlikely a lock and key mechanism of binding is occurring. It is more likely that there is initial complementarity, but not a perfect lock and key fit. Some amount of conformational flexibility allows for high affinity and highly specific binding to take place. This is true of p007 and likely, all of the p007 variants that bind with nearly as high affinity and specificity.

The relationship between inherent alpha helicity and DNA binding
Figure 2.10  Plot of $\Delta G_{\text{hsCRE}}$ vs. $\Delta \Delta G_{\text{spec}}$ of p007 and all variants, illustrating the relationship between affinity and specificity.
CD data of GCN4 in the presence and absence of DNA shows that the leucine zipper of GCN4 is fully helical and the basic region exhibits nascent helicity in the absence of DNA; upon binding specific DNA, the basic region becomes fully helical \((40,41)\). Many other DNA-binding proteins observe this induced fit mechanism of binding. Free energy drives the folding of the DNA-binding domain, so the entropic cost of binding DNA with a flexible domain can be reduced by stabilizing nascent helicity in the absence of DNA. When the recognition sequence is found, the native alpha helical conformation is stabilized by interactions with the bases and the phosphodiester backbone. Specific DNA binding of proteins is accompanied by a large heat capacity change \((\Delta C_{\text{assoc}})\), associated with the burial of large amounts of nonpolar surface area. The binding free energy \(\Delta G_{\text{hsCRE}}\) drives the folding upon binding of a specific sequence. This coupled conformational change is characterized by a reduction in entropy, consistent with folding of previously disordered regions of protein or DNA \((42)\).

While p007’s precursor, PPBR4, binds hsCRE DNA well, it was minimally structured in the absence of DNA. P007 had significantly more alpha helical content in the absence of DNA and a greater affinity for specific DNA than PPBR4. This higher affinity for DNA can partly be attributed to stabilization of the alpha helix, reducing the entropic penalty for folding upon binding DNA. This is supported by CD data in which p007 displays an increase in MRE\(_{222}\) from \(-18,000\) to \(-28,000\) upon the addition of specific DNA, corresponding to an ordering of 19 residues in the alpha helix \((5)\). These results provide evidence that p007 folds when bound to specific DNA. Mutation of any of the folding residues Pro8, Leu17, Phe20, and Leu24 to alanine disrupts folding. Variants that differ at these residues are less structured than p007 in the absence of DNA,
and presumably, they are unable to reach the optimally folded state in the presence of DNA or must pay a higher energetic cost to reach this folded state. This is manifested in their lowered affinity for hsCRE$_{25}$ relative to p007.

However, variants as minimally structured as the folding variants could bind DNA with high affinity. For example, the least structured variant p007$^{\text{D11A}}$ (–6214 deg•cm$^2$•dmol$^{-1}$) binds DNA with $K_d$ only 10-fold greater than that of p007. Yet, the most structured variant p007$^{\text{P13A}}$ (–16,537 deg•cm$^2$•dmol$^{-1}$) binds with the same affinity. The extent of alpha helicity in the absence of DNA does not absolutely correlate with affinity (Figure 2.11). Furthermore, the three variants that bound DNA as well as p007 (p007$^{\text{D10A}}$, p007$^{\text{E15A}}$, and p007$^{\text{D16A}}$) are all less alpha helical and display varying degrees of structure relative to each other (MRE$_{222}$ = –10,855, –7149, and –9351, respectively). Again, structure in the absence of DNA does not correlate with function.

One might hypothesize that the loss of one of the alpha helix residues could be compensated by benefit of preorganization due to substitution with the amino acid with the highest alpha helical propensity, alanine. Yet, this does not seem to be the case, as judged by of MRE$_{222}$ of the p007 alanine variants. All these variants are less structured than p007. However, it appears that in variants that bind specific DNA well, it costs little additional free energy to fold and then bind. There may be a threshold of nascent helicity that is enough for specific DNA binding to occur in addition to the ability to adopt a final stable conformation in complex with DNA. Furthermore, the threshold level of preorganization that is necessary to bind DNA with high affinity cannot be accurately predicted simply by CD measurements of alpha helicity.

