Gregory H. Bird
Ajay R. Lajmi
Jumi A. Shin*
Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260

Received 8 February 2002; accepted 19 April 2002

Sequence-Specific Recognition of DNA by Hydrophobic, Alanine-Rich Mutants of the Basic Region/ Leucine Zipper Motif Investigated by Fluorescence Anisotropy

Abstract: *We generated minimalist proteins capable of sequence-specific, high-affinity binding of DNA to probe how proteins are used and can be used to recognize DNA. In order to quantify binding affinities and specificities in our protein–DNA system, we used fluorescence anisotropy to measure in situ the thermodynamics of binding of alanine-rich mutants of the GCN4 basic region/leucine zipper (bZIP) domain to DNA duplexes containing target sites AP-1 (5′-TGACTCA-3′) or ATF/CREB (5′-TGACGTCA-3′). We simplified the α-helical bZIP molecular recognition scaffold by alanine substitution:* **4A**, **11A**, *and* **18A** *contain four, eleven, and eighteen alanine mutations in their DNA-binding basic regions, respectively. DNase I footprinting analysis demonstrates that all bZIP mutants retain the sequence-specific DNA-binding function of native GCN4 bZIP. Titration of fluorescein-labeled oligonucleotide duplexes with increasing amounts of protein yielded low nanomolar dissociation constants for all bZIP mutants in complex with the AP-1 and ATF/CREB sites: binding to the nonspecific control duplex was > 1000-fold weaker. Remarkably, the most heavily mutated protein* **18A**, *containing 24 alanines in its 27-residue basic region, still binds AP-1 and ATF/CREB with dissociation constants of 15 and 7.8 nM, respectively. Similarly, wild-type bZIP binds these sites with* K_d *values of 9.1 and 14 n*M. **11A** *also displays low nanomolar dissociation constants for AP-1 and ATF/CREB, while* **4A** *binds these sites with ~ 10-fold weaker* K_d *values. Thus, both DNA-binding specificity and affinity are maintained in all our bZIP derivatives. This Ala-rich scaffold may be useful in design and synthesis of small α-helical proteins with desired DNA-recognition properties capable of serving as therapeutics targeting transcription.* © 2002 Wiley Periodicals, Inc. Biopolymers 65: 10–20, 2002

Keywords: *basic region/leucine zipper; GCN4; fluorescence anisotropy; hydrophobic; alanine-scanning mutant; protein–DNA binding affinity*

Correspondence to: Jumi A. Shin, email: jshin@utm.utoronto.ca
**Current address:* Department of Chemistry, University of Toronto, Mississauga, ON L5L 1C6, Canada
Contract grant sponsor: National Science Foundation (NSF) and the University of Pittsburgh
Contract grant number: MCB-9733410 (NSF)
Biopolymers, Vol. 65, 10–20 (2002)

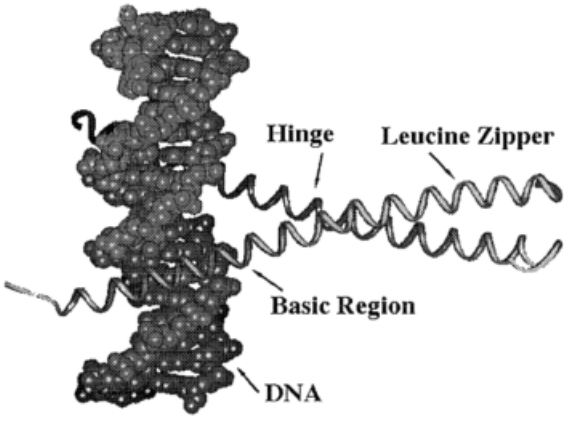

FIGURE 1 GCN4 bZIP in complex with the AP-1 DNA site, 5′-TGACTCA.[11] DNA is the vertical double helix at the left of the figure, and the bZIP is the horizontal α-helical dimer. The leucine zipper dimerizes into the coiled-coil structure shown at the right of the figure; the helical zipper then smoothly forks to either side of the DNA major groove.

INTRODUCTION

Our work aims to contribute to understanding the relationship between a protein's structure and its DNA-binding function. We exploit the protein α-helix, a structure used ubiquitously for sequence-specific DNA recognition, and one that chemists have successfully used in design and synthesis studies for many years (examples include Refs. 1–6). The α-helix is maximally utilized by the basic region/leucine zipper (bZIP) structure for protein dimerization and DNA recognition. In order to begin to probe protein–DNA recognition, we ask how DNA can be recognized by the bZIP structure: Can we create a minimalist α-helical protein structure that retains desired DNA-binding function?

To begin investigation of the minimal protein determinants for sequence-specific, high-affinity recognition of the DNA major groove, we generated proteins with α-helical structure and DNA-recognition capabilities from a core scaffold based on the GCN4 bZIP.[7,8] GCN4 is a dimeric transcriptional regulator that governs histidine biosynthesis in yeast under conditions of amino acid starvation (Figure 1).[9] The full-length GCN4 monomer is 281 amino acids, and the bZIP comprises a dimer of $\sim$ 60-residue monomers. Crystal structures of the GCN4 bZIP domain bound to two different DNA sites,[10–12] as well as the Jun–Fos heterodimeric bZIP-DNA crystal,[13] show that a continuous α-helix of $\sim$ 60 amino acids provides both the basic region for binding specific DNA sites and the leucine zipper coiled-coil dimerization structure.

Of the naturally occurring amino acids, alanine possesses the highest propensity for forming and stabilizing α-helical protein structures.[14,15] Interestingly, the bZIP basic region is disordered until binding to DNA: both NMR and CD demonstrate that while the leucine zipper is intrinsically stable and helical, the basic region remains only loosely helical until binding to DNA.[16–20] This folding transition may enhance control of gene transcription. Thus, the basic region of bZIP proteins requires DNA binding to achieve stability and helicity, and this energetic requirement may be circumvented by design of preorganized Ala-based scaffolds. We substituted alanines into the basic regions of bacterially expressed GCN4 bZIP derivatives comprising GCN4 basic region residues 226–252; the leucine zipper hails from C/EBP, residues 310–338 (Figure 2).[7] The wt (wild-type) bZIP is the "native" variant comprising the GCN4 basic region and C/EBP zipper. These Ala-based mutants are short ($\sim$ 100 amino acids) and hