Relationship between Folding and Function in a Sequence-Specific Miniature DNA-Binding Protein†

Loretta Yang‡,§ and Alanna Schepartz*,||,⊥

Departments of Molecular Biophysics and Biochemistry, Chemistry, and Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520

Received January 20, 2005; Revised Manuscript Received March 23, 2005

ABSTRACT: Previously, we have described a miniature protein-based approach to the design of molecules that bind DNA or protein surfaces with high affinity and specificity. In this approach, the small, well-folded protein avian pancreatic polypeptide acts as a scaffold to present and stabilize an α-helical or PPII-helical recognition epitope. The first miniature protein designed in this way, a molecule called p007, presents the α-helical recognition epitope found on the bZIP protein GCN4 and binds DNA with nanomolar affinity and exceptional specificity. In this work we use alanine-scanning mutagenesis to explore the contributions of 29 p007 residues to DNA affinity, specificity, and secondary structure. Virtually every residue within the p007 α-helix, and most residues within the p007 PPII helix, contribute to both DNA affinity and specificity. These residues include those introduced to make specific and nonspecific DNA contacts, as well as those that complete the miniature protein core. Moreover, there exists a direct correlation between the affinity of a p007 variant for specific DNA and the ability of that variant to select for specific DNA over nonspecific DNA. Although we observe no correlation between α-helicity and affinity, we observe a limited correlation between α-helicity and sequence specificity that emphasizes the role of coupled binding/folding in the function of p007. Our results imply that formation of a highly evolved set of protein·DNA contacts in the context of a well-packed hydrophobic core, and not the extent of intrinsic α-helical structure, is the primary determinant of p007 function.

The development of general strategies for the selective recognition of macromolecular targets remains a major challenge for chemical biology and a fundamental postgenome goal. Molecules capable of tight and selective macromolecular recognition have utility in the validation of potential therapeutic targets and can, in certain cases, function as therapeutics in their own right (*1*−*3*). Our laboratory has developed a general approach toward the design of small, well-folded proteins that bind macromolecular targets with high affinity and selectivity (*4*−*11*). This approach is often referred to as protein grafting, and the molecules that result are called miniature proteins: miniature because they contain fewer than 40 amino acids and proteins because they often fold cooperatively. In a protein grafting experiment, those residues that comprise a natural α-helical or type II polyproline (PPII)-helical recognition epitope are introduced onto the solvent-exposed α- or PPII-helical face of the small yet stable protein avian pancreatic polypeptide (aPP)[1] (*12*, *13*). Our laboratory has used this procedure, often in combination with directed evolution, to engineer miniature protein ligands with high affinity for a variety of targets including the duplex DNAs recognized by GCN4 and CREB (*4*, *5*) and the Q50K engrailed homeodomain (*7*), the antiapoptotic proteins Bcl-X_L (*6*) and Bcl-2 (*10*), the oncoprotein hDM2 (manuscript in preparation), the coactivator CBP (*8*), the EVH1 domain in Mena (*9*), and, in combination with an active site directed inhibitor, a selective protein kinase inhibitor (*11*). Similar procedures have been used by Shimba et al. to generate an aPP-based ligand for the Karposi sarcoma-associated herpesvirus protease that inhibited protein dimerization in the micromolar concentration range (*14*) and by Cobos et al. to generate a 20 μM ligand for the Abl SH3 domain (*15*). In addition to exhibiting high affinity for their targets, many of the miniature proteins studied in our laboratory display high levels of specificity and distinguish effectively between closely related DNA sequences or protein family members. In this work we use alanine-scanning mutagenesis to explore the relationship between folding and function in the first molecule designed using this approach, the DNA-binding miniature protein p007 (*4*, *5*).

The design of p007 proceeded through two distinct stages (*4*, *5*). In the first stage, all 12 residues comprising the DNA-binding surface of GCN4 (*16*−*18*) were substituted for 12 solvent-exposed positions along the aPP α-helix to produce the polypeptide PPBR4 (*4*). These substitutions by necessity

† This work was supported by the NIH (Grant GM 65453) and the National Foundation for Cancer Research and in part by a grant to Yale University, in support of A.S., from the Howard Hughes Medical Institute.

* To whom correspondence should be addressed: phone, (203) 432-5094; fax, (203) 432-3486; e-mail, alanna.schepartz@yale.edu.

‡ Department of Molecular Biophysics and Biochemistry.

§ Current address: La Jolla Bioengineering Institute, 505 Coast Blvd. South, La Jolla, CA 92037.

|| Department of Chemistry.

⊥ Department of Molecular, Cellular and Developmental Biology.

[1] Abbreviations: PPII helix, type II polyproline helix; aPP, avian pancreatic polypeptide; CD, circular dichroism; hsCRE, half-site CRE; K_d, dissociation constant; CT DNA, calf thymus DNA; MRE_{222}, mean residue ellipticity at 222 nm.

10.1021/bi050121h CCC: $30.25
Published on Web 04/30/2005

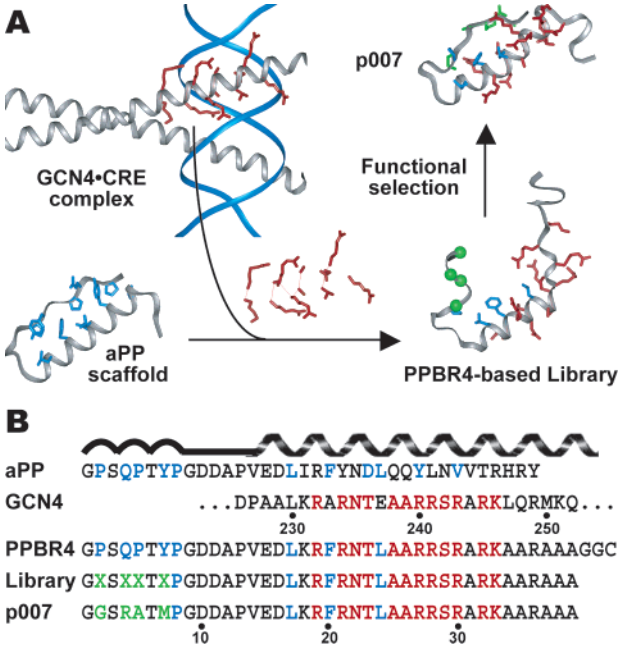

FIGURE 1: (A) Design and optimization of p007. (B) Alignment of the primary sequences of aPP, GCN4$_{226-252}$, PPBR4, and p007. Residues that contact DNA in the GCN4•CRE complex are shown in red, residues that contribute to formation of the aPP hydrophobic core are in blue, and residues selected from the miniature protein library during the evolution of PPBR4 into p007 are in green.

altered three residues derived from the aPP α-helix that contribute to hydrophobic core formation. PPBR4 binds specific DNA with nanomolar affinity at 4 °C, forming a DNA complex that is approximately 3 kcal·mol^{-1} more stable than that of an analogous peptide lacking the PPII helix. However, PPBR4 failed to bind specific DNA detectably at room temperature, and circular dichroism (CD) and NMR experiments indicated that it possessed only nascent α-helicity in the absence of DNA (*19*). Building on the hypothesis that the lack of structure in PPBR4 resulted from disruption of the hydrophobic core, in the second stage of design (*20*) we used phage display to combinatorially vary residues on the N-terminal PPII helix that could repack the core. This selection led to a significantly more helical (*21*) miniature protein, p007, which binds specific DNA with picomolar affinity at 4 °C and with nanomolar affinity at 25 °C and which displays greatly enhanced specificity when compared with PPBR4 (Figure 1A) (*5*). p007 and PPBR4 differ at only four residues, all within the N-terminal PPII helix region (Figure 1B).

In an attempt to deepen our understanding of the interplay between structure and function in p007, we set out to characterize a set of 31 alanine and/or sarcosine variants of p007 with respect to DNA affinity, DNA specificity, and protein secondary structure. Quantitative equilibrium binding experiments indicate that virtually every residue within the p007 α-helix, and the vast majority of residues within the p007 PPII helix, contributes significantly to both DNA affinity and specificity. These residues include those positioned to make specific and nonspecific DNA contacts, as well as those positioned to complete the miniature protein core. At the same time, we observe no single relationship between the DNA affinity or specificity of a given variant and the α-helicity of that variant in the absence of DNA. Taken together, these results imply that formation of a well-packed hydrophobic core upon DNA binding, and not the extent of α-helical structure, is the primary determinant of p007 function.