Preorganization is also important in terms of specificity. The alpha helix content
**Figure 2.11** Plot of $\Delta G_{\text{hsCRE}}$ vs. $\text{MRE}_{222}$ of p007 and all variants, illustrating the relationship between affinity and inherent alpha helicity.
of the GCN4 basic region, largely disordered in the absence of DNA, increases upon the addition of specific DNA, but also increases slightly upon the addition of nonspecific DNA (40,41). The advantages of induced helicity upon binding nonspecific DNA are numerous. For one, having a flexible DNA-binding domain enables scanning of large amounts of genomic DNA. Furthermore, if a small amount of helicity is induced in the presence of DNA, the extra cost of specific binding is reduced (43). Specific binding is also enhanced because the overall free energy of binding is reduced when folding and binding are coupled, so high affinity binding can only occur when complementarity is maximal (39).

However, the amount of alpha helicity of the p007 variants in the absence of DNA did not correlate with specificity (Figure 2.12). The most apparently preorganized variant in terms of alpha helicity, p007P13A, was not nearly as specific as p007. In addition, the most p007-like variant in terms of affinity and specificity, p007M7A, was considerably less structured than p007, with MRE_{222} = −7327. Furthermore, CD of peptides simply composed of alanines and lysines demonstrate increased alpha helicity upon binding of dsDNA, another indication that the folded state in the absence of DNA does not correlate with specificity (43).

The effects of individual alanine substitutions on the affinity, specificity, and structure of p007 can be mapped onto p007 (Figure 2.13). As predicted, we found that variants containing alanine in place of putative DNA-binding residues possess severely diminished specific DNA affinity and specificity. In addition, we found that variants containing alanine in place of residues within the hydrophobic core also possess diminished specific DNA affinity and specificity. The importance of most of the selected
Figure 2.12  Plot of $\Delta \Delta G_{\text{spec}}$ vs. MRE$_{222}$ of p007 and all variants, illustrating the relationship between specificity and inherent alpha helicity.
**Figure 2.13** Mapping the effect of individual alanine substitutions on the affinity, specificity, and structure of p007.
ΔΔG_{hsCRE} < 1.0
ΔΔG_{hsCRE} = 1.0 to 2.9
ΔΔG_{hsCRE} > 2.9

K_{rel} > 1000
K_{rel} = 1000 to 1
K_{rel} < 1

MRE_{222} < −16,000
MRE_{222} = −16,000 to −10,000
MRE_{222} > −10,000

Mutation to alanine results in a small effect
Mutation to alanine results in a moderate effect
Mutation to alanine results in a large effect
residues in the PPII helix was confirmed by diminished affinity and specificity of the corresponding alanine containing variants. While all variants were less structured than p007 as determined by the extent of alpha helicity at 4 °C from CD spectroscopy, there was little correlation between structure and DNA-binding affinity or specificity. This alanine scanning mutagenesis study of the miniature DNA-binding protein p007 has demonstrated that nearly every residue contributed to affinity and specificity for hsCRE DNA. No mutation enhanced the function of p007, nor improved its specificity. One implication for miniature protein design is that potentially all residues can contribute to structure, function, and specificity. The residues chosen for the initial grafting approach and for subsequent molecular evolution must be judiciously chosen.

**Experimental**

**General**

All DNA sequences are written in the 5’ to 3’ direction. Polypeptides are written from the amino terminus to the carboxy terminus. All peptides are amidated at the carboxy termini and have free amino termini.

Standard sterile technique was maintained throughout. Media and equipment were sterilized by autoclaving at 121 °C and 20 p.s.i. for 20 min.

