Protein Expression and Purification **18,** 394–403 (2000)
doi:10.1006/prep.2000.1209, available online at http://www.idealibrary.com on IDEAL®

Short, Hydrophobic, Alanine-Based Proteins Based on the Basic Region/Leucine Zipper Protein Motif: Overcoming Inclusion Body Formation and Protein Aggregation during Overexpression, Purification, and Renaturation

Ajay R. Lajmi, Timothy R. Wallace, and Jumi A. Shin[1]
Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Received December 21, 1999, and in revised form January 13, 2000

GCN4 is a yeast transcriptional regulatory protein; its DNA-binding domain is a basic region/leucine zipper (bZIP) structure that comprises a dimer of α-helices capable of high-affinity, sequence-specific recognition of the DNA major groove. We are exploiting what nature has evolved by manipulating the bZIP motif as a molecular recognition scaffold; thus we reduced the elegantly minimal bZIP to an even more simplified structure by substitution with alanine residues—hence, a generic, Ala-based, helical scaffold. These Ala-based mutants are unusual proteins for expression as they are short (~100 amino acids) and hydrophobic (Ala-mutated basic regions, leucine-zipper dimerization domains). Hydrophobicity posed a major problem throughout the expression, isolation, and purification stages; inclusion body formation and protein aggregation were significant hurdles throughout protein production. We describe measures that solved these problems, including use of high concentrations of denaturant in all steps of protein isolation and purification and use of temperature-dependent renaturing techniques to obtain folded, functional protein. Despite these difficulties, we ultimately retrieved 5–10 mg/L of broth of active, correctly folded protein after the complete purification procedure. Homogeneity of the proteins was established by chromatography, electrophoresis, and mass spectrometry. Furthermore, characterization by circular dichroism and DNase footprinting analysis demonstrates that these alanine-based mutants retain the structure and function of the native GCN4 DNA-binding domain. Remarkably, the most heavily mutated protein, containing 24 alanines of 27 total amino acids in the DNA-binding basic region, still binds the AP-1 site, the target of native GCN4.

GCN4 is a dimeric transcriptional regulatory protein that governs histidine biosynthesis in yeast under conditions of amino acid starvation (1). The full-length GCN4 monomer is 281 amino acids, and the basic region/leucine zipper (bZIP[2]) structure comprises a dimer of 60-residue monomers. The bZIP structure is responsible for the protein dimerization and DNA-recognition activities of GCN4. Cocrystal structures of the bZIP domain of GCN4 bound to two different DNA sites (2–4), as well as the Jun–Fos heterodimer bZIP–DNA cocrystal (5), show that a continuous α-helix of over 50 amino acids provides both the interface for binding to specific DNA sites and the leucine zipper coiled-coil dimerization structure (Fig. 1). The positively charged basic regions interact with DNA, while the amphipathic leucine zippers dimerize through the hydrophobic leucine interface.

Our work aims to contribute to understanding the structural and functional aspects of recognition of the DNA major groove by α-helical bZIP proteins. Notably, nature's use of the protein α-helix for specific DNA recognition is ubiquitous; other helical protein motifs—including the helix–turn–helix (HTH), zinc finger, and basic helix–loop–helix (bHLH)—have also been found to be responsible for the high affinity and sequence specificity intrinsic to a number of DNA-binding proteins (6). We have chosen to examine the α-helical bZIP motif, as it is the simplest protein structure that recognizes specific DNA sites with high binding affinity. Additionally, the bZIP cocrystal structures demonstrate astonishing conservation of protein backbone structure between the two yeast GCN4 (2–4) and the avian Jun–Fos structures (5); the protein scaffold is

[1] To whom correspondence should be addressed. Fax: (412) 624-4255. E-mail: jumi+@pitt.edu.

[2] Abbreviations used: bZIP, basic region/leucine zipper protein; CD, circular dichroism; SEC, size-exclusion chromatography; HPLC, high-performance liquid chromatography; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight.

1046-5928/00 $35.00

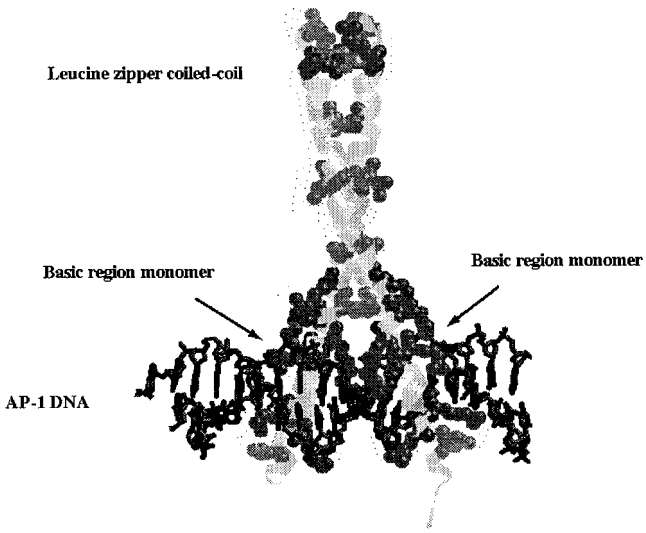
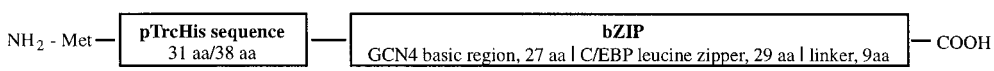


FIG. 1. GCN4 bZIP in complex with the AP-1 DNA site (3). DNA is the dark double helix at the bottom of the figure, and the bZIP is the paler, α-helical dimer at the top. The leucine zipper dimerizes into the coiled-coil structure shown at the top of the figure; the helical zipper then smoothly forks to either side of the DNA, thus allowing the helical basic region dimer to bind opposite sides of the DNA major groove.

composed of extremely regular and predictable α-helices that lie similarly in the major groove. Thus, the simplicity and tractability of the bZIP make it the ideal system for molecular design and analysis of binding specificity and affinity.

If we can begin to understand how nature uses a pair of α-helices to bind DNA, we then may ask how DNA can be recognized by the bZIP structure. To this end, we generated proteins with desired DNA recognition capabilities from a core, alanine-based scaffold. These Ala-based proteins serve to begin investigation of the minimal protein determinants for sequence-specific, high-affinity recognition of the DNA major groove. Of the naturally occurring amino acids, alanine possesses the highest propensity for forming and stabilizing α-helical protein structures (7,8). Interestingly, the bZIP basic region is disordered until binding specific DNA sites: 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 sequence (9–12). Thus, the basic region of bZIP proteins requires site-specific DNA binding to achieve stability and helicity, and this energetic requirement may be circumvented by alanine scanning mutagenesis.