MATERIALS AND METHODS

Peptides. 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. Crude peptides were purified by reverse-phase HPLC on a Dynamax preparative C18 column (300 Å, 5 μm, 21.4 mm × 250 mm) and a Vydac semipreparative C18 column (300 Å, 5 μm, 10 mm × 150 mm). Purity was confirmed by reinjection on a Vydac analytical C18 column (300 Å, 5 μm, 4.6 mm × 150 mm). The identities of all peptides were confirmed by amino acid analysis and matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry.

DNA. An oligonucleotide containing the half-site CRE (hsCRE) target (hsCRE$_{25A}$, 5′-AGTGGAGATGACGAGCTACTCGTGC-3′) and its complementary strand (hsCRE$_{25B}$, 5′-GCACGAGTAGCTCGTCATCTCCACT-3′) were synthesized on a 1.0 μmol scale at the W. M. Keck Foundation Biotechnology Resource Laboratory. The oligonucleotides were purified by preparative denaturing 20% polyacrylamide gel (19:1 acrylamide:bisacrylamide) electrophoresis. hsCRE$_{25A}$ was radiolabeled at the 5′ end with [γ-^{32}P]ATP and T4 polynucleotide kinase and annealed to the complementary strand by heating the mixture to 95 °C for 2 min and then slowly cooling to room temperature.

Electrophoretic Mobility Shift Assays. Binding reactions were performed in PBS buffer [1.4 mM KH_2PO_4, 4.3 mM Na_2HPO_4, 2.7 mM KCl, and 137 mM NaCl (pH 7.4)] supplemented with 1 mM EDTA, 0.1% NP-40, 0.4 mg·mL^{-1} BSA, and 5% glycerol. In brief, a set of serially diluted peptides was incubated with ≤40 pM ^{32}P-labeled duplex DNA for 1 h at 25 °C and then applied to a preequilibrated nondenaturing 8% polyacrylamide gel (79:1 acrylamide:bisacrylamide) prepared in 10 mM Tris, pH 8.1. The temperature of the gel and the running buffer (10 mM Tris, pH 8.1) were maintained at 25 °C using a circulating, temperature-controlled water bath. The samples were electrophoresed at 500 V for 30 min. After drying, the gels were exposed to a phosphor screen overnight and analyzed using a Storm 840 Phosphorimager (Molecular Dynamics). Free and bound DNA were quantitated using ImageQuant software (Molecular Dynamics) and analyzed using Kaleidagraph (Abelbeck Software). The dissociation constant (K_d) was determined by fitting the data to the Langmuir equation $\Theta = c\{1/(1 + K_d/[\text{peptide}]_T)\}$. In this equation, Θ = (cpm in peptide·DNA complex)/(cpm in peptide·DNA complex + cpm in free DNA), $[\text{peptide}]_T$ is the total peptide concentration, and c is an adjustable parameter representing the maximum value of Θ. Values reported represent the average ± standard error of three independent trials. Error bars shown represent the standard error for each data point. ΔG values were calculated from the relationship $\Delta G = -RT \ln K_d^{-1}$, where R is the universal gas constant (1.987×10^{-3} kcal·mol^{-1}·K^{-1}) and T is the temperature in kelvin.

Table 1: Dissociation Constants of Complexes between p007 and Variants Thereof and Both Specific (hsCRE) and Nonspecific Calf Thymus DNA (CT DNA) at 25 °C

	K_d (nM)	ΔG_{hsCRE} (kcal·mol^{-1})	$K_{d,ns}$ (nM)	$\Delta G_{CT\ DNA}$ (kcal·mol^{-1})	K_{rel} ($K_{d,ns}/K_d$)	$\Delta\Delta G_{spec}$ (kcal·mol^{-1})	MRE_{222}[a] (deg·cm^2·dmol^{-1})
p007	1.6 ± 0.1	−12.0	6670 ± 1200	−7.0	4169	−4.9	−18000
G1A	18.5 ± 5.5	−10.5	13400 ± 2940	−6.6	724	−3.9	−10435
G2A	91.0 ± 28.0	−9.6	11700 ± 2280	−6.7	129	−2.9	−8841
S3A	16.5 ± 1.4	−10.6	10400 ± 1400	−6.8	630	−3.8	−8787
R4A	427 ± 66	−8.7	44.0 ± 2.3	−10.0	0.1	1.3	−9919
T6A	39.3 ± 9.1	−10.1	13700 ± 3070	−6.6	349	−3.5	−12397
M7A	9.5 ± 1.7	−10.9	23000 ± 3140	−6.3	2429	−4.6	−7327
P8A	90.7 ± 19.0	−9.6	14900 ± 2260	−6.6	164	−3.0	−7044
P8Z	31.0 ± 6.0	−10.2	15000 ± 2580	−6.6	478	−3.6	−9674
G9A	14.8 ± 1.3	−10.6	12100 ± 1090	−6.7	818	−4.0	−9151
D10A	2.3 ± 0.5	−11.7	403 ± 54	−8.7	174	−3.0	−10855
D11A	17.1 ± 5.1	−10.6	4160 ± 333	−7.3	243	−3.2	−6214
P13A	20.0 ± 2.6	−10.5	10900 ± 1050	−6.7	545	−3.7	−16537
P13Z	12.0 ± 1.0	−10.8	6000 ± 692	−7.1	500	−3.7	−9598
V14A	219 ± 32	−9.0	22500 ± 3500	−6.3	103	−2.7	−9805
E15A	1.8 ± 0.2	−11.9	744 ± 112	−8.3	418	−3.6	−7149
D16A	1.5 ± 0.1	−12.0	647 ± 43	−8.4	428	−3.6	−9351
L17A	168 ± 21	−9.2	18200 ± 3040	−6.4	108	−2.8	−7975
K18A	239 ± 39	−9.0	33.7 ± 9.1	−10.2	0.1	1.2	−8955
R19A	633 ± 81	−8.4	72.7 ± 16.8	−9.7	0.1	1.3	−8402
F20A	93.3 ± 12.8	−9.6	15300 ± 4840	−6.5	164	−3.0	−8110
R21A	245 ± 25	−9.0	40.3 ± 6.8	−10.0	0.2	1.1	−8984
N22A	608 ± 121	−8.4	91.2 ± 17.4	−9.6	0.2	1.1	−11329
T23A	9.0 ± 1.0	−10.9	7280 ± 339	−7.0	806	−3.9	−10556
L24A	125 ± 21	−9.4	6540 ± 902	−7.0	52	−2.3	−8067
R27A	551 ± 146	−8.5	160 ± 11	−9.2	0.3	0.7	−10112
R28A	503 ± 108	−8.6	81.5 ± 21.3	−9.6	0.2	1.1	−9004
S29A	86.5 ± 10.9	−9.6	27800 ± 2490	−6.2	321	−3.4	−11361
R30A	691 ± 173	−8.4	184 ± 13	−9.2	0.3	0.8	−10699
R32A	422 ± 83	−8.7	26.8 ± 4.3	−10.3	0.1	1.6	−9362
K33A	593 ± 111	−8.5	223 ± 47	−9.0	0.4	0.6	−12214
R36A	85.0 ± 14.4	−9.6	26.4 ± 7.5	−10.3	0.3	0.7	−8759

[a] Mean residue ellipticity at 222 nm of each variant in the absence of DNA (PBS, 4 °C).

