Minimalist, Alanine-Based, Helical Protein Dimers Bind to Specific DNA Sites

Ajay R. Lajmi, Michael E. Lovrencic, Timothy R. Wallace, Robyn R. Thomlinson, and Jumi A. Shin*

Department of Chemistry, University of Pittsburgh Pittsburgh, Pennsylvania 15260

Received August 20, 1999

The wealth of information about protein−DNA interactions demonstrates the tremendous complexity and subtlety Nature uses in molecular recognition. We hypothesize that we can exploit what Nature has evolved by manipulating the α-helical basic region/leucine zipper motif (bZIP) as a molecular recognition scaffold. We designed preorganized, alanine-based bZIP proteins that retain the α-helical structure and DNA-binding function of the wild-type bZIP. We chose the bZIP because it is the smallest, simplest protein structure that binds specific DNA sequences.[1,2] In our most highly mutated bZIP, 24 of 27 amino acids are alanine; this protein retains helical structure and DNA-binding sequence specificity similar to the wild-type. Our strategy suggests that dimeric, alanine-based α-helical proteins may be designed that recognize the DNA major groove with high specificity and affinity.

Our design concentrates on mutants of the well-characterized bZIP protein GCN4.[3−5] We reduced the elegantly minimal bZIP structure by substitution with alanines (Ala) to afford a generic, helical scaffold. Of the naturally occurring amino acids, Ala possesses the highest propensity for forming and stabilizing α-helical structures.[6,7] Interestingly, both NMR and circular dichroism (CD) demonstrate that, while the leucine zipper is intrinsically stable and helical, the basic region remains only loosely helical until binding to a specific DNA target.[8−11] Thus, the basic region of bZIP proteins requires site-specific DNA binding to achieve stability and helicity; this energetic requirement may be circumvented by alanine scanning mutagenesis.

We substituted alanines into the basic regions of bacterially expressed GCN4 bZIP derivatives containing GCN4 basic region residues 226−252 (Figure 1). The GCN4 bZIP-DNA cocrystal structures show that only 4 highly conserved amino acids per basic region monomer make direct contacts to DNA bases: Asn[235], Ala[238], Ala[239], and Arg[243].[3−5,12] The basic region mutant with the highest Ala content, **18A**, retains only these 4 amino acids from native GCN4, plus Lys[246], which is believed to improve protein solubility.[5] wt (wild-type) is the "native" variant comprising the GCN4 basic region and the C/EBP leucine zipper. The GCN4−C/EBP fusion was demonstrated to bind to GCN4-binding sites as tightly and specifically as the native GCN4 bZIP.[13,14] Our protein constructs contain a 6-histidine tag at the N terminus that aids in purification.[15,16] **4A** and **11A** contain 4 and 11 Ala substitutions, respectively: both specific interactions with DNA bases and nonspecific electrostatic interactions with the DNA phosphodiester backbone are maintained.[3−5] **11A** is also mutated in the hinge region between the leucine zipper and basic region; the hinge is important for spacing the basic region monomers properly on the DNA site.[17] In the **18A** mutant, only base-specific interactions are conserved; note that only 3 of 27 amino acids in the **18A** basic region are non-alanine, and electrostatic protein−DNA interactions have been virtually abolished.

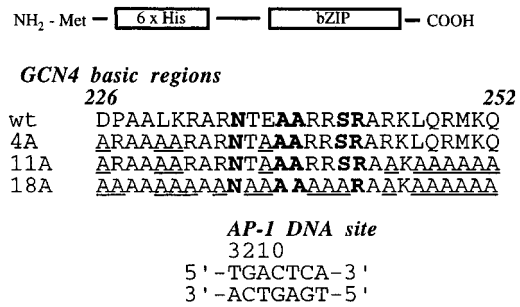

Figure 1. (Top) Schematic of expressed protein. The bZIP domains comprise the basic region mutants of GCN4 (residues 226−254), leucine zipper from C/EBP (residues 312−338) plus ∼35 residues from the pTrcHis B expression vector. Molecular weights of expressed proteins were confirmed by MALDI-TOF (Middle) Sequences of the bZIP domains. **4A**, **11A**, and **18A** are the same as wt, except for the mutated basic regions. Alanine substitutions are underlined; highly conserved residues are in bold. The leucine zipper lies at the carboxyl termini of the basic regions. (Bottom) Sequence of the AP-1 DNA site. Numbering begins at the central CG base pair.