We substituted alanines into the basic regions of bacterially expressed GCN4 bZIP derivatives comprising GCN4 basic region residues 226–252. These proteins are α-helical dimers based on the bZIP scaffold wherein the leucine zipper hails from C/EBP and the basic region comprises Ala-based derivatives of GCN4. The wild-type GCN4 derivative and alanine mutants are shown in Fig. 2; alanine replacements are underlined. wt (wild type) is the "native" variant comprising the GCN4 basic region and the C/EBP leucine zipper.



pTrcHis B sequence

MGGSHHHHHHGMASMTGGQQMGRDLYDDDDK (31 aa; wt)
MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPDDDDK (38aa; 4A, 11A, 18A)

bZIP domain

GCN4 basic region *C/EBP leucine zipper* *linker*
DPAALKRARNTEAARRSRARKLQRMKQ-LEQKVLELTSDNDRLRKRVEQLSRELDTL-GGCGGYYYY

GCN4 basic regions

	226	*252*
wt	DPAALKRARNTEAARRSRARKLQRMKQ	
4A	ARAAAARARNTAAARRSRARKLQRMKQ	
11A	ARAAAARARNTAAARRSRAAKAAAAAA	
18A	AAAAAAAAANAAAAAAARAAKAAAAAA	

FIG. 2. (Top) Schematic of expressed protein. Basic region/leucine zipper (bZIP) proteins were cloned into expression vector pTrcHis B (Invitrogen), which contains a six-histidine tag for protein purification and an enterokinase cleavage site (DDDDK). The bZIP is at the carboxyl termini of the expressed proteins, which is the same positioning of the bZIP domain in native GCN4. The fully expressed bZIP comprises approximately 35 residues from the pTrcHis expression vector, the basic region mutants of GCN4 (residues 226–252), leucine zipper from C/EBP (residues 312–338), and a linker (GGCGGYYYY) for covalent protein dimerization (cysteine) and chemical derivatization (tyrosines) (12). (Middle) Sequences of pTrcHis B and bZIP domain sequences. The pTrcHis sequences for 4A, 11A, and 18A contain an additional enterokinase site and are therefore longer than that for wt. (Bottom) Sequences of the bZIP domains. The sequences for alanine mutants 4A, 11A, and 18A are shown below wt; these proteins are the same as wt, except for the mutated basic regions. Alanine substitutions are underlined. We note that Pro 227 is arginine in both 4A and 11A; this is a cloning artifact, and this residue has no interaction with DNA (2–4).

Our basic regions are fused to the C/EBP leucine zipper at the same junction where Agre *et al*. fused the GCN4 basic region to the C/EBP zipper (residues 312–338) (13). The GCN4–C/EBP fusion was demonstrated to bind to GCN4-binding sites as tightly and specifically as the native GCN4 bZIP (13,14).

The GCN4 bZIP–DNA cocrystal structures show that only four amino acids in each bZIP basic region monomer make direct contacts to bases in the DNA major groove: Asn^{235}, Ala^{238}, Ala^{239}, and Arg^{243} (2–4). These four amino acids are also highly conserved among bZIP proteins (15). Our basic region mutant with the highest Ala content, 18A, retains only these four amino acids from native GCN4, plus Lys^{246} due to concerns about solubility of hydrophobic proteins (Fig. 2); the refined cocrystal structure of the GCN4 bZIP with the ATF/CREB DNA site shows that Lys^{246}, which lies in the hinge region between the leucine zipper and the basic region, is involved in a water-mediated hydrogen-bonding network in the major groove (4). Note that only 3 of 27 amino acids in the 18A basic region are non-alanine. 4A and 11A contain 4 and 11 Ala substitutions, respectively: in these proteins, both specific interactions with DNA bases and nonspecific electrostatic interactions with the DNA phosphodiester backbone are maintained (2–4). 11A is also mutated in the hinge region between the leucine zipper and the DNA-binding basic region; the hinge region is important for spacing the basic region monomers properly on the DNA site.

These Ala-based mutants are unusual proteins for expression in that they are short (~100 amino acids) and hydrophobic (Ala-mutated basic regions, leucine-zipper dimerization domains). Hydrophobicity was a significant issue throughout the expression and purification stages. We describe overcoming major problems with inclusion body formation and protein aggregation; this includes use of high concentrations of denaturant in all steps of protein isolation and purification. After chromatographic purification, protein stock solutions had a strong tendency to form aggregates when stored in less than 4 M urea. We therefore utilized the temperature-leap tactic that aids in protein renaturation from concentrated stock solutions (16). Despite inclusion body formation and protein aggregation during the purification process, we ultimately were able to retrieve 5–10 mg/L of broth of active, correctly folded protein after the complete purification procedure. Additionally, when using our expression protocol, we did not observe any protease degradation of these short proteins. The genes encoding the proteins were synthesized by the gene assembly/gene amplification method, also known as the recursive-PCR procedure (17). In this procedure, a duplex DNA template was formed by specifically annealing two partially overlapping long oligonucleotides by thermal denaturation and renaturation steps, followed by chain extension from the 3′-hydroxyl end using thermophilic polymerase. The duplex product was amplified by thermal cycling in the presence of excess amounts of short terminal primers. Finally, we characterized the wt and Ala mutant proteins by circular dichroism to demonstrate that these proteins assume the properly folded helical structure, and we performed DNase footprinting analysis to show that these mutant proteins also retain sequence-specific DNA-binding function.

MATERIALS AND METHODS

Materials

[γ-^{32}P]ATP was supplied by Amersham, and radioactivity was monitored on a Beckman LS 6500 scintillation counter. Water was purified through a Milli Q filtration system (Millipore). Reagents were purchased from Aldrich or Acros/Fisher; enzymes were purchased from New England Biolabs and Promega.