Competition Electrophoretic Mobility Shift Assays. An aliquot of 2 μL of a serial dilution of calf thymus DNA (CT DNA) was added to 8 μL of binding buffer [1.4 mM KH_2PO_4, 4.3 mM Na_2HPO_4, 2.7 mM KCl, 137 mM NaCl (pH 7.4), 1 mM EDTA, 0.1% NP-40, 0.4 mg·mL^{-1} BSA, and 5% glycerol] containing a fixed concentration of peptide sufficient for a 50−80% mobility shift in the absence of competitor DNA and ≤80 pM ^{32}P-labeled $hsCRE_{25}$ DNA. The binding reactions were equilibrated for 1 h at 25 °C, applied to a preequilibrated polyacrylamide gel, and electrophoresed as described above. Competition data were fit as previously described using the equation $C\Theta = \{-K_{d,ns}/(1 - \Theta_0)\}\Theta + \{K_{d,ns}/((1 - \Theta_0)/\Theta_0)\}$ (*22*). In this equation, C is the molar concentration of competitor sites, Θ is the fraction of protein bound to the specific DNA site, Θ_0 is the fraction of peptide bound to the specific DNA site in the absence of competitor, and $K_{d,ns}$ is the dissociation constant for binding the competitor site. A plot of $C\Theta$ versus Θ is linear, and $K_{d,ns}$ can be extracted from the gradient or the intercept on the ordinate axis. The $K_{rel} = K_{d,ns}/K_d$, where $K_{d,ns}$ is the dissociation constant for binding the nonspecific site and K_d is the dissociation constant for binding $hsCRE_{25}$. The molar concentration of competitor sites in CT DNA was calculated from its concentration in milligrams per milliliter and the average molar mass of a base, assuming that every base of DNA constitutes the start of a potential competitor site on either strand of duplex DNA.

Circular Dichroism. CD experiments were performed in PBS buffer [1.4 mM KH_2PO_4, 4.3 mM Na_2HPO_4, 2.7 mM KCl, and 137 mM NaCl (pH 7.4)], which contains the same salt concentration as binding buffer. CD experiments were conducted at 4 °C on an Aviv Model 202 spectrometer using a 2 mm path length cell. Samples were scanned from 200 to 250 nm, with a 6 s averaging time and 1 nm step size.

RESULTS

Quantitative Analysis of DNA Affinity among p007 Variants. To study the energetic contribution of individual p007 side chains to DNA affinity, we synthesized 29 variants of p007 in which a single residue was substituted with alanine and used a quantitative electrophoretic mobility shift assay to determine the affinity of each variant for DNA containing the specific half-site CRE (hsCRE) sequence ATGAC (Table 1). We also prepared two variants in which a single proline residue was replaced with sarcosine. The equilibrium dissociation constants of the variant·hsCRE complexes determined in this way ranged from 1.5 to 691 nM, corresponding to binding free energies (ΔG_{hsCRE}) between −12.0 and −8.4 kcal·mol^{-1} (Table 1, Figure 2A).

Contributions of Residues Identified by Directed Evolution. The first variants we consider are the three containing alanine in place of a non-alanine residue selected during the evolution of PPBR4 into p007: G2A, R4A, and M7A (one of the selected residues was alanine and was therefore not studied). All of these residues (shown in green in Figure 2A) are located on the N-terminal PPII helix of p007 (*5*). Each of the variants binds hsCRE DNA with significantly lower affinity, and thus with less favorable binding free energy,

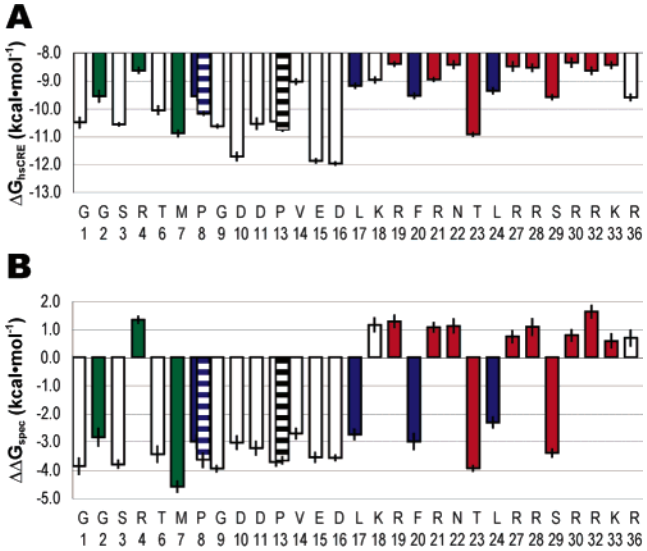

FIGURE 2: (A) Plot illustrating the affinity of each variant·hsCRE complex ($\Delta G_{hsCRE} = -RT \ln K_d^{-1}$). The value for p007 is −12.0 kcal·mol^{-1}. (B) Plot illustrating the specificity of each variant for hsCRE and CT DNA ($\Delta\Delta G_{spec} = -RT \ln K_{rel}$). The value for p007 is −4.9 kcal·mol^{-1}. Residues identified during evolution of p007 (*5*) are in green, residues that contribute to the aPP hydrophobic core are in blue, residues that contact DNA in the context of the GCN4·CRE structure (*18*) are in red, and residues in which a proline residue is substituted with sarcosine are hatched.

than does p007. Of the three variants, R4A binds hsCRE with the lowest affinity (K_d = 427 ± 66 nM), a 3.3 kcal·mol^{-1} loss relative to p007. G2A also binds hsCRE poorly (K_d = 91 ± 28 nM), a 2.4 kcal·mol^{-1} loss relative to p007. M7A binds hsCRE with the highest affinity (K_d = 9.5 ± 1.7 nM), a loss of only 1.0 kcal·mol^{-1} relative to p007. Thus, each of the three non-alanine residues identified during the evolution of PPBR4 into p007 contributes significantly, albeit unequally, to specific DNA affinity. This result provides independent support for the strategy of enhancing the DNA affinity of a miniature protein indirectly by optimizing residues essential to intramolecular packing.

Contributions of Base and Phosphate Contact Residues. Next we consider the ten p007 variants containing alanine in place of a non-alanine residue expected to contact hsCRE DNA, as predicted by homology with the well-characterized GCN4·CRE structure (*18*): R19A, R21A, N22A, T23A, R27A, R28A, S29A, R30A, R32A, and K33A (shown in red in Figure 2A). With one exception, these p007 variants all bind hsCRE DNA with significantly lower affinity than p007 itself, with K_d values between 86.5 and 691 nM. These K_d values correspond to binding energies at least 2.4 kcal·mol^{-1} less favorable than that of p007.

Four of the ten variants in this set, N22A, S29A, R30A, and K33A, contain alanine in place of a highly conserved residue that contacts a base pair in the GCN4·CRE structure. In this structure, the side chain of Asn235 (Asn22 in p007) contacts O4 of T_{-4} and N4 of C_3, the side chain of Ser242 (Ser29 in p007) contacts the T_{-4} methyl group, the side chain of Arg243 (Arg30 in p007) contacts N7 of G_{-4}, and the side chain of Lys246 (Lys33 in p007) interacts through a water molecule with N7 of G_{-3}. Gel shift and yeast complementation assays indicate that substitution of Asn235, Arg243, or Ser242 with a variety of amino acids results in GCN4 variants that bind DNA poorly if at all (*23*−*25*). We find that p007 variants containing alanine in place of Asn22, Ser29, Arg30, or Lys33 bind specific DNA with greatly reduced affinity, with equilibrium dissociation constants of 608 ± 121, 86.5 ± 10.9, 691 ± 173, and 593 ± 111 nM, respectively. These K_d values correspond to free energy losses of 2.4−3.6 kcal·mol^{-1} relative to the wild-type complex.

Five of the remaining six p007 variants, R19A, R21A, R27A, R28A, and R32A, contain alanine in place of a basic residue that contacts the DNA phosphodiester backbone in the GCN4·CRE structure (*18*). In this structure, the side chain of Arg232 (Arg19 in p007) contacts the phosphate of T_2, the side chain of Arg234 (Arg21 in p007) contacts the phosphate of G_{-6}, the side chain of Arg240 (Arg27 in p007) contacts the phosphates of G_1 and C_{-1}, the side chain of Arg241 (Arg28 in p007) contacts the phosphate of A_{-5}, and the side chain of Arg245 (Arg32 in p007) contacts the phosphate of T_2. Although no quantitative data exist describing the effect of these mutations in the context of GCN4, studies of minimalist alanine-based GCN4 mimics that retain at least four of the five base contact residues suggest that each phosphate contact contributes to affinity in a significant way (*26*, *27*). We find that p007 variants containing alanine in place of Arg19, Arg21, Arg27, Arg28, and Arg32 bind specific DNA with greatly reduced affinity, with equilibrium dissociation constants ranging from 245 to 633 nM ($\Delta\Delta G_{hsCRE}$ = 3.0−3.5 kcal·mol^{-1} less favorable than that of p007). It is notable that alanine variants of phosphate-contacting and base-contacting residues bind specific DNA with approximately equivalent affinities.