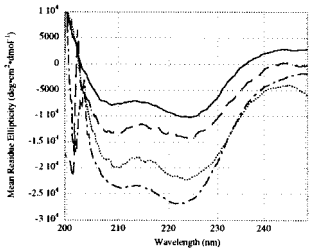

Figure 2. Circular dichroism on wt (solid line), 4A (- - -), 11A (·····) and 18A (·-·-·-·) proteins. 1 mM protein dimer was placed in 20 mM phosphate buffer, pH 7.5, 100 mM NaCl and equilibrated for 15 min at 25 °C prior to CD measurement.

CD demonstrates that protein α-helicity increases with increasing Ala content (Figure 2). Because we conducted our CD measurements well below the apparent dimerization constants for the proteins (see Supporting Information),[8] we conclude that the increased protein helicity is due to enhanced stabilization of the basic region rather than enhanced dimerization through the alanine-substituted interface. Mean residue ellipticity values at Θ_{222} for these mutants may be compared to Θ_{222} for an ideal α-helix, calculated to be −37,500 deg·cm²·dmol⁻¹.[18] wt and **4A** have intrinsic helical character of 27 and 38%, respectively, whereas **11A** and **18A** possess substantially more helicity of 59 and 71%, respectively. Therefore, increasing Ala content in the bZIP basic region generates proteins of higher α-helical stability with potentially more favorable energetics for binding to DNA.

(1) Struhl, K. *Trends Biochem. Sci.* **1989**, *14*, 137.
(2) Landschulz, W. H.; Johnson, P. F.; McKnight, S. L. *Science* **1988**, *240*, 1759.
(3) König, P.; Richmond, T. J. *J. Mol. Biol.* **1993**, *233*, 139.
(4) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223.
(5) Keller, W.; König, P.; Richmond, T. J. *J. Mol. Biol.* **1995**, *254*, 657.
(6) O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250*, 646.
(7) Luque, I.; Mayorga, O. L.; Freire, E. *Biochemistry* **1996**, *35*, 13681.
(8) O'Neil, K. T.; Shuman, J. D.; Ampe, C.; DeGrado, W. F. *Biochemistry* **1991**, *30*, 9030.
(9) Saudek, V.; Pasley, H. S.; Gibson, T.; Gausepohl, H.; Frank, R.; Pastore, A. *Biochemistry* **1991**, *30*, 1310.
(10) Weiss, M. A.; Ellenberger, T.; Wobbe, C. R.; Lee, J. P.; Harrison, S. C.; Struhl, K. *Nature* **1990**, *347*, 575.
(11) Shin, J. A. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2367.
(12) Johnson, P. F. *Mol. Cell. Biol.* **1993**, *13*, 6919.
(13) Agre, P.; Johnson, P. F.; McKnight, S. L. *Science* **1989**, *246*, 922.
(14) We use the C/EBP zipper because in related experiments, we affix our proteins to solid support by diazotization through tyrosine; histidine may interfere with the diazotization reaction. See ref 11. The GCN4 zipper contains a histidine, whereas C/EBP does not.
(15) Hochuli, E.; Bannwarth, W.; Döbeli, H.; Gentz, R.; Stüber, D. *Biotechnology* **1988**, 1321.
(16) We were concerned about the His tag; however, similar constructs were generated by Abate et al. who expressed the bZIP domains of Fos and Jun with 6-His tags at the N termini and found no interference with Jun−Fos interaction with the AP-1 site (see Abate, C.; Luk, D.; Gentz, R.; Rauscher, F. J., III; Curran, T. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 1032).
(17) We note that amino acid 227 is arginine in both **4A** and **11A**; this is a cloning artifact, and this residue has no interaction with DNA (see refs 3−5).
(18) Scholtz, J. M.; Qian, H.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers* **1991**, *31*, 1463.

10.1021/ja993025a CCC: $19.00
Published on Web 05/23/2000

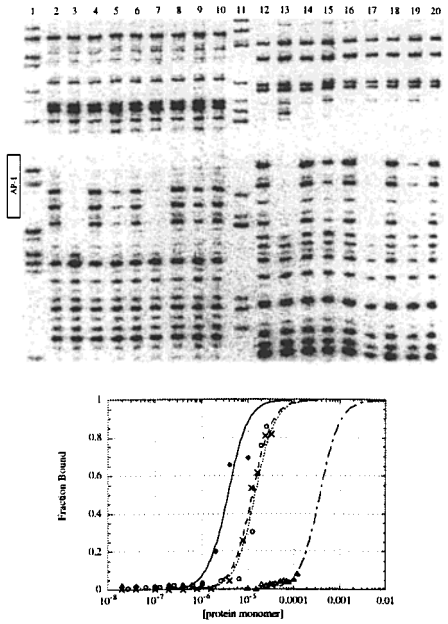

Figure 3. (Top) DNAse I footprinting reactions on wt, **4A**, **11A**, and **18A** proteins bound to the AP-1 DNA site. Lanes 1−10 present data for 3′ end-labeled DNA; lanes 11−20 5′ end-labeled DNA. Lanes 1 and 11, chemical sequencing G reaction;[30] lanes 2 and 12, DNAse I cleavage control. Lanes 3−10 and 13−20, DNAse I cleavage reactions. Lanes 3 and 13, 2 μM wt; lanes 4 and 14, 0.2 μM wt. Lanes 5 and 15, 2 μM **4A**; lanes 6 and 16, 0.2 μM **4A**. Lanes 7 and 17, 5 μM **11A**; lanes 8 and 18, 0.5 μM **11A**. Lanes 9 and 19, 20 μM **18A**; lanes 10 and 20, 2 μM **18A**. The bar on the left indicates AP-1 site. (Bottom) The fraction of AP-1 DNA bound (54 base-pair duplex) as a function of protein concentration: (◆) wt; (×) 4A; (○) 11A; (Δ) 18A. Binding reactions were performed at 37 °C in 20 mM Tris, pH 7.5, 4 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, 0.5 mM EDTA, 5% glycerol, 0.5 mg/mL BSA, and 1 mM DTT.

DNAse footprinting and electrophoretic mobility shift assay (EMSA) show that all three Ala-mutated proteins retain DNA-binding specificity for the AP-1 site (Figure 3). Therefore, all mutants retain α-helical structure (Figure 2) and DNA-binding function (Figure 3).[19] EMSA was used to determine apparent dissociation constants of protein dimers to AP-1.[20] The wt monomer has an apparent K_d of 0.4 μM for AP-1, whereas **4A** and **11A** bind with 7-fold decreased affinities at 2.5 and 3.1 μM, respectively. In **18A**, 24 of 27 amino acids in the basic region are Ala, yet **18A** still retains specificity for the AP-1 site with apparent dissociation constant of ∼490 μM, despite the loss of virtually all electrostatic contacts.[21] We note these modest K_d values, which we attribute to high concentrations of urea denaturant (1−1.5 M urea). High urea concentrations were necessary to overcome protein aggregation (see Supporting Information); it also causes denaturation of protein and DNA structure, and a decrease in binding affinities is expected.

CD indicates that **18A** is maximally helical in the bZIP domain (the expressed proteins contain an extra ∼35 residues from the pTrcHis vector). Thus, the **18A** is a preorganized helix that will not pay a folding penalty (entropy) upon DNA binding. Numerous electrostatic interactions have been removed, however, in the **18A** cocomplex with DNA; thus, the enthalpic contribution to binding has been diminished. **18A** retains binding specificity to AP-1, but affinity decreases 1000-fold; preorganization (entropy) does not entirely compensate for loss of electrostatic interactions (enthalpy). In contrast, the binding affinities of **4A** and **11A** for the AP-1 site are diminished only 7-fold from that of wt; 11A's α-helicity approaches that for **18A**, so preorganization is beneficial to the binding energetics, but **11A** also maintains specific and nonspecific protein−DNA interactions. **11A** appears to be an optimal balance of preorganizational/entropic vs electrostatic/enthalpic components.