Gene Assembly and Gene Amplification by PCR

DNA oligonucleotides were synthesized at the DNA Synthesis Facility, University of Pittsburgh (PE Biosystems Expedite 8909). All oligonucleotides over 50 bases in length were purified by nondenaturing polyacrylamide gel electrophoresis (18). The gel (15% acrylamide, 1:29 cross-link) was run for ~3 h at 160 V in 1× TBE buffer; DNA bands in the gel were visualized with a UV lamp and excised. Oligonucleotides were eluted from the gel by crushing and soaking in oligo elution buffer (0.1% SDS, 0.5 M NH_4OAc, 10 mM $Mg(OAc)_2$) at 37°C overnight. The eluted DNA was desalted on a NAP-25 gel column (Pharmacia), eluted with water, and dried under vacuum.

Genes for expression of bZIP proteins were constructed by mutually primed synthesis (19,20) for gene construction, followed by polymerase chain reaction with terminal primers for gene amplification (17,21,22). Gene assembly and amplification were performed in two separate polymerase chain reactions on a Perkin-Elmer GeneAmp 2400. Gene Jockey II software (Biosoft) was used in design of oligonucleotides and determination of annealing temperatures. Overlapping oligonucleotides ranging from 55 to 110 bases in length were used to assemble the genes by mutually primed synthesis, and strand hybridization occurred through 21-base-pair annealing regions. The genes for each protein were constructed from two unique overlapping oligonucleotides: wt (61 and 55 bases), 4A (82 and 55 bases), 11A (82 and 55 bases), and 18A (82 and 55 bases). Each gene was assembled by an initial gene-assembly PCR that contained 8 pmol of each oligonucleotide, 2.5 U of *Vent* DNA polymerase (New England Biolabs), and 250 μM each of the four dNTPs in a total volume of 100 μL of PCR buffer (10 mM KCl, 20 mM Tris, pH 8.8, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1%

Triton X-100). The gene-assembly PCR began with a hot start at 80°C followed by 5 cycles of 1 min at 95°C, 1 min at 59°C, and 1 min at 72°C. The PCR was completed with a 5-min incubation at 72°C. The PCR products were purified by nondenaturing polyacrylamide gel electrophoresis (8% acrylamide, 1:29 cross-link). The gel was run for ~3 h at 160 V, and bands were visualized with a UV lamp and excised. The duplex DNA was eluted from the gel by crushing and soaking in elution buffer (0.5 M NH_4OAc, 10 mM $Mg(OAc)_2$, 1 mM EDTA, 0.1% SDS) at 37°C overnight. The eluted DNA was desalted on a NAP-25 column, eluted with water, and dried under vacuum.

Ten microliters of the gene-assembly reaction mixture was then transferred for gene-amplification PCR with terminal oligonucleotide primers (17,21,22). These second polymerase chain reactions contained the same cocktail as above but also included 100 pmol each of the 20-base terminal primers. The gene-amplification PCR began with a hot start at 80°C followed by 25 cycles of 1 min at 90°C, 1 min at 45°C, and 1 min at 72°C. The PCR was completed with a 5-min incubation at 72°C. Final PCR products were purified by nondenaturing polyacrylamide gel electrophoresis (8% acrylamide, 1:29 cross-link). The gel ran ~3 h at 160 V, and bands were visualized with a UV lamp and excised. The duplex DNA was eluted by crushing and soaking as described in the previous paragraph and then cloned into protein expression vectors.

The genes for the basic regions and leucine zipper were joined by restriction and ligation at the designed *Xho*I site. The synthesized genes were inserted into the *Bam*HI and *Eco*RI sites of vector pTrcHis B (Invitrogen); *E. coli* strain BL21(DE3) (Stratagene) was transformed by electroporation (Bio-Rad *E. coli* Gene Pulser) with the recombinant pTrcHis plasmids. Cloned inserts were sequenced by dideoxy-DNA sequencing (T7 Sequenase kit, USB).

Bacterial Expression of bZIP Proteins

Initial protein expression screenings were performed to determine if the clones produced any recombinant protein. Samples were taken before and 8–9 h after induction, resuspended in loading buffer, and loaded onto an SDS–PAGE gel to determine whether overexpression had occurred. Selected clones were stored at −80°C in 20% glycerol (w/v).

Bacterial strain BL21(DE3) expressing pTrcHis recombinant plasmids was grown in 12 mL of LB medium containing 50 μg/mL ampicillin for ~16 h at 30°C; this overnight culture was then added to 1 L of the same medium and grown at 30°C until late log phase (OD_{600} ~1.4). IPTG was added to a final concentration of 1 mM, and the cells were grown for an additional 9 h at 30°C. Cells were harvested (Beckman J2-HC centrifuge), washed once with water, spun down in a centrifuge, and resuspended in ~20 mL of sonication buffer (50 mM Hepes, pH 7.4, 300 mM NaCl, 10 mM imidazole, 5 mM 2-mercaptoethanol) containing 1 mM PMSF (Acros) and 0.5 μg/mL pepstatin (Boehringer Mannheim). The cells were sonicated five times for 1 min with a 2-min incubation on ice between sonications (Sonic Dismembrator 60, Fisher Scientific). Sonicated cells were centrifuged in 50-mL conical vials (Corning) at 9000*g* for 15 min, and the supernatant was recentrifuged at 13,000*g* for 15 min to further clarify the supernatant.

PAGE Analysis of Expressed Proteins

Cell pellets to be analyzed by polyacrylamide gel electrophoresis were thawed and resuspended in 100 μL of 10% glycerol, 2% SDS, 0.05% bromophenol blue, 0.05 M Tris, pH 6.8, and 15% 2-mercaptoethanol. The cell-pellet mixtures were spun down, and 5 μL of each lysate was loaded onto a Bio-Rad 16.5% Tris–Tricine Ready Gel in 1× Tris–Tricine buffer. The gel was run at 100 V for 80 min. It was then fixed in 40% methanol/10% acetic acid for 20 min at 30°C, shaking at 40 rpm. The gel was stained in 0.025% Coomassie brilliant blue staining solution (10% acetic acid, 0.25 g/L Coomassie brilliant blue) for 1 h. The gel was washed three times for 15 min each in 10% acetic acid, followed by a final wash in 10% acetic acid until the gel achieved the desired color intensity. The gel was set between cellophane support sheets and clamped into Tut's Tomb (Idea Scientific, Minneapolis, MN) for drying.