The final residue that contributes to DNA binding by GCN4, Thr236 (Thr23 in p007), makes no direct or water-mediated base contacts but is believed to align nearby residues Arg232, Glu237, and Arg243 for binding DNA (*18*). GCN4 variants containing asparagine, glutamine, arginine, or methionine at position 236 bind DNA with wild-type affinity (*28*), and most AP-1 family members, such as Fos and Jun, contain arginine or lysine at the equivalent position. We find that the p007 variant containing alanine in place of Thr23 binds specific DNA with virtually wild-type affinity, with an equilibrium dissociation constant of 9.0 ± 1.0 nM ($\Delta\Delta G_{hsCRE}$ = 1.0 kcal·mol^{-1} less favorable than that of p007). Taken as a whole, the data reveal remarkable agreement between the effects of amino acid substitutions within GCN4 and the engineered miniature protein p007, providing strong support for a model in which the α-helices of these two proteins bind in a structurally analogous manner.

Contributions of Putative Folding Residues. Next we consider those five variants, P8A, P8Z, L17A, F20A, and L24A, in which alanine is substituted for a residue that could contribute to formation of the p007 hydrophobic core, based on the crystal structure of aPP (Figure 3A) (*12*) and of the NMR structure p007 in the absence of DNA (Figure 3B) (*5*). In each of these structures, the hydrophobic core is comprised of side chains from residues at positions 5, 7, and 8 on the PPII helix and residues 17, 20, and 24 on the α-helix. Each of these five variants binds specific DNA significantly more poorly than p007, with equilibrium dissociation constants between 31.4 and 168 nM. These values correspond to binding free energies that are between 1.8 and 2.7 kcal·mol^{-1} lower than that of p007, values only slightly lower than those observed with variants containing

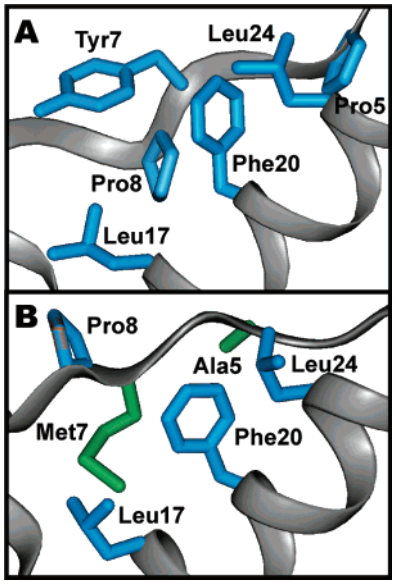

FIGURE 3: (A) Close-up view of side chain packing in the aPP hydrophobic core, illustrating the relative orientation of Pro8, Tyr7, and Pro5 on the PPII helix and Leu17, Phe20, and Leu24 on the α-helix. (B) Close-up view of side chain packing in a model of the p007 hydrophobic core (*5*), illustrating the relative orientation of Pro8 and selected residues Met7 and Ala5 (green) on the PPII helix and Leu17, Phe20, and Leu24 on the α-helix.

alanine in place of a direct DNA contact residue. Interestingly, although substitution of Pro8 with alanine has a significant effect on specific DNA binding ($\Delta\Delta G_{hsCRE} = 2.4$ kcal·mol^{-1} lower than that of p007), substitution with sarcosine has a smaller effect ($\Delta\Delta G_{hsCRE} = 1.8$ kcal·mol^{-1} lower than that of p007). Though Pro8 appears to be solvent-exposed in the average structure of p007 (*21*), long-range NOEs are found between Pro8 and Leu17 and between Pro8 and Phe20. The difference between the functions of alanine- and sarcosine-substituted variants of p007 is consistent with a defined role for the aliphatic side chain of Pro8 in protein stability.

Contributions of Other Residues. To complete our analysis, we also determined the specific DNA affinity of the remaining 13 p007 variants in which alanine is substituted for a residue that was neither selected for nor presumed to be involved in DNA binding or hydrophobic core formation. These variants vary greatly in terms of their affinity for specific DNA. Those variants that possess alanine substitution along the PPII helix (G1A, S3A, and T6A) show modestly reduced specific DNA affinity, with dissociation constants between 16.5 and 39.3 nM. These values correspond to binding energies that are 1.4−1.9 kcal·mol^{-1} less favorable than that of p007. Three of the five variants that possess alanine substitution along the p007 α-helix (V14A, K18A, and R36A) show greatly reduced specific DNA affinity, with dissociation constants between 85 and 239 nM. These values correspond to binding energies that are 2.3−3.0 kcal·mol^{-1} less favorable than that of p007. The remaining two variants, E15A and D16A, which possess alanine substitution at the very N-terminus of the p007 α-helix, bind specific DNA with virtually wild-type affinity, with dissociation constants of 1.8 and 1.5 nM. Variants G9A, D10A, D11A, P13A, and P13Z, which alter the β-turn linking the PPII and α-helices, bind with dissociation constants between 2.3 and 20.0 nM ($\Delta\Delta G_{hsCRE} = 0.2-1.5$ kcal·mol^{-1} less favorable than that of p007). In general, the data suggest that (1) residues within the PPII helix that were not varied combinatorially during directed evolution of p007 contribute less to specific DNA affinity than those that were varied and (2) residues within the p007 α-helix that are not presumed to contact DNA nor contribute to hydrophobic core formation can nevertheless contribute significantly to specific recognition of hsCRE.

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 exceptional specificity for DNA containing the sequence ATGAC. For example, p007 binds the wild-type hsCRE sequence 200−800 times better than it binds analogous DNA sequences containing two base pair mutations (*5*). Moreover, p007 displays a >4000-fold preference for hsCRE over nonspecific calf thymus DNA (CT DNA) (*5*). This preference is considerably greater than the number of potential competitor sites in a random five base pair recognition site ($4^5 = 1024$) (*22*). To determine the extent to which each residue within p007 contributes to specificity, we used a standard competitive binding assay to determine the relative affinity of each p007 variant for CT and hsCRE DNA (Table 1) (*22*). These data allow us to calculate values of $\Delta\Delta G_{spec}$ defined by the relationship $\Delta\Delta G_{spec} = -RT \ln K_{rel}$ (Table 1, Figure 2B).

The 31 variants fall into three categories with regard to their affinity for specific hsCRE and nonspecific CT DNA relative to p007. The first category includes those variants that exhibit far lower specificity (K_{rel} values < 1, $\Delta\Delta G_{spec}$ values > 0 kcal·mol^{-1}) than p007 ($K_{rel} = 4169$, $\Delta\Delta G_{spec} = -4.9$ kcal·mol^{-1}), suggesting that the residue substituted for alanine in that variant plays an important direct or indirect role selecting for hsCRE DNA and/or against the random sequences found in CT DNA. The second category includes those variants that exhibit a level of specificity that equals p007 (K_{rel} values > 1000, $\Delta\Delta G_{spec}$ values < −4 kcal·mol^{-1}), suggesting that the residue substituted for alanine in that variant plays little or no role in hsCRE recognition. In the final category are those variants with intermediate values of K_{rel} ($1000 > K_{rel} > 1$, $\Delta\Delta G_{spec}$ values between 0 and −4 kcal·mol^{-1}). Variants in the first category include most (but not all) of the residues on the p007 α-helix that are expected to contact DNA based on analogy with GCN4 (Arg19, Arg21, Asn22, Arg27, Arg28, Arg30, Arg32, and Lys33) and two residues on the p007 α-helix that were not designed to recognize DNA (Lys18 and Arg36), as well as one residue on the p007 PPII helix (Arg4) that was identified by functional selection. All of these variants exhibit lower specificity than p007 by virtue of a significant decrease in affinity for hsCRE and increase in affinity for CT DNA. It is notable that a large fraction of the DNA contact residues included in this set are predicted to contact the phosphodiester backbone, and not DNA bases, providing direct evidence for a classic case of "indirect readout" (*29*).