UV/Visible spectrophotometry was carried out on a Beckman DU 640B instrument. Lyophilization was carried out using a Savant SC100 Speed Vac. Peptide synthesis and amino acid analysis were performed by the W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University School of Medicine (New Haven,
MALDI-TOF was performed on a Voyager-DE Pro mass spectrometer. Nondenaturing polyacrylamide gel electrophoresis was performed in a model SE600 Dual-Controller Vertical Slab Unit (Hoefer) using 14x16 cm plates. The temperature was controlled at 25 °C using a constant temperature bath (Model RTE-10 (Neslab Instruments)). Gels were maintained at constant temperature by recirculating the buffer between the two chambers. TE contained 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0). TBE contained 45 mM Tris-borate, 1 mM EDTA (pH 8.0).

**Peptide synthesis**

Peptides were synthesized on a 25 µmol scale using Fmoc chemistry at the W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University School of Medicine (New Haven, CT) or on a Symphony/Multiplex synthesizer (Protein Technologies, Inc.). All peptides contained an N-terminal amine and a C-terminal amide. The synthesis of all peptides was completed on a 25 µmol scale using Rink amide resin which resulted in an amidated carboxy terminus using standard Fmoc chemistry. Fmoc amino acids and HBTU were used in five-fold excess (125 µmol per coupling). HBTU was used as the activating reagent without addition of supplemental hydroxybenzotriazole. The peptide synthesis cycle was performed as follows: resin was swelled for 3 x 10 min with NMP; 20% piperidine in DMF was added to the resin and allowed to mix for 2 min, followed by 2 x 7 min; resin was washed for 6 x 30 s with NMP; 400 µmol of NMM and 25 µmol of HBTU in NMP was added to the Fmoc amino acid; the activated amino acids were allowed to react for 20 min; after completion of the reaction, the resin was washed for 30 s with NMP, and coupled again; resin was washed
for 3 x 30 s with NMP; capping was carried out with 5% NMM and 5% acetic anhydride in NMP for 2 x 2.5 min; resin was washed for 6 x 30 s. Peptide cleavage and side chain deprotection were achieved by treatment with a cocktail comprised of 95% TFA, 1% phenol, 1% ethanedithiol, 1% thioanisole, 1% triisopropylsilane, and 1% water. The peptide was precipitated by the addition of cold diethylether. The mixture was centrifuged at 3000 rpm for 5 min, and the supernatant was discarded. The precipitated peptide was washed with cold diethylether, centrifuged, and the supernatant was discarded. The peptides were dissolved in 25% acetonitrile, lyophilized, and stored in the freezer until purification.

**Peptide purification**

Peptides were dissolved in acetonitrile and water, then filtered through a 0.2 micron filter. Peptides were purified by reverse phase HPLC on a Rainin Dynamax Solvent Delivery System Model SD-200 with a Dynamax PDA-2 Diode Array Detector or on a Varian ProStar 210 Solvent Delivery System with a ProStar 330 PDA Detector. Peptide purifications were performed on a Dynamax preparative C18 column (300Å, 5 µm, 21.4 mm x 250 mm) at a flow rate of 15 mL/min or on a Vydac semi-preparative C18 column (300Å, 5 µm, 10 mm x 150 mm) at a flow rate of 5 mL/min. Solvent A contained 98% HPLC grade water, 2% HPLC grade acetonitrile, 0.06% trifluoroacetic acid, and solvent B contained 20% HPLC grade water, 80% HPLC grade acetonitrile, and 0.05% trifluoroacetic acid. Peptide purity was confirmed by reinjection on a Vydac analytical C18 column (300Å, 5 µm, 4.6 mm x 150 mm). Concentrated peptide stocks in dH₂O were made from lyophilized HPLC purified peptides. Peptide concentrations were
determined by amino acid analysis, using homoserine and norleucine as internal standards, on a Beckman 7300 instrument. Mass spectrometry on an Applied Biosystems (Foster City, CA) Voyager-DE-Pro MALDI-TOF instrument using an $\alpha$-cyano-4-hydroxy-cinnamic acid matrix confirmed the identity of all peptides.