Western Blot Analysis

After a standard PAGE was run as described in the previous paragraph, a transfer sandwich was prepared. A nitrocellulose transfer and immobilization membrane (Protran, Schleicher & Schuell) was placed on top of the gel and sandwiched between two sheets of Whatman filter paper that had been presoaked in transfer buffer (25 mM Tris, 200 mM glycine, 20% methanol, 0.11% SDS, ~pH 8.3). The gel sandwich was placed in a Mini Trans-Blot electrophoretic transfer cell (Bio-Rad). The proteins were transferred to the nitrocellulose membrane at 0.36 A for 1.5 h in transfer buffer. The membrane was then soaked in Ponceau S solution (Sigma) for 2 min, rinsed with water until the bands were visible, and marked with a pencil for future reference. The membrane was incubated with Blocko (2% BSA in 5 mM sodium azide, 50 mM Tris, pH 7.4, 150 mM NaCl, 0.2% v/v Tween-20) for 20 min. The membrane was then incubated with 6 μg of mouse Penta-His antibody (Qiagen) in Blocko overnight at 4°C. The membrane was washed with TBST (50 mM Tris, pH 7.4, 150 mM NaCl, 20% v/v Tween 20) three times for 5 min each. The membrane was incubated with anti-mouse IgG horseradish peroxidase (HRP) linked antibody (Amersham, 5:3000 dilution) for 1.5 h

in Blotto (TBST plus 2% nonfat dry milk). The membrane was washed three times for 5 min each in TBST. HRP was activated with the Super Signal Substrate Western blotting kit (Pierce) and immediately exposed to Kodak X-OMAT film and developed.

Protein Purification on TALON

The harvested cells were resuspended in 20 mL of sonication buffer (20 mM Tris, pH 8.0, 300 mM NaCl, 10 mM 2-mercaptoethanol, 10% glycerol, 6 M guanidinium-HCl) containing 1 mM PMSF (Acros) and 0.5 μg/mL pepstatin (Boehringer Mannheim). The cells were sonicated 8–10 times for 1 min each with a 2-min incubation on ice between sonications. Sonicated cells were spun down in a centrifuge at 9000*g* for 15 min and the supernatant was spun down at 13,000*g* for 15 min. The resulting supernatant was added to 3 mL of TALON cobalt metal affinity resin slurry suspension (Clontech) and gently agitated on a platform rocker at room temperature for 30 min. Separation of the 6×His-tagged proteins was performed using the batch/gravity flow method (Clontech). The proteins were eluted with four bed volumes (12 mL) of elution buffer (20 mM Tris, pH 8.0, 300 mM NaCl, 100 mM imidazole, 10 mM 2-mercaptoethanol, 10% glycerol, 8 M urea) in 1-mL fractions. All washes and elutions contained 1 mM PMSF and 0.5 μg/mL pepstatin. Fractions were analyzed by SDS–PAGE (16.5% Tris–Tricine Ready Gel, Bio-Rad) followed by Coomassie brilliant blue staining and immunoblot analysis using anti-His_5 antibody (Qiagen). Peak fractions were pooled and concentrated by Amicon centrifugation (Ultrafree 15, Millipore) at 4°C and further purified on a size-exclusion chromatography column at room temperature.

Protein Purification by Size-Exclusion Chromatography (SEC)

The fractions eluted by metal ion affinity chromatography (TALON) were analyzed by SDS–PAGE and immunoblot analysis; those fractions containing 6×His-tagged proteins were consolidated and concentrated to 0.1- to 0.2-mL total volume. In order to solubilize proteins, up to 8 M urea was added. Approximately 0.1–0.2 mL (8–10 mg) of concentrated TALON-purified protein in 8 M urea solution was loaded onto a Superdex 75 HR 10/30 column (Pharmacia) using a manual injector (0.2-mL loop, Upchurch Scientific) on a Beckman System Gold HPLC. Proteins were eluted from the column using 10% acetonitrile in buffer (50 mM phosphate buffer, pH 6.8, 150 mM NaCl, 4 M urea) over 340 min at 0.05 mL/min. The proteins were stored with 1 mM PMSF and 1 μg/mL pepstatin at −20°C.

Temperature-Leap Renaturation of Proteins

Protein stocks were stored at 4°C; only the amount of protein stock to be used for that day's experiments was renatured to active form following the temperature-leap tactic developed by Xie and Wetlaufer (16). Protein concentrations were determined by UV/vis absorbance (Beckman DU 640 spectrometer) at 275 nm for tyrosine ($\epsilon_{275} = 1405\ M^{-1} \cdot cm^{-1}$ per tyrosine). Refolding was performed by rapid manual dilution of purified, denatured proteins in 8 M urea stocks to the desired protein concentrations in dilution buffer (20 mM phosphate buffer, pH 7.5, 100 mM NaCl, 1 M urea). All plastic containers were preequilibrated at the desired temperatures before mixing. The purified protein solutions to be used for CD and DNase footprinting studies were prepared by dilution of an appropriate amount of the purified stocks. The dilutions were performed by adding the required amount of dilution buffer at 4°C and then incubating the diluted protein at 4°C for 2 h, followed by rapid heating to 37°C for 1 h. This protein was then immediately used in CD and footprinting experiments.

Circular Dichroism

CD was performed on a JASCO J-710 with a jacketed, Suprasil, 1-mm-path length cell (Hellma). The CD was blanked using the standard D-camphor-10-sulfonic acid. The optimized parameters were as follows: sensitivity, 20 mdeg; response, 8 s; scan speed, 50 nm/min; step resolution, 0.2 nm; scale, 300–175 nm; accumulation, 5 (average of five readings at the same wavelength).

Protein dimer was placed in 200 μL of 20 mM phosphate buffer, pH 7.5, 100 mM NaCl such that the final concentration of protein dimer was 5 μM and equilibrated for 10 min at 16°C prior to CD measurement.

Radiolabeling of Restriction Fragments

Plasmid pUC19 was linearized by digestion with *Hin*dIII. Linearized pUC19 was 5′-end-labeled by dephosphorylation with calf intestinal alkaline phosphatase, followed by phosphorylation using [γ-^{32}P]ATP and T4 polynucleotide kinase (23). Labeled linearized plasmid pUC19 was digested with *Ssp*I, and the resulting radiolabeled ~650-bp DNA fragment was isolated using agarose gel electrophoresis.