It is remarkable that only one variant, M7A, containing alanine in place of a residue identified by functional selection, segregates into the second category (along with p007 itself). Variants in the third category are numerous and include those with alanine in place of (1) one residue on the p007 PPII helix selected by directed evolution (Gly2), (2) two residues on the p007 α-helix that are expected to contact DNA based

on analogy with GCN4 (Thr23 and Ser29), (3) four residues on the p007 PPII or α-helix that contribute to formation of the hydrophobic core (Pro8, Leu17, Phe20, and Leu24), (4) three residues on the p007 PPII helix (Gly1, Ser3, and Thr6), (5) all residues within the *β*-turn linking the p007 PPII and α-helix (Gly9, Asp10, Asp11, and Pro13), and (6) several residues on the p007 α-helix that were not designed to recognize DNA (Val14, Glu15, and Asp16). Many of these variants (those with alanine in place of Pro8, Gly9, Asp11, Pro13, Val14, Lys17, Ser29, and Leu24, a set that includes all folding residues derived from aPP) exhibit lower specificity than p007 by virtue of a significant decrease in affinity for hsCRE with little or no change in affinity for CT DNA. Several other variants (those with alanine in place of Asp10, Glu15, and Asp16) exhibit lower specificity than p007 by virtue of a significant increase in affinity for CT DNA (presumably the result of altered electrostatic interactions) with little or no change in affinity for hsCRE. Taken together, these data emphasize in a dramatic way that no single residue or segment of p007 is sufficient to recapitulate high DNA specificity and that the miniature protein encodes specific hsCRE recognition as a single, functional unit.

Circular Dichroism Analysis of Miniature Protein Secondary Structure. The basic, DNA-binding region of a bZIP protein such as GCN4 exists as an ensemble of transient, minimally α-helical structures in the absence of DNA, achieving the characteristic α-helical form revealed by circular dichroism and crystallography (*16*−*18*, *30*) only when DNA bound (*31*−*39*). The entropic cost associated with this helix−coil transition has been estimated from calorimetric measurements and NMR studies to be as large as $\Delta S_{conf} = -0.29$ kcal·mol^{-1}·K^{-1}, with 40−45% of this amount attributable to loss of backbone conformational entropy (*40*). This entropic cost translates into a free energy cost of 35−40 kcal·mol^{-1} at 25 °C. It is therefore not surprising that alanine substitutions within the basic segment increase the DNA affinities of bZIP peptides related to GCN4 (*19*). The effects of alanine substitutions in the context of p007 are more complicated, however, because of the interplay between the α-helix and the PPII helix and because of the potential for dimerization at high concentration (*41*); substitution of alanine for a solvent-exposed residue within the α-helix may stabilize α-helical secondary structure, whereas substitution of alanine for a residue that contributes to hydrophobic core formation may destabilize α-helical secondary structure if that residue lowers the overall stability of the folded form (even if the change would increase intrinsic α-helical structure). To explore the relationship between DNA affinity and specificity and the extent of α-helix secondary structure, we characterized each p007 variant in the absence of DNA at 4 °C using circular dichroism spectroscopy (Figure 4A).

Surprisingly, all 31 p007 variants displayed less α-helical character than p007 itself in the absence of DNA, with values of mean residue ellipticity at 222 nm (MRE_{222}) that varied between −6214 and −16537 deg·cm^2·dmol^{-1} (Table 1, Figure 4B). The MRE_{222} value of p007 is −18000 deg·cm^2·dmol^{-1}, whereas the value for PPBR4 (which differs from p007 at only four positions) is −8000 deg·cm^2·dmol^{-1}. All eight variants which possess alanine in place of a residue presumed to contribute to p007's hydrophobic core, including those derived directly from aPP (P8A, $MRE_{222} = -7044$; P8Z, $MRE_{222} = -9674$; L17A, $MRE_{222} = -7975$; F20A, $MRE_{222} = -8110$; and L24A, $MRE_{222} = -8067$ deg·cm^2·dmol^{-1}) and those selected during evolution (G2A, $MRE_{222} = -8841$; R4A, $MRE_{222} = -9919$; and M7A, $MRE_{222} = -7327$ deg·cm^2·dmol^{-1}), display relatively low levels of α-helical character ($MRE_{222} > -10000$ deg·cm^2·dmol^{-1}). Other variants with exceptionally low levels of α-helix structure include those containing alanine in place of residues in the *β*-turn (G9A, $MRE_{222} = -9151$; D11A; $MRE_{222} = -6214$; and P13Z, $MRE_{222} = -9598$ deg·cm^2·dmol^{-1}). Interestingly, the variant containing the highest level of α-helix structure (MRE_{222} values < −16000 deg·cm^2·dmol^{-1}) is the alanine variant of Pro13 in the *β*-turn (P13A; $MRE_{222} = -16537$ deg·cm^2·dmol^{-1}). Variants containing moderate levels of α-helix structure (−16000 < MRE_{222} < −10000 deg·cm^2·dmol^{-1}) include those with alanine in place of DNA contact residues (Asn22, Thr23, Arg27, Ser29, Arg30, and Lys33), remaining residues in the PPII helix (Gly1 and Thr6), and one residue in the *β*-turn (Asp10).

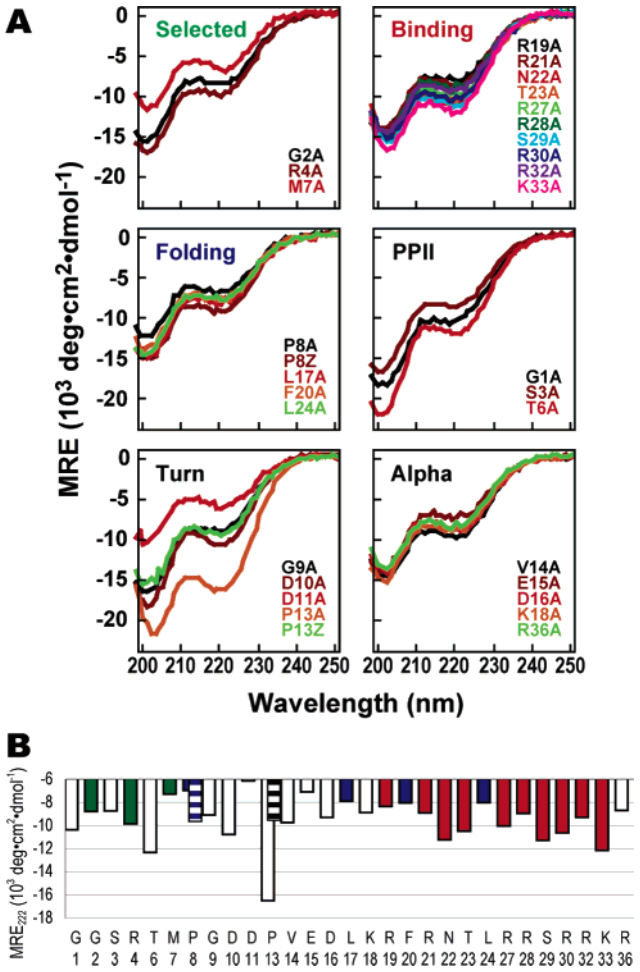

FIGURE 4: (A) CD spectra of p007 variants (10 μM) in PBS buffer at 4 °C. (B) Plot comparing values of MRE_{222} for all p007 variants. The MRE_{222} value for p007 is −18000 deg·cm^2·dmol^{-1}.