\textbf{P007}^{G1A}: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly2 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.4 Thr1.9 Ser1.8 Glx1.2 Pro2.1 Gly2.1 Ala11.5 Val1.1 Met0.9 Leu2.2 Phe1.1 Lys2.2 Arg7.2. Mass expected: 4181.8; observed: 4178.8.

\textbf{P007}^{G2A}: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly2 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.4 Thr1.8 Ser1.4 Glx1.2 Pro2.1 Gly2.2 Ala11.9 Val1.0 Met0.7 Leu2.3 Phe1.1 Lys2.2 Arg7.3. Mass expected: 4181.8; observed: 4178.8.

\textbf{P007}^{S3A}: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly2 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.4 Thr1.9 Ser0.9 Glx1.1 Pro2.2 Gly3.0 Ala11.5 Val1.1 Met0.7 Leu2.2 Phe1.1 Lys2.2 Arg7.0. Mass expected: 4151.8; observed: 4149.0.

\textbf{P007}^{R4A}: Amino acid analysis expected: Asx4 Thr2 Ser1 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.4 Thr1.9 Ser1.7 Glx1.1 Pro2.0 Gly3.0 Ala11.6 Val1.1 Met0.9 Leu2.2 Phe1.1 Lys2.2 Arg6.4. Mass expected: 4082.6; observed: 4078.3.

\textbf{P007}^{T6A}: Amino acid analysis expected: Asx4 Thr1 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.4 Thr1.0 Ser1.6 Glx1.1 Pro1.9 Gly3.0 Ala11.8 Val1.0 Met0.8 Leu2.3 Phe1.1 Lys2.2 Arg7.3. Mass expected: 4137.7; observed:
P007M7A: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.2 Thr1.9 Ser1.8 Glx1.1 Pro2.1 Gly2.9 Ala11.6 Val1.1 Leu2.2 Phe1.1 Lys2.2 Arg7.2. Mass expected: 4107.6; observed: 4103.9.

P007P8A: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro1 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.1 Thr1.9 Ser1.8 Glx1.1 Pro1.0 Gly2.7 Ala12.9 Val1.1 Met0.8 Leu2.2 Phe1.1 Lys2.2 Arg7.1. Mass expected: 4141.7; observed: 4138.0.

P007P8Z: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro1 Gly3 Ala10 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.4 Thr2.0 Ser1.8 Glx1.2 Pro1.0 Gly3.3 Ala10.6 Val1.1 Met0.9 Leu2.3 Phe1.1 Lys2.2 Arg7.4. Mass expected: 4141.7; observed: 4138.8.

P007G9A: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly2 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.2 Thr2.0 Ser1.8 Glx1.1 Pro2.1 Gly1.9 Ala11.2 Val1.1 Met1.0 Leu2.2 Phe1.1 Lys2.3 Arg7.3. Mass expected: 4181.8; observed: 4177.6.

P007D10A: Amino acid analysis expected: Asx3 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx3.2 Thr1.9 Ser1.7 Glx1.1 Pro2.0 Gly3.0 Ala11.9 Val1.0 Met0.9 Leu2.2 Phe1.1 Lys2.2 Arg7.3. Mass expected: 4123.7; observed: 4122.2.

P007D11A: Amino acid analysis expected: Asx3 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx3.3 Thr1.9 Ser1.7 Glx1.1 Pro2.1 Gly2.9 Ala11.8 Val1.1 Met0.9 Leu2.2 Phe1.1 Lys2.3 Arg7.2. Mass expected: 4123.7; observed:
P007p13A: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro1 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.5 Thr2.0 Ser1.6 Glx1.1 Pro0.6 Gly2.9 Ala11.9 Val1.0 Met0.8 Leu2.4 Phe1.1 Lys2.3 Arg7.4. Mass expected: 4141.7; observed: 4139.3.