DNase Footprinting

Reaction mixtures (10-μL total volume) contained ^{32}P-end-labeled DNA fragment (~20,000 cpm), bZIP dimer protein, TKMC buffer (20 mM Tris, pH 7.5, 4 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$), 0.1 mg/mL tRNA (Sigma Chemical, Type XX), and 0.17 μg/mL DNase I (Boehringer Mannheim, grade I). All components except DNase I were incubated for 1 h at 22°C. Footprinting reactions were initiated by addition of DNase I and allowed to proceed for 2 min at 22°C. Reactions were terminated by addition of 2 μL of DNase stop solution

(3 M NH_4OAc, 250 mM EDTA), followed by phenol/chloroform extraction, ethanol precipitation, drying, and resuspension in formamide loading buffer (18). Reaction products were analyzed by electrophoresis on 8% polyacrylamide denaturing gels (GIBCO BRL Gel-Mix 8, 1:20 cross-linked, 7 M urea) run at 1000–2000 V. After electrophoresis, gels were dried and autoradiographed.

Mass Spectrometry

MALDI-TOF mass spectrometry was performed on a PerSeptive Biosystems Voyager Elite at the Center for Molecular Analysis, Carnegie Mellon University. Proteins were dissolved at a concentration of 20 pmol/μL in 1:1 water:acetonitrile with 0.1% trifluoroacetic acid and loaded onto a 100-well sample plate (PerSeptive). Equal parts of sample and matrix (α-cyano-4-hydroxycinnamic acid, Aldrich) were used.

RESULTS

Design and Assembly of Synthetic Genes for Protein Expression

The synthetic bZIP genes were designed to avoid palindromic and repetitive sequences that could lead to intramolecular hairpin formation or mispriming, respectively; to minimize G/C content to aid efficient gene transcription and sequencing; and to bias toward those codons whose tRNAs are of high concentration in *E. coli* (21). As only a subset of all possible codons is employed by *E. coli* during highly efficient protein expression, codon usage was largely restricted to the two most commonly used *E. coli* codons for each amino acid (24,25).

Initial assembly of the gene was performed by mutually primed synthesis (19,20). As shown in Fig. 3, the basic regions and leucine zipper were constructed separately; each segment comprised two unique overlapping oligonucleotides with an annealing region of 21 base pairs capable of mutually primed synthesis. Each duplex segment was purified by polyacrylamide gel electrophoresis. Aliquots of this purified duplex were then amplified by polymerase chain reaction, with 18-base terminal primers representing the outermost flanking sequences of each duplex. Again, the amplified duplexes were purified by PAGE. The basic regions and leucine zipper were designed to be joined by restriction with *Xho*I followed by ligation. The full bZIP genes were inserted into the *Bam*HI and *Eco*RI sites of vector pTrcHis B, which provides a six-histidine tag that aids in protein purification. *E. coli* strain BL21(DE3) was transformed by electroporation with the recombinant pTrcHis B plasmids, and the identities of the cloned inserts were ascertained by dideoxy-DNA sequencing.

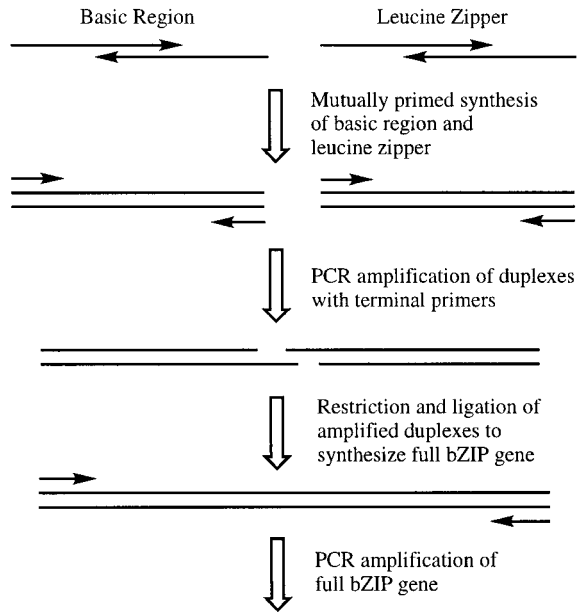


FIG. 3. Gene assembly and amplification. Overlapping oligonucleotides ranging from 55 to 110 bases in length were used to assemble the genes by mutually primed synthesis, and strand hybridization occurred through 21-base-pair annealing regions. Twenty-base terminal primers were used for PCR amplification of duplexes. Direction of PCR-catalyzed chain extension is indicated by arrows.

Optimization of Protein Expression

The genes for expression of bZIP proteins were cloned into both the pRSET B and pTrcHis B vectors, and the levels of expression were examined for both vectors transformed into BL21(DE3) cells. Overexpression was induced with 1 mM IPTG. Pilot studies using timed samples during the course of protein expression showed very low levels of expression of these proteins in pRSET B vector, which uses the T7 promoter for protein expression; however, excellent expression levels were obtained in pTrcHis B vector, which uses the *trc* promoter, a hybrid derived from the *trp* and *lac* promoters. Other than the difference in promoters, the pRSET and pTrcHis vectors are extremely similar.

Expression of our short, hydrophobic proteins worked poorly when transformed cells were grown in LB medium at 37°C; virtually all of the protein remained insoluble in the cell pellet, and still remained largely insoluble even after sonication in high concentrations of denaturant. Thus, protein yields were very low. We therefore optimized our protein expression protocol by conducting growth and induction at 30°C rather than 37°C. The lysis buffer contained 6 M guanidine in order to solubilize the expressed proteins. Most of the protein was now found in the soluble fraction of the cell lysate, and this dramatically increased our protein yields; even more protein could be retrieved by resolubilizing and extracting the cell pellet in 6 M

guanidine. In general, 6 M guanidine was found to be a more powerful chaotrope for our bZIP proteins than 8 M urea. Growth at 30°C aided expression of soluble bZIP proteins and also minimized protease activity, a potential problem when expressing such short proteins.

Optimization of Protein Purification

The hydrophobicity of the bZIP proteins posed a significant problem in the expression, isolation, and purification protocols. After solving difficulties with inclusion body formation during protein expression, we found aggregation still remained an issue during purification. Because these proteins were expressed with a six-histidine tag, an initial purification on immobilized metal affinity resin was performed. Protein solubilized in 6 M guanidine (see previous section) was loaded onto TALON resin, which uses octahedral cobalt ion with coordination sites for the histidine tag. We initially tried nickel-based metal affinity resins, but we discovered that these resins bound significant amounts of cellular, non-His-tagged proteins. TALON cobalt metal affinity resin was much more selective for purification of our His-tagged proteins, and protein purity and yields were vastly superior with TALON. Protein aggregation necessitated the use of 8 M urea throughout TALON purification in order to maintain solubility of bZIP proteins. The fractions eluted from metal affinity chromatography were analyzed by SDS–PAGE and immunoblot analysis; those fractions containing His-tagged proteins were consolidated. As can be seen in the SDS–PAGE in Fig. 4, protein is substantially cleaner after TALON purification.