DISCUSSION

p007 Binds Specific DNA in a Manner Highly Analogous to GCN4. Examination of the relative affinities of the p007 variants provides strong evidence that p007 binds hsCRE DNA in a manner highly analogous to that of GCN4. For example, GCN4 variants containing virtually any amino acid

in place of the invariant residues Asn235 or Arg243 bind poorly if at all to the CRE site in traditional gel shift assays (*25*). p007 variants containing alanine in place of the corresponding residues Asn22 and Arg30 exhibit hsCRE affinities reduced by 3.5 and 3.6 kcal·mol^{-1}, respectively, emphasizing the importance of these residues to hsCRE recognition. Similarly, GCN4 variants containing alanine in place of Ser242 bind DNA poorly in gel shift assays, and replacement of the corresponding residue within p007 (Ser29) reduces hsCRE affinity by 2.4 kcal·mol^{-1}. Finally, every potential phosphate contact residue derived from a basic residue of GCN4 (p007 residues Arg19, Arg21, Arg27, Arg28, and Arg32) contributes significantly to DNA affinity. These data, combined with the observed high specificity of p007 for the hsCRE sequence, provide strong evidence that p007 and the GCN4 basic segment interact with DNA in the same binding mode, making an analogous set of base and phosphate contacts. This result emphasizes how protein grafting and molecular evolution can generate a molecule that mimics the structure of a natural protein and improve its function.

Residues throughout the p007 Sequence Contribute to hsCRE Affinity. The data also reveal that residues not presumed to bind DNA also contribute significantly to hsCRE affinity. With one exception, all of these residues contribute, either directly or indirectly, to formation of the miniature protein's hydrophobic core. The data demonstrate that energetic contributions from the PPII helix derive from residues that were selected during the directed evolution of PPBR4 into p007 (Gly2, Arg4, and Met7) and one residue derived from aPP itself (Pro8). The interior surface of the p007 α-helix, which packs against the PPII helix, contains three aPP-derived residues that contribute to core formation **(**Leu17, Phe20, and Leu24). Even residues located along the β-turn separating the α-helix and the PPII helix of p007 **(**Gly9, Asp11, and Pro13) appear to help to stabilize the tertiary fold, promoting DNA binding. The one exception to this trend is the variant containing alanine in place of a residue at the very N-terminal end of the α-helix (Val14). Considering the poorer target affinity of p007's unstructured predecessor, PPBR4, these data provide strong evidence that the large increase in hsCRE affinity seen with p007, an increase of greater than 3 kcal·mol^{-1} at 4 °C relative to PPBR4, results from preorganization of the p007 α-helix into a DNA-binding conformation. This increase in equilibrium binding affinity is significantly larger than those seen with GCN4 peptides containing α-helix-promoting lactam bridges (*42*), for which a free energy difference of 1.1 kcal·mol^{-1} is attributed to α-helix stabilization. It is notable that the increases in affinity observed in both systems represent only a fraction of the 35−40 kcal·mol^{-1} cost of the helix−coil transition estimated by NMR methods (*40*), suggesting that it might be possible to increase affinity even further.

Residues throughout the p007 Sequence Contribute to hsCRE Specificity. Examination of the relative specificities of the p007 variants provides strong evidence that both direct miniature protein·DNA contacts and a well-formed hydrophobic core contribute to sequence-specific DNA recognition in this system. The specificity of p007 for specific hsCRE versus nonspecific CT DNA (−4.9 kcal·mol^{-1}) exceeds that achieved by other designed DNA-binding proteins (*22*, *43*, *44*), many transcription factors (*45*), and GCN4 itself. GCN4, which recognizes the palindromic sequence ATGACGTCAT, prefers CRE to CT DNA by 3.3 kcal·mol^{-1} (*46*). p007 residues that contribute the greatest to specificity include those involved in base (Asn22, Arg30, and Lys33) or phosphate contacts (Arg19, Arg21, Arg27, Arg28, and Arg32) and one residue selected during p007 evolution (Arg4). Interestingly, the NMR structure of p007 suggests that the Arg4 side chain is directed away from the protein core, not toward the protein interior, and molecular modeling suggests that it is within reach of the hsCRE phosphodiester backbone. It is notable that these residues, though highly basic in nature, contribute not only to DNA affinity but to target site specificity and that base and phosphate contact residues appear to contribute approximately equally to the high specificity of p007.

Residues that contribute more modestly, but still significantly, to specificity include virtually every residue that contributes to formation of the p007 hydrophobic core. This set includes the remaining residue selected during the directed evolution of PPBR4 into p007 (Gly2) as well as those residues derived from aPP and located on the interior surface of the p007 α-helix **(**Pro8, Leu17, Phe20, and Leu24), those located along the β-turn separating the α-helix and the PPII helix of p007 (Gly9, Asp10, Asp11, and Pro13), and those located at the very N-terminal region of the α-helix (Val14, Glu15, and Asp16). This remarkable result indicates that p007 is highly optimized to recognize its target site. Moreover, p007 makes use of its relatively limited primary sequence space in an extraordinarily efficient manner: almost every side chain is used constructively to preorganize the fold and/or bind target DNA.

No Relationship between DNA Affinity and α-Helical Secondary Structure in the Absence of DNA. Circular dichroism analysis of the secondary structure of GCN4 indicates that the leucine zipper of the dimeric protein is fully helical, with only nascent helicity within the DNA-binding basic region. The basic DNA-binding region becomes fully helical only in complex with DNA (*32*, *36*). Alanine favors the α-helical conformation (*47*), and numerous studies, including those involving variants of PPBR4 (*19*), document the increase in DNA affinity that results from an increase in the intrinsic α-helicity of a GCN4-derived basic segment peptide (*19*, *42*, *48*). By contrast, we find no relationship between the α-helicity of unbound p007 variants and hsCRE affinity (Figure 5A**)**. For example, the p007 variant possessing the lowest level of α-helix structure as judged by CD (D11A) possessed an affinity for hsCRE DNA only 10-fold lower than that of p007. In an analogous way, several p007 variants possessing moderately high levels of α-helix structure (G1A, T6A, and P13A) bind hsCRE DNA poorly. Furthermore, three variants that bind DNA as well as p007 (D10A, E15A, and D16A) all possess varying degrees of structure relative to each other (MRE_{222} = −10855, −7149, and −9351 deg·cm^2·dmol^{-1}, respectively). This result contrasts with similar results obtained with bZIP variants and DNA, in which the effects of analogous alanine substitutions on DNA affinity correlated well to intrinsic α-helicity (*48*). These differing outcomes reflect the fact that, within the DNA-binding region, p007 possesses a tertiary fold while GCN4 contains only secondary structure. Both molecules should benefit from substitutions that increase binding free energy. Within GCN4, whose DNA-bound form

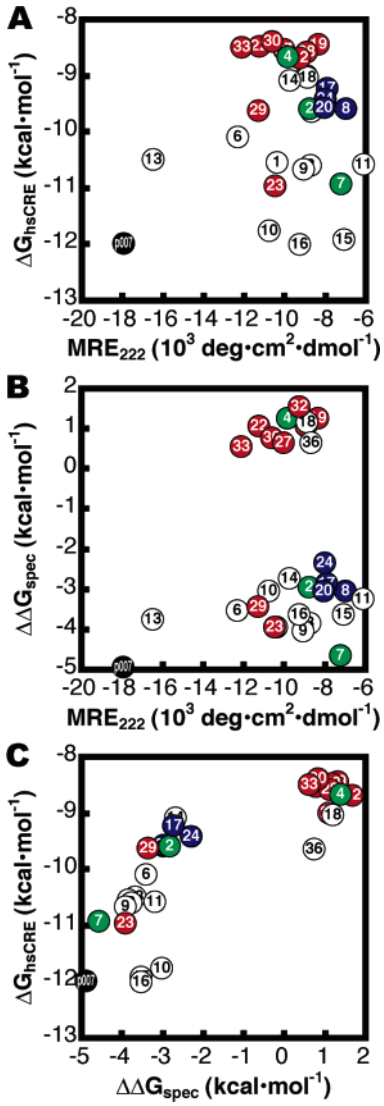

FIGURE 5: (A) Plot of ΔG_{hsCRE} versus MRE_{222} of p007 and alanine variants thereof, illustrating the relationship between affinity and the extent of α-helical structure in the absence of DNA. (B) Plot of $\Delta\Delta G_{spec}$ versus MRE_{222} of p007 and variants thereof, illustrating the relationship between specificity and the extent of α-helical structure in the absence of DNA. (C) Plot of ΔG_{hsCRE} versus $\Delta\Delta G_{spec}$ of p007 and variants thereof, illustrating the relationship between affinity for the hsCRE sequence and the ability to differentiate this sequence from CT DNA. The data for p007 are shown as a black circle, residues identified during the evolution of p007 (*5*) are shown in green, residues that contribute to the aPP hydrophobic core are shown in blue, and residues that contact DNA in the context of the GCN4·CRE structure (*18*) are shown in red.

is a canonical α-helical secondary structure, any substitution that promotes α-helical secondary structure should, in theory, promote DNA recognition. In p007, however, the DNA-bound form is a tertiary structure whose stability depends not only on intrinsic α-helix stability but on a complex set of interactions between the α- and PPII helices. Thus, the presence of the tertiary fold likely accounts for the observed differences in relative side chain contributions to DNA affinity between p007 and GCN4.