P007p13Z: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro1 Gly3 Ala10 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.4 Thr1.9 Ser1.8 Glx1.1 Pro1.2 Gly3.2 Ala10.7 Val1.0 Met0.9 Leu2.2 Phe1.1 Lys2.2 Arg7.4. Mass expected: 4141.7; observed: 4138.3.

P007v14A: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.4 Thr2.0 Ser1.8 Glx1.1 Pro2.1 Gly3.1 Ala11.5 Met1.0 Leu2.2 Phe1.1 Lys2.2 Arg7.2. Mass expected: 4139.7; observed: 4136.1.

P007e15A: Amino acid analysis expected: Asx4 Thr2 Ser2 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.4 Thr2.0 Ser1.7 Pro2.0 Gly3.0 Ala11.7 Val1.1 Met1.0 Leu2.2 Phe1.1 Lys2.2 Arg7.2. Mass expected: 4109.7; observed: 4108.9.

P007d16A: Amino acid analysis expected: Asx3 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx3.3 Thr2.0 Ser1.9 Glx1.2 Pro1.8 Gly3.2 Ala11.6 Val1.1 Met0.9 Leu2.3 Phe1.1 Lys2.1 Arg7.2. Mass expected: 4123.7; observed: 4121.7.

P007l17A: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu1 Phe1 Lys2 Arg8; observed: Asx4.4 Thr1.8 Ser1.5 Glx1.1 Pro2.3 Gly3.3 Ala11.8 Val1.1 Met0.9 Leu1.1 Phe1.1 Lys2.2 Arg7.2. Mass expected: 4125.7; observed:
P007

P007\textsuperscript{R19A}: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys1 Arg8; observed: Asx4.4 Thr2.0 Ser1.9 Glx1.1 Pro2.2 Gly3.1 Ala11.6 Val1.1 Met1.0 Leu2.2 Phe1.1 Lys1.0 Arg7.1. Mass expected: 4110.6; observed: 4107.2.

P007\textsuperscript{R19A}: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg7; observed: Asx4.4 Thr2.0 Ser1.8 Glx1.1 Pro1.9 Gly3.1 Ala11.7 Val1.0 Met0.8 Leu2.3 Phe1.1 Lys2.2 Arg6.2. Mass expected: 4082.6; observed: 4075.9.

P007\textsuperscript{F20A}: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Lys2 Arg8; observed: Asx4.4 Thr1.8 Ser1.5 Glx1.1 Pro2.3 Gly3.3 Ala11.8 Val1.1 Met0.9 Leu2.2 Lys2.1 Arg7.2. Mass expected: 4091.6; observed: 4092.3.

P007\textsuperscript{R21A}: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg7; observed: Asx4.3 Thr1.9 Ser1.8 Glx1.1 Pro2.2 Gly3.1 Ala11.5 Val1.0 Met0.9 Leu2.2 Phe1.1 Lys2.2 Arg6.3. Mass expected: 4082.6; observed: 4077.8.

P007\textsuperscript{N22A}: Amino acid analysis expected: Asx3 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx3.3 Thr2.0 Ser1.8 Glx1.1 Pro2.0 Gly3.2 Ala11.5 Val1.1 Met1.0 Leu2.2 Phe1.1 Lys2.2 Arg7.3. Mass expected: 4124.7; observed: 4122.0.

P007\textsuperscript{T23A}: Amino acid analysis expected: Asx4 Thr1 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.4 Thr1.0 Ser1.7 Glx1.1 Pro2.0 Gly3.1 Ala11.6 Val1.1 Met1.0 Leu2.2 Phe1.1 Lys2.2 Arg7.2. Mass expected: 4137.7; observed:
P007L24A: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu1 Phe1 Lys2 Arg8; observed: Asx4.3 Thr1.8 Ser1.5 Glx1.2 Pro2.3 Gly3.2 Ala11.6 Val1.1 Met1.0 Leu1.1 Phe1.1 Lys2.1 Arg7.1. Mass expected: 4125.7; observed: 4127.2.