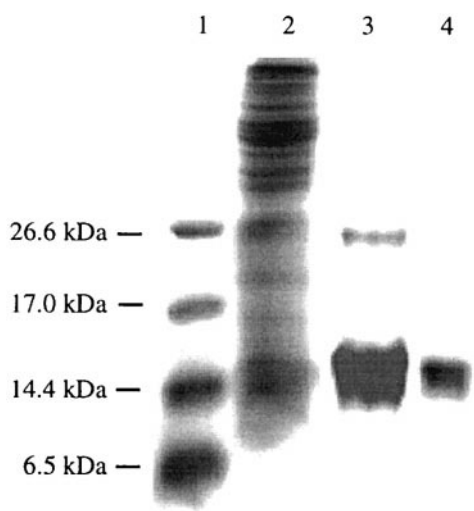


FIG. 4. Expression and purification of wt bZIP protein analyzed by SDS–PAGE (16.5% Tris–Tricine Ready Gel, Bio-Rad) stained with Coomassie brilliant blue. Lane 1, molecular weight marker; lane 2, crude extract; lane 3, after purification on TALON metal affinity resin; lane 4, after purification by size-exclusion chromatography.

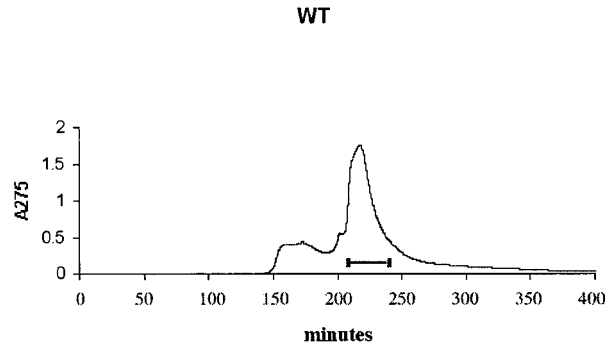


FIG. 5. Purification of wt bZIP on a Superdex 75 HR 10/30 size-exclusion chromatography column using 10% acetonitrile in buffer (50 mM phosphate buffer, pH 6.8, 150 mM NaCl, 4 M urea) over 340 min at 0.05 mL/min. The large peak (marked with horizontal bar) in the above chromatogram was saved as purified wt bZIP.

After a first purification on TALON resin, the proteins were further purified on a size-exclusion column (SEC) using 10% acetonitrile in phosphate buffer containing 4 M urea. Initially, we tried using ion-exchange chromatography, but only marginal yields of protein were obtained. Our proteins were more successfully purified by SEC. Although only the wt bZIP is shown in the SDS–PAGE in Fig. 4 and the chromatogram shown in Fig. 5, all four bZIP proteins were successfully purified by these procedures (Table 1), and the SEC chromatograms looked very similar for all proteins. The SDS–PAGE shows our proteins being expressed with apparent molecular weights of approximately 14 kDa, whereas the actual molecular weights are around 11 kDa. There is also a protein contaminant that shows up at around 27 kDa in the SDS–PAGE; this protein is not present in the Western blot, indicating that it is not a His-tagged protein, but more likely a cellular protein contaminant. MALDI-TOF mass spectrometry was performed on the purified wt bZIP protein; the calculated molecular weight was 11,073 Da, and the observed molecular weight was 11,079 Da. SDS–PAGE and Western blot analyses both showed no protease degradation of freshly expressed and purified bZIP proteins; over time, however, protease degradation was found to be problematic, so proteins were stored with 1 mM PMSF and 1 μg/mL pepstatin at −20°C.

TABLE 1

Purification of bZIP Proteins

	70 g of wet cells per liter of broth	
Purification step	Total protein (mg)	Recovery (%)
Crude extract	11.2	100
TALON affinity column	10.1	90
Size-exclusion column	8.0	79

Temperature-Leap Renaturation of Proteins

After purification by SEC, protein concentrations were assessed by UV/vis measurement of tyrosine at 275 nm. Protein stocks were stored at −20°C in the same buffer used in SEC containing 1 mM PMSF and 1 μg/mL pepstatin. We found that diluting an aliquot of this concentrated protein stock caused protein aggregation and precipitation, as the 4 M urea in the buffer is diluted as well. Therefore, only the amount of protein stock to be used for that day's experiments was renatured to active form following the temperature-leap tactic developed by Xie and Wetlaufer (16). The purified protein solutions to be used for CD and DNase footprinting studies were prepared by dilution of an appropriate amount of the purified stocks. The dilutions were performed by adding the required amount of buffer (the buffer used in either CD or footprinting experiments) at 4°C and incubating the diluted protein at 4°C for 2 h, followed by rapid heating to 37°C for 1 h. This protein was then immediately used in CD and footprinting experiments.

Characterization of bZIP Structure and Function by Circular Dichroism and DNase Footprinting Analysis

Circular dichroism (CD) is an excellent method for characterizing the structure of bZIP proteins, as the α-helix displays distinctive minima at 208 and 222 nm. We can also take advantage of the fact that the native GCN4 basic region is intrinsically unstable and disordered (9–12), and therefore, CD can be utilized to monitor changes in bZIP helical structure. Because alanine has the highest propensity for forming and stabilizing α-helices (7,8), our alanine mutant bZIP proteins should become increasingly helical in structure as alanine content increases; the CD data in Fig. 6 affirm this expectation. Calculations show that 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 · cm^2 · dmol^{-1} (26). On the basis of this calculated ideal value, wt bZIP and 4A have comparable intrinsic helical character of 34 and 37%, respectively, whereas 11A and especially 18A possess substantially more helicity of 51 and 79%, respectively. Note that 18A possesses approximately twice the helical content of wt, indicating that the 18A basic region is a preorganized α-helix in comparison with the more disordered wt basic region.