Limited Relationship between DNA Specificity and α-Helical Secondary Structure in the Absence of DNA. The data suggest a weak relationship between the α-helicity of certain p007 variants and the ability of that variant to discriminate hsCRE from CT DNA (Figure 5B). This relationship holds for the subset of variants in which a GCN4-derived DNA-binding residue has been changed to alanine (R19A, R21A, N22A, R27A, R28A, R30A, R32A, and K33A) and those in which a potential new DNA-binding residue has been changed to alanine (R4A, K18A, and R36A). In these cases, alanine substitution leads to a significant loss in specificity ($\Delta\Delta G_{spec}$) and a significant decrease in α-helical structure in the absence of DNA. Little or no relationship exists for the remaining variants, including those in which alanine is substituted for a residue involved in formation of the hydrophobic core (P8A, P8Z, L17A, F20A, and L24A), a GCN4-derived residue that contributes little to affinity (T23A and S29A), a residue selected during evolution (G2A and M7A), or one with no presumed function (G1A, S3A, T6A, G9A, D10A, D11A, P13A, P13Z, V14A, E15A, and D16A). In these cases, alanine substitution leads to a smaller loss in specificity ($\Delta\Delta G_{spec}$) and a widely variable decrease in α-helical structure in the absence of DNA.

There are three differences between the residues in these categories. The first is their location within the structure of p007 when DNA-bound, given the reasonable assumption of a structurally analogous binding mode for the p007 and GCN4 α-helices. Most of the residues in the first category, where specificity and intrinsic α-helicity are weakly correlated, are located on the solvent-exposed face of the DNA-bound α-helix and are not involved in intramolecular packing, whereas those in the second class are scattered throughout the structure. In addition, residues in the first category perform a function that has, in the context of GCN4, been improved through natural evolution, whereas those in the second have not. Finally, as pointed out earlier, all residues in the first category achieve specificity by virtue of selection for hsCRE *and* against CT DNA, whereas residues in the second category do so by virtue of selection for hsCRE *or* against CT DNA. Although p007 is significantly more α-helical in the absence of DNA than is PPBR4 (*4*, *19*, *21*) or the basic segment of GCN4, the level of α-helix structure in p007 increases further upon binding specific DNA (*21*). Thus, we propose that the weak correlation between specificity and intrinsic α-helicity observed in the first case reflects the extent to which the identities of Arg19, Arg21, Asn22, Arg27, Arg28, Arg30, Arg32, and Lys33 have been optimized to select the hsCRE sequence while maintaining a high level of α-helical propensity in a reaction in which binding is coupled to local folding (*49*). We also propose that this relationship allows us to assign the function of sequence-selective recognition to p007 residue Arg4 (a residue identified during directed evolution), as well as to Lys18 and Arg36 (which do not make direct DNA contacts in complexes of known structure). Notably, the GCN4 residue corresponding to Arg36 contributes significantly to the half-site specificity of GCN4, its ability to discriminate between the sequences ATGACGTCAT and ATGACTCAT (*50*).

A Direct Relationship Is Observed between Affinity and Specificity. The data suggest a direct relationship between the affinity of a given p007 variant for specific hsCRE DNA and the ability of that variant to discriminate hsCRE from CT DNA (Figure 5C). Although the laws of thermodynamics provide no inherent relationship between values of ΔG_{hsCRE} and $\Delta\Delta G_{spec}$, such a relationship has been reported in several diverse systems, and in each case the relationship has been rationalized on the basis of ligand preorganization. A classic

example is from the ionophore field, in which molecules possessing sterically and electronically preorganized cavities discriminate exceedingly well between the alkali metal cations Na^+ and K^+ (*51*). More recently, it was reported that affinity-matured antibodies that display high affinities and specificities are preorganized for hapten binding. By contrast, the germ-line antibodies from which the affinity-matured antibodies arose underwent significant conformational changes upon binding and displayed lower specificity (*52*). Relationships between affinity and specificity have also been explored using RNA aptamers optimized by in vitro evolution as ligands for HIV reverse transcriptase, basic fibroblast growth factor, and vascular endothelial growth factor (*53*). The high specificities measured for protein-binding RNA aptamers have been rationalized on the basis of the limited torsional freedom of the nucleic acid, which results in a large energetic cost for the conformational changes required to bind incorrect targets. We reason that similar energetic costs are responsible for the high specificities of miniature proteins such as p007. The relationship between the affinity and specificity of p007 variants observed in this work provides strong evidence that formation of a highly evolved set of protein·DNA contacts in the context of a well-packed hydrophobic core, and not the extent of intrinsic α-helical structure, is the primary determinant of p007 function.

ACKNOWLEDGMENT

Oligonucleotide synthesis, peptide synthesis, and amino acid analysis were performed by the W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University School of Medicine (New Haven, CT). We thank D. Daniels, J. Kritzer, and N. Luedtke for comments on the manuscript.