P007R27A: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg7; observed: Asx4.3 Thr1.9 Ser1.8 Glx1.1 Pro2.0 Gly3.1 Ala11.7 Val1.1 Met0.9 Leu2.2 Phe1.1 Lys2.3 Arg6.3. Mass expected: 4082.6; observed: 4080.8.

P007R32A: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg7; observed: Asx4.3 Thr1.9 Ser1.8 Glx1.1 Pro1.9 Gly3.2 Ala11.6 Val1.1 Met0.9 Leu2.2 Phe1.1 Lys2.2 Arg6.3. Mass expected: 4082.6; observed: 4081.4.

P007S29A: Amino acid analysis expected: Asx4 Thr2 Ser1 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg8; observed: Asx4.3 Thr1.9 Ser0.9 Glx1.1 Pro2.0 Gly3.1 Ala11.6 Val1.1 Met0.9 Leu2.2 Phe1.1 Lys2.2 Arg7.2. Mass expected: 4151.8; observed: 4150.8.

P007R30A: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg7; observed: Asx4.3 Thr1.9 Ser1.8 Glx1.1 Pro1.9 Gly3.0 Ala11.7 Val1.1 Met0.9 Leu2.2 Phe1.1 Lys2.3 Arg6.3. Mass expected: 4082.6; observed: 4081.7.

P007R32A: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg7; observed: Asx4.3 Thr1.9 Ser1.8 Glx1.1 Pro1.9 Gly3.0
Ala11.7 Val1.1 Met0.9 Leu2.2 Phe1.1 Lys2.2 Arg6.3. Mass expected: 4082.6; observed: 4080.6.

**P007**<sup>K33A</sup>: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys1 Arg8; observed: Asx4.3 Thr1.9 Ser1.8 Glx1.1 Pro2.1 Gly3.1 Ala11.7 Val1.1 Met0.9 Leu2.2 Phe1.1 Lys1.1 Arg7.2. Mass expected: 4110.6; observed: 4107.8.

**P007**<sup>R36A</sup>: Amino acid analysis expected: Asx4 Thr2 Ser2 Glx1 Pro2 Gly3 Ala11 Val1 Met1 Leu2 Phe1 Lys2 Arg7; observed: Asx4.2 Thr1.9 Ser1.9 Glx1.1 Pro2.2 Gly3.1 Ala11.5 Val1.1 Met1.0 Leu2.2 Phe1.1 Lys2.1 Arg6.2. Mass expected: 4082.6; observed: 4075.2.

**DNA**

The oligonucleotides hsCRE25A and hsCRE25B were synthesized on a 1.0 µmol scale at the HHMI Biopolymer/Keck Foundation Biotechnology Resource Laboratory (Yale University School of Medicine, New Haven, CT). hsCRE<sub>25</sub> contained the following sequence: 5’-AGTGGAGATGACGAGCTACTCGTGC-3’ and its complementary strand. The oligonucleotides were resuspended in 150 µL dH<sub>2</sub>O and purified on a denaturing acrylamide gel. To a 25 µL aliquot, an equal volume of formamide loading buffer was added and heated to 95 °C for 2 min and cooled on ice. The gel was pre-electrophoresed at 65 W for 30 min. Samples were loaded on a 20% (19:1 acrylamide:bisacrylamide) polyacrylamide preparative gel with 7M urea in 1X TBE. Running buffer was 1X TBE. The DNA was loaded in the absence of current and was run at 65 W for 5 h. Bands were visualized by UV shadow, excised, and eluted in...
1X TE for 12 h at room temperature. The oligonucleotides were desalted by gel filtration on a NAP-10 column (Pharmacia) and stored in water at –20 °C. The concentrations of DNA were determined by the UV absorbance at 260 nm.