As expected, the alanine mutant bZIPs become more stable and helical as alanine content increases. This favorable energetic feature in bZIP structure may assist in DNA-binding function. DNase footprinting analysis presented in Fig. 7 shows that all of our bZIP proteins bind specifically to the AP-1 DNA site (5′-TGACTCA-3′), which is the *in vivo* target site of native GCN4 in yeast. The wt bZIP comprises the GCN4 DNA-binding domain and the C/EBP leucine zipper; as shown by Agre *et al.* (13), this fusion hybrid still retains GCN4 DNA-binding function, and our wt bZIP strongly footprints at 1 μM concentration. 4A bZIP also footprints well at 1 μM, and increasing the 11A concentration to 5 μM gives a strong footprint. The concentration of 18A must be increased 20-fold over that of wt bZIP to give a strong footprint; thus, 20 μM 18A footprints approximately as well as lower concentrations of the other proteins. We used a ~650-bp restriction fragment in our footprinting experiments, and we note that the mutants consistently footprinted only at the AP-1 site and that their DNA-binding pattern mimicked that of wt bZIP. Therefore, despite elimination of numerous protein–DNA interactions by alanine mutagenesis, these alanine-based bZIP mutants still retain the sequence-specific DNA-binding function of native GCN4.

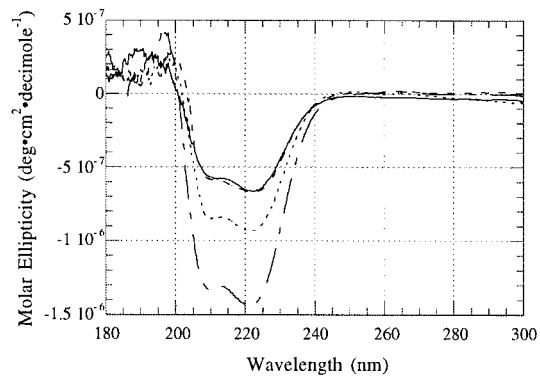


FIG. 6. Circular dichroism on wt (—), 4A (· - ·), 11A (· · ·), and 18A (– — –) proteins. Final concentration of each protein dimer was 5 μM in 20 mM phosphate buffer, pH 7.5, 100 mM NaCl. Measurements were performed at 25°C.

DISCUSSION

We are examining mutants of the GCN4 bZIP dimer to seek the minimal determinants for a dimer of α-helices to recognize the major groove; our strategy has been to generate minimalist, alanine-based helical proteins capable of binding specific DNA sites. To this end, we synthesized the genes for expression of bZIP mutants comprising the C/EBP leucine zipper joined to mutant basic regions; we used mutually primed synthesis (21) and recursive PCR methods (21) to synthesize the genes. Our design allows versatility of mixing and matching different leucine zippers and basic regions by restriction with *Xho*I endonuclease. The proteins generated therefore contain a full bZIP protein structure with a histidine tag at the amino terminus.

Expression of these hydrophobic proteins in pRSET vector in BL21(DE3) *E. coli* cells was unsuccessful; time-course samples taken during pilot induction studies showed no production of protein even after 6 h of

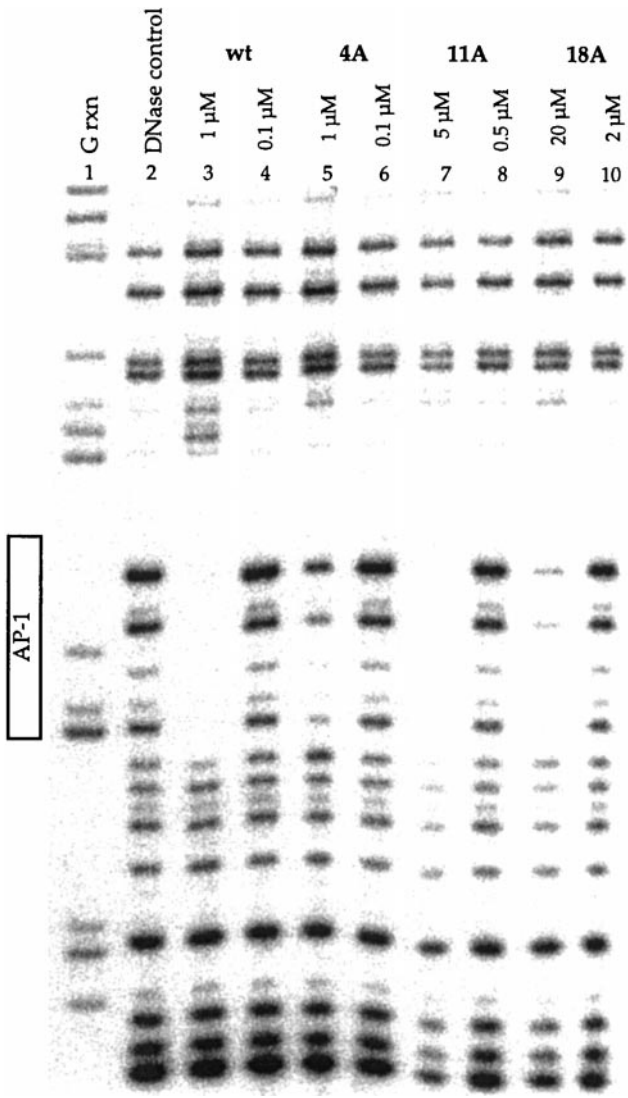


FIG. 7. Autoradiogram of a high-resolution denaturing polyacrylamide gel of DNase I footprinting reactions on wt, 4A, 11A, and 18A proteins bound to the AP-1 DNA site. Data presented for 5′-end-labeled DNA. Lane 1, chemical sequencing G reaction (23); lane 2, DNase I cleavage control. Lanes 3–10, DNase I cleavage reactions in the presence of various concentrations of protein. Lane 3, 1 μM wt; lane 4, 0.1 μM wt. Lane 5, 1 μM 4A; lane 6, 0.1 μM 4A. Lane 7, 5 μM 11A; lane 8, 0.5 μM 11A. Lane 9, 20 μM 18A; lane 10, 2 μM 18A. The box drawn on the left of the autoradiogram indicates the AP-1 site.

induction with IPTG. However, excellent yields of 8–40 mg per liter of all four proteins were obtained by overexpression in pTrcHis B vector in BL21(DE3) cells at 30°C. One might have expected intracellular protease degradation of these relatively short proteins (~100 amino acids), but this we did not observe. Inclusion body formation was a tremendous problem during protein expression; however, inclusion bodies can protect short proteins from proteolysis during expression (27), and perhaps these problematic inclusion bodies served some positive function. Interestingly, the bZIP mutants with higher alanine content, 11A and 18A, did not exhibit any higher tendency for inclusion body formation or aggregation in comparison to wt bZIP; we suspect that the hydrophobic leucine zipper may play the major role in protein aggregation.