The DNA-binding region of **18A** is prefolded and notably hydrophobic, in contrast to the wild-type GCN4 DNA-binding region, which is disordered and highly cationic. Hydrophobic protein−DNA interactions contribute to binding specificity and affinity. Ser^{242} makes a nonspecific phosphodiester interaction, as well as a hydrophobic contact between the Ser side chain methylene unit and the C_5 methyl group on T_3.[4] **18A** substitutes Ser^{242} with Ala, which is still capable of maintaining van der Waals contact with the thymine methyl group but loses the hydrogen bond to the DNA backbone. Struhl and co-workers have shown that maintaining hydrophobic interactions, especially those at Ala^{238} and Ala^{239}, can greatly influence GCN4 recognition of DNA.[22,23] Alanine scanning has been performed on the basic helix-loop-helix/leucine zipper (bHLH/ZIP) protein TFEB,[24] which also uses a dimer of α-helices to bind specific sites in the DNA major groove.[25−27] As many as 12 alanines could be substituted into the 18-residue TFEB basic region, and binding to the E box sequence 5′-CACGTG is maintained.[24] More stable helices can be generated with alanine replacements, but Nature may employ the α-helical folding transition to enhance regulation of cellular processes.

Only 4 residues per bZIP monomer make base-specific contacts, and a few more make nonspecific electrostatic interactions with the DNA backbone.[3−5,28] These amino acids can be considered "responsible" for bZIP function, that is, sequence-specific recognition of the DNA major groove. For the minimalist bZIP, the backbone α-helix can be maintained by alanines. Thus, those amino acids necessary for maintenance of the α-helix can be considered responsible for bZIP structure. The GCN4 bZIP may rely on the highly conserved Asn^{235}, Ala^{238}, Ala^{239}, and Arg^{243} to provide binding specificity, whereas binding affinity may be maintained, at least partly, by a preorganized protein helix (structure) in lieu of basic residues making electrostatic contacts to DNA (function). Structure and function have been successfully dissected by other approaches; Zondlo and Schepartz have shown that DNA-binding proteins can be generated by grafting those residues from GCN4 required for DNA recognition onto a miniature protein scaffold.[29] Our experiments suggest that judiciously placed amino acids embedded within a stable α-helix may constitute the minimal requirements for sequence-specific, high-affinity DNA recognition.

Acknowledgment. We appreciate helpful discussions with Jennifer Goeckeler and Jeffrey Brodsky. MALDI-TOF was obtained at the Center for Molecular Analysis, Carnegie Mellon University. J.A.S. is grateful for support from NSF CAREER MCB-9733410, University of Pittsburgh, and University of Pittsburgh Medical Center.

Supporting Information Available: Protocols for protein cloning, expression, and purification (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA993025A

(19) We note that 20 μM **18A** does not appear to produce equally strong DNAse footprints on the 3′-radiolabeled DNA strand (Figure 3, lane 9) as on the 5′-radiolabeled strand (Figure 3 lane 19). This may be due to differences in the degree of protection that **18A** affords to either strand in the major groove.

(20) Metallo, S. J.; Schepartz, A. *Chem. Biol.* **1994**, *1*, 143.
(21) Poor solubility of **18A** precludes obtaining data at high protein concentrations. This limitation necessitated extrapolation of the binding curve.
(22) Pu, W. T.; Struhl, K. *Mol. Cell. Biol.* **1991**, *11*, 4918.
(23) Kim, J.; Tzamarias, D.; Ellenberger, T.; Harrison, S. C.; Struhl, K. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4513.
(24) Takemoto, C.; Fisher, D. E. *Gene Expression* **1995**, *4*, 311.
(25) Ferre-D'Amare, A. R.; Prendergast, G. C.; Ziff, E. B.; Burley, S. K. *Nature* **1993**, *363*, 38.
(26) Ma, P. C. M.; Rould, M. A.; Weintraub, H.; Pabo, C. O. *Cell* **1994**, *77*, 451.
(27) Ferre-D'Amare, A. R.; Pogonec, P.; Roeder, R. G.; Burley, S. K. *EMBO J.* **1994**, *13*, 180.
(28) Glover, J. N. M.; Harrison, S. C. *Nature* **1995**, *373*, 257.
(29) Zondlo, N. J.; Schepartz, A. *J. Am. Chem. Soc.* **1999**, *121*, 6938.
(30) Maxam, A.; Gilbert, W. *Methods Enzymol.* **1980**, *65*, 499.