**Radiolabeling DNA**

The A strand of hsCRE<sub>25</sub> was radiolabeled at the 5’ end with [γ-<sup>32</sup>P]-ATP and T4 polynucleotide kinase. Labeling reactions were carried out in a reaction volume of 10 µL containing 100 pmol of DNA, 1X kinase buffer (70 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 5 mM DTT, pH 7.6), 10 U of T4 polynucleotide kinase (New England Biolabs), and 10 µCi of [γ-<sup>32</sup>P]-ATP (New England Nuclear). The reactions were incubated for at least 1 h at 37 °C and the reaction volume brought up to a volume of 50 µL with dH<sub>2</sub>O. Reactions were extracted with an equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol, followed by an equal volume of 24:1 chloroform:isoamyl alcohol. Unincorporated label was removed by passage through a G-25 spin column (Pharmacia). 1 µL of a 1 µM solution of the complementary strand was added and the mixture heated to 95 °C for 2 min in a hot block. The block was then removed from heat and the DNA strands were allowed to anneal by slow cooling to room temperature.

**Electrophoretic mobility shift assay**

To quantify the affinity of the selected peptides for hsCRE DNA, each peptide was titrated with a fixed concentration of radiolabeled hsCRE<sub>25</sub> and the free and bound DNA separated by gel electrophoresis. In a typical binding procedure, a peptide was serially diluted from a stock of known concentration into binding buffer [1.4 mM
KH$_2$PO$_4$, 4.3 mM Na$_2$HPO$_4$, 2.7 mM KCl, 137 mM NaCl (pH 7.4), 1 mM EDTA, 0.1% NP-40, 0.4 mg/ml BSA, and 5% glycerol]. To 8 µL of this serial dilution was added 2 µL of [$^{32}$P]-labeled double stranded hsCRE$_{25}$. The final concentration of DNA was ≤ 80 pM, ensuring that the concentration of labeled hsCRE was at least 10-fold lower than the $K_d$ of the peptide being tested. The binding reactions were equilibrated for 1 h at 25 °C. 8 µL of each reaction was loaded in a single well of the pre-equilibrated gel.

Polyacrylamide gels were 8% w/v with 79:1 acrylamide (ICN ultrapure):bisacrylamide (American Bioanalytical) crosslinking in 10 mM Tris, pH 8.1. Gels were by deoxygenation of the acrylamide solution by bubbling nitrogen followed by polymerization with 0.15% of a 20% ammonium persulfate solution and 0.075% tetramethylenediamine (TEMED) (American Bioanalytical). Gels were pre-equilibrated to 25 °C and pre-electrophoresed at 100 V for 30 min. After samples were loaded the gels were run at 500 V for 30 min. After electrophoresis, gels were transferred to Whatman filter paper and dried with vacuum at 80 °C on a Model 583 Gel Dryer (Bio-Rad). The dry gels were exposed to Phosphor screens (Molecular Dynamics) and the intensity of bands imprinted on the Phosphor screen imaged using a Storm 840 Phosphoimager (Molecular Dynamics). The amounts of free and bound DNA were quantified using ImageQuant software (Molecular Dynamics). Dissociation constants ($K_d$) were calculated by fitting the data to the Langmuir equation (2.1), using non-linear least squares fitting in the program Kaleidagraph 3.6 (Synergy Software), where $\Theta$ is the fraction DNA bound = (counts bound DNA/counts bound DNA + counts free DNA; [peptide]$_T$ = total peptide concentration and c is an adjustable parameter representing the maximum fraction bound (44,45).
Competition electrophoretic mobility shift assays were performed to quantify the specificity of the selected peptides for hsCRE with respect to mutant DNA sites or nonspecific DNA. Peptide at a fixed concentration below its $K_d$ was equilibrated with a fixed concentration of radiolabeled specific DNA. Nonspecific DNA was titrated with the binding mixture and the free and bound labeled DNA separated by gel electrophoresis.