SDS–PAGE analysis after purification of these His-tagged recombinant proteins using Ni-NTA metal affinity resin (Qiagen) gave the desired bZIP protein products with substantial levels of three high molecular weight protein contaminants (>20 kDa). Purification of the same His-tagged proteins on cobalt-based TALON resin yielded only a small amount of one high molecular weight band (~27 kDa) in addition to the desired proteins. TALON resin therefore proved to be more efficient in purifying these hydrophobic proteins. Size-exclusion chromatography (SEC) was the secondary purification method of choice as the highly hydrophobic nature of these proteins resulted in the loss of product during ion-exchange chromatography. Acetonitrile (10%) and 4 M urea were necessary to obtain high yields from the SEC column. Even more critical was use of high concentrations of denaturant, either 6 M guanidine or 8 M urea, throughout all isolation and purification steps.

It is noteworthy that in 18A, 24 of 27 amino acids in the basic region are Ala (i.e., three non-Ala residues), yet the hydrophobic 18A mutant still retains DNA-binding specificity to the AP-1 site, despite loss of virtually all electrostatic contacts. DNase footprinting analysis shows that qualitatively, wt bZIP and 4A strongly footprint at the AP-1 site at the same concentration, whereas 11A binds the AP-1 site well upon increasing protein concentration fivefold. The heavily mutated 18A still binds specifically to AP-1, albeit with reduced binding affinity. Therefore, all three alanine mutants retain α-helical structure, as shown by CD, and DNA-binding function, as shown by DNase footprinting.

Increasing alanine content in the bZIP basic region generates proteins of higher α-helical stability, with potentially more favorable energetics for binding to DNA. Much more stable helices can be generated with alanine replacements, but nature may employ the α-helical folding transition to enhance regulation of cellular processes. By substituting alanines into GCN4′s basic region, we hope to uncover the minimal recognition elements for GCN4–DNA complex formation and to exploit the bZIP structure as a molecular recognition scaffold from which we may design helical protein dimers capable of tight-affinity recognition of desired DNA sequences.

ACKNOWLEDGMENTS

We are grateful to Michael Lovrencic and Robyn Thomlinson for scientific expertise and helpful discussion and to Jennifer Goeckeler and Jeffrey Brodsky for helpful discussion on protein expression. We

appreciate support from NSF CAREER Grant MCB-9733410 to J.A.S. and the University of Pittsburgh.

REFERENCES

1. Hill, D. E., Hope, I. A., Macke, J. P., and Struhl, K. (1986) Saturation mutagenesis of the yeast *his3* regulatory site: Requirements for transcriptional induction and for binding by GCN4 activator protein. *Science* **234,** 451–457.
2. König, 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.
3. 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.
4. Keller, W., König, P., and Richmond, T. J. (1995) Crystal structure of a bZIP/DNA complex at 2.2 Å: Determinants of DNA specific recognition. *J. Mol. Biol.* **254,** 657–667.
5. 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.
6. Pabo, C. O., and Sauer, R. T. (1992) Transcription factors: Structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61,** 1053–1095.
7. O'Neil, K. T., and DeGrado, W. F. (1990) A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* **250,** 646–651.
8. Luque, I., Mayorga, O. L., and Freire, E. (1996) Structure-based thermodynamic scale of α-helix propensities in amino acids. *Biochemistry* **35,** 13681–13688.
9. O'Neil, K. T., Shuman, J. D., Ampe, C., and DeGrado, W. F. (1991) DNA-induced increase in the α-helical content of C/EBP and GCN4. *Biochemistry* **30,** 9030–9034.
10. 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.
11. 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.
12. Shin, J. A. (1997) Specific DNA binding peptide derivatized solid support. *Bioorg. Med. Chem. Lett.* **7,** 2367–2372.
13. Agre, P., Johnson, P. F., and McKnight, S. L. (1989) Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP. *Science* **246,** 922–926.
14. We use the C/EBP zipper because in related experiments, we covalently affix our proteins to solid support by diazotization through tyrosine; histidine may, however, interfere with the diazotization reaction. See Ref. 12. The GCN4 zipper contains a histidine, whereas the C/EBP zipper does not.
15. Johnson, P. F. (1993) Identification of C/EBP basic region residues involved in DNA sequence recognition and half-site spacing preference. *Mol. Cell. Biol.* **13,** 6919–6930.
16. Xie, Y., and Wetlaufer, D. B. (1996) Control of aggregation in protein refolding: The temperature-leap tactic. *Protein Sci.* **5,** 517–523.
17. Casimiro, D. R., Wright, P. E., and Dyson, H. J. (1997) PCR-based gene synthesis and protein NMR spectroscopy. *Structure* **5,** 1407–1412.
18. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Press, Cold Spring Harbor, NY.
19. Oliphant, A. R., and Struhl, K. (1987) The use of random-sequence oligonucleotides for determining consensus sequences. *Methods Enzymol.* **155,** 568–582.
20. Oliphant, A. R., Nussbaum, A. L., and Struhl, K. (1986) Cloning of random-sequence oligodeoxynucleotides. *Gene* **44,** 177–183.
21. Casimiro, D. R., Toy-Palmer, A., Blake, R. C., II, and Dyson, H. J. (1995) Gene synthesis, high-level expression, and mutagenesis of *Thiobacillus ferrooxidans* rusticyanin: His 85 is a ligand to the blue copper center. *Biochemistry* **34,** 6640–6648.
22. Prytulla, S., Dyson, H. J., and Wright, P. E. (1996) Gene synthesis, high-level expression and assignment of backbone ^{15}N and ^{13}C resonances of soybean leghemoglobin. *FEBS Lett.* **399,** 283–289.
23. Maxam, A., and Gilbert, W. (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65,** 499–560.
24. Kane, J. F. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli. Curr. Opin. Biotechnol.* **6,** 494–500.
25. Ikemura, T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* **146,** 1–21.
26. Scholtz, J. M., Qian, H., York, E. J., Stewart, J. M., and Baldwin, R. L. (1991) Parameters of helix–coil transition theory for alanine-based peptides of varying chain lengths in water. *Biopolymers* **31,** 1463–1470.
27. Rudolph, R., and Lilie, H. (1996) *In vitro* folding of inclusion body proteins. *FASEB J.* **10,** 49–56.