SUPPORTING INFORMATION AVAILABLE

K_d plots and $K_{d,ns}$ plots for p007 and all variants thereof. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Shogren-Knaak, M. A., Alaimo, P. J., and Shokat, K. M. (2001) Recent advances in chemical approaches to the study of biological systems, *Annu. Rev. Cell Dev. Biol. 17*, 405−433.
2. McCormick, F. (2000) Small-molecule inhibitors of cell signaling, *Curr. Opin. Biotechnol. 11*, 593−597.
3. Huang, Z. (2000) Structural chemistry and therapeutic intervention of protein−protein interactions in immune response, human immunodeficiency virus entry, and apoptosis, *Pharmacol. Ther. 86*, 201−215.
4. Zondlo, N. J., and Schepartz, A. (1999) Highly specific DNA recognition by a designed miniature protein, *J. Am. Chem. Soc. 121*, 6938−6939.
5. Chin, J. W., and Schepartz, A. (2001) Concerted evolution of structure and function in a miniature protein, *J. Am. Chem. Soc. 123*, 2929−2930.
6. Chin, J. W., and Schepartz, A. (2001) Design and Evolution of a Miniature Bcl-2 Binding Protein, *Angew. Chem., Int. Ed. Engl. 20*, 3806−3809.
7. Montclare, J. K., and Schepartz, A. (2003) Miniature homeodomains: high specificity without an N-terminal arm, *J. Am. Chem. Soc. 125*, 3416−3417.
8. Rutledge, S. E., Volkman, H. M., and Schepartz, A. (2003) Molecular recognition of protein surfaces: high affinity ligands for the CBP KIX domain, *J. Am. Chem. Soc. 125*, 14336−14347.
9. Golemi-Kotra, D., Mahaffy, R., Footer, M. J., Holtzman, J. H., Pollard, T. D., Theriot, J. A., and Schepartz, A. (2004) High affinity, paralog-specific recognition of the Mena EVH1 domain by a miniature protein, *J. Am. Chem. Soc. 126*, 4−5.
10. Gemperli, A. C., Rutledge, S. E., Maranda, A., and Schepartz, A. (2005) Paralog-selective ligands for Bcl-2 proteins, *J. Am. Chem. Soc. 127*, 1596−1597.
11. Schneider, T. L., Mathew, R. S., Rice, K. P., Tamaki, K., Wood, J. L., and Schepartz, A. (2005) Increasing the kinase specificity of K252a by protein surface recognition, *Org. Lett.* (in press).
12. Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P., and Wu, C.-W. (1981) X-ray analysis (1.4-Å resolution) of avian pancreatic polypeptide: Small globular protein hormone, *Proc. Natl. Acad. Sci. U.S.A. 78*, 4175−4179.
13. Glover, I., Haneef, I., Pitts, J., Wood, S., Moss, D., Tickle, I., and Blundell, T. (1983) Conformational flexibility in a small globular hormone: X-ray analysis of avian pancreatic polypeptide at 0.98-Å resolution, *Biopolymers 22*, 293−304.
14. Shimba, N., Nomura, A. M., Marnett, A. B., and Craik, C. S. (2004) Herpesvirus protease inhibition by dimer disruption, *J. Virol. 78*, 6657−6665.
15. Cobos, E. S., Pisabarro, M. T., Vega, M. C., Lacroix, E., Serrano, L., Ruiz-Sanz, J., and Martinez, J. C. (2004) A miniprotein scaffold used to assemble the polyproline II binding epitope recognized by SH3 domains, *J. Mol. Biol. 342*, 355−365.
16. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: Crystal structure of the protein-DNA complex, *Cell 71*, 1223−1237.
17. Konig, P., and Richmond, T. J. (1993) The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility, *J. Mol. Biol. 233*, 139−154.
18. Keller, W., König, P., and Richmond, T. (1995) Crystal structure of a bZIP/DNA complex at 2.2 Å: Determinants of DNA specific recognition, *J. Mol. Biol. 254*, 657−667.
19. Zondlo, N. J. (1999) Miniaturized DNA-binding peptides, Ph.D. Thesis, Yale University, New Haven, CT.
20. Chin, J. W., Grotzfeld, R. M., Fabian, M. A., and Schepartz, A. (2001) Methodology for optimizing functional miniature proteins based on avian pancreatic polypeptide using phage display, *Bioorg. Med. Chem. Lett. 11*, 1501−1505.
21. Chin, J. W. (2001) Design and evolution of functional miniature proteins, Ph.D. Thesis, Yale University, New Haven, CT.
22. Greisman, H. A., and Pabo, C. O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites, *Science 275*, 657−661.
23. Pu, W. T., and Struhl, K. (1991) Highly conserved residues in the bZIP domain of yeast GCN4 are not essential for DNA binding, *Mol. Cell. Biol. 11*, 4918−4926.
24. Tzamarias, D., Pu, W. T., and Struhl, K. (1992) Mutations in the bZIP domain of yeast GCN4 that alter DNA-binding specificity, *Proc. Natl. Acad. Sci. U.S.A. 89*, 2007−2011.
25. Suckow, M., Schwamborn, K., Kisters-Woike, B., von Wilcken-Bergmann, B., and Müller-Hill, B. (1994) Replacement of invariant bZip residues within the basic region of the yeast transcriptional activator GCN4 can change its DNA binding specificity, *Nucleic Acids Res. 22*, 4395−4404.
26. Lajmi, A. R., Lovrencic, M. E., Wallace, T. R., Thomlinson, R. R., and Shin, J. A. (2000) Minimalist, alanine-based, helical protein dimers bind to specific DNA sites, *J. Am. Chem. Soc. 122*, 5638−5639.
27. Bird, G. H., Lajmi, A. R., and Shin, J. A. (2002) Sequence-specific recognition of DNA by hydrophobic, alanine-rich mutants of the basic region/leucine zipper motif investigated by fluorescence anisotropy, *Biopolymers 65*, 10−20.
28. Suckow, M., von Wilcken-Bergmann, B., and Müller-Hill, B. (1993) Identification of three residues in the basic regions of the bZIP proteins GCN4, C/EBP and TAF-1 that are involved in specific DNA binding, *EMBO J. 12*, 1193−1200.
29. Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) Crystal structure of trp repressor/operator complex at atomic resolution, *Nature 335*, 321−329.
30. Glover, J. N. M., and Harrison, S. C. (1995) Crystal-structure of the heterodimeric Bzip transcription factor C-Fos-C-Jun bound to DNA, *Nature 373*, 257−261.
31. O'Neil, K. T., Hoess, R. H., and DeGrado, W. F. (1990) Design of DNA-binding peptides based on the leucine zipper motif, *Science 249*, 774−778.

32. O'Neil, K. T., Shuman, J. D., Ampe, C., and DeGrado, W. F. (1991) DNA-induced increase in the alpha-helical content of C/EBP and GCN4, *Biochemistry 30*, 9030−9034.
33. Talanian, R. V., McKnight, C. J., and Kim, P. S. (1990) Sequence-specific DNA binding by a short peptide dimer, *Science 249*, 769−771.
34. Talanian, R. V., McKnight, C. J., Rutkowski, R., and Kim, P. S. (1992) Minimum length of a sequence-specific DNA binding peptide, *Biochemistry 31*, 6871−6875.
35. Weiss, M. A. (1990) Thermal unfolding studies of a leucine zipper domain and its specific DNA complex: implications for scissor's grip recognition, *Biochemistry 29*, 8020−8024.
36. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., and Struhl, K. (1990) Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA, *Nature 347*, 575−578.
37. Saudek, V., Pasley, H. S., Gibson, T., Gausepohl, H., Frank, R., and Pastore, A. (1991) Solution structure of the basic region from the transcriptional activator GCN4, *Biochemistry 30*, 1310−1317.
38. Cuenoud, B., and Schepartz, A. (1993) Design of a metallo-bZIP protein that discriminates between CRE and AP1 target sites: Selection against AP1, *Proc. Natl. Acad. Sci. U.S.A. 90*, 1154−1159.
39. Cuenoud, B., and Schepartz, A. (1993) Altered specificity of DNA-binding proteins with transition metal dimerization domains, *Science 259*, 510−513.
40. Bracken, C., Carr, P. A., Cavanagh, J., and Palmer, A. G., III (1999) Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA, *J. Mol. Biol. 285*, 2133−2146.
41. Chang, P. J., Noelken, M. E., and Kimmel, J. R. (1980) Reversible dimerization of avian pancreatic polypeptide, *Biochemistry 19*, 1844−1849.
42. Zhang, M., Wu, B., Zhao, H., and Taylor, J. W. (2002) The effect of C-terminal helix stabilization on specific DNA binding by monomeric GCN4 peptides, *J. Pept. Sci. 8*, 125−136.
43. Morii, T., Sato, S., Hagihara, M., Mori, Y., Imoto, K., and Makino, K. (2002) Structure-based design of a leucine zipper protein with new DNA contacting region, *Biochemistry 41*, 2177−2183.
44. Turner, E. C., Cureton, C. H., Weston, C. J., Smart, O. S., and Allemann, R. K. (2004) Controlling the DNA binding specificity of bHLH proteins through intramolecular interactions, *Chem. Biol. 11*, 69−77.
45. Hope, I. A., and Struhl, K. (1985) GCN4 protein, synthesized in vitro, binds HIS3 regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast, *Cell 43*, 177−188.
46. Metallo, S. J., Paolella, D. N., and Schepartz, A. (1997) The role of a basic amino acid cluster in target site selection and non-specific binding of bZIP peptides to DNA, *Nucleic Acids Res. 25*, 2967−2972.
47. Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties*, 2nd ed., W. H. Freeman and Co., New York.
48. Hollenbeck, J. J., McClain, D. L., and Oakley, M. G. (2002) The role of helix stabilizing residues in GCN4 basic region folding and DNA binding, *Protein Sci. 11*, 2740−2747.
49. Spolar, R. S., and Record, M. T., Jr. (1994) Coupling of local folding to site-specific binding of proteins to DNA, *Science 263*, 777−784.
50. Metallo, S. J., and Schepartz, A. (1994) Distribution of labor among bZIP segments in the control of DNA affinity and specificity, *Chem. Biol. 1*, 143−151.
51. Cram, D. J. (1988) The design of molecular hosts, guests, and their complexes, *Angew. Chem., Int. Ed. Engl. 27*, 1009−1020.
52. Wedemayer, G. J., Patten, P. A., Wang, L. H., Schultz, P. G., and Stevens, R. C. (1997) Structural insights into the evolution of an antibody combining site, *Science 276*, 1665−1669.
53. Eaton, B. E., Gold, L., and Zichi, D. A. (1995) Let's get specific: the relationship between specificity and affinity, *Chem. Biol. 2*, 633−638.

BI050121H