In a typical binding procedure, 2 µL of a serial dilution of calf thymus DNA was added to 8 µL of binding buffer [1.4 mM KH$_2$PO$_4$, 4.3 mM Na$_2$HPO$_4$, 2.7 mM KCl, 137 mM NaCl (pH 7.4), 1 mM EDTA, 0.1% NP-40, 0.4 mg/ml BSA, and 5% glycerol] containing a fixed concentration of peptide sufficient for a 50 to 80% mobility shift in the absence of competitor DNA and ≤ 80 pM [$^{32}$P]-labeled hsCRE DNA. The binding reactions were equilibrated for 1 h at 25 °C. 8 µL of each reaction was loaded in a single well of the pre-equilibrated gel.

Polyacrylamide gels were 8% w/v with 79:1 acrylamide (ICN ultrapure):bisacrylamide (American Bioanalytical) crosslinking in 10 mM Tris, pH 8.1. Gels were by deoxygenation of the acrylamide solution by bubbling nitrogen followed by polymerization with 0.15% of a 20% ammonium persulfate solution and 0.075% tetramethylenediamine (TEMED) (American Bioanalytical). Gels were pre-equilibrated to 25 °C and pre-electrophoresed at 100 V for 30 min. After samples were loaded the
gels were run at 500 V for 30 min. After electrophoresis, gels were transferred to Whatman filter paper and dried with vacuum at 80 °C on a Model 583 Gel Dryer (Bio-Rad). The dry gels were exposed to Phosphor screens (Molecular Dynamics) and the intensity of bands imprinted on the Phosphor screen imaged using a Storm 840 Phosphoimager (Molecular Dynamics). The amounts of free and bound DNA were quantified using ImageQuant software (Molecular Dynamics).

**Calculation of specificity**

The fraction of DNA bound (\(\Theta\)) was calculated as the cpm contained within the band corresponding to the bound DNA divided by the sum of the cpm present in the bands corresponding to the bound and the free DNA within each lane. Competition data were fit as previously described (24), using equation 2.2.

\[
C\Theta = \left[ \frac{-K_{d,ms}}{1 - \Theta_0} \right] \Theta + \left[ \frac{K_{d,ns}}{(1 - \Theta_0)/\Theta_0} \right]
\]

Where \(C\) is the molar concentration of competitor sites, \(\Theta\) is the fraction of protein bound to the specific DNA site, \(C\Theta\) is the fraction of peptide bound to the specific DNA site in the absence of competitor, and \(K_{d,ms}\) is the dissociation constant for binding the competitor site. A plot of \(C\Theta\) versus \(\Theta\) is linear and \(K_{d,ms}\) can be extracted from the gradient or the intercept on the ordinate axis. The \(K_{rel} = K_{d,ms}/K_d\), where \(K_{d,ms}\) is the dissociation constant for binding the nonspecific site, and \(K_d\) is the dissociation constant for binding hsCRE25.

For experiments involving calf thymus DNA, the molar concentration of competitor sites was calculated from its concentration in mg/mL and the average molar
mass of the base, assuming that every base of DNA constitutes the start of a potential competitor site on either strand of duplex DNA (24).

**Far UV circular dichroism**

Far UV circular dichroism was performed on an Aviv 62DS spectrometer using a 0.2 cm pathlength cell. All spectra were recorded in 1.4 mM KH$_2$PO$_4$, 4.3 mM Na$_2$HPO$_4$, 2.7 mM KCl, and 137 mM NaCl (pH 7.4). Spectra were background corrected but not smoothed. For each sample, wavelength scans were performed between 200 and 260 nm with a 1 nm step size and an averaging time of 6 s with the temperature set to 4 °C. The observed ellipticity at a given wavelength was converted to mean residue ellipticity (MRE) using equation 2.3.

\[
\text{MRE} = \frac{\Theta}{10 \times n \times c \times l}
\]

Where $\Theta$ is the observed ellipticity (mdeg), $n$ is the number of amino acids in a protein monomer, $l$ is the path length of the cuvette (cm) and $c$ is the concentration (M).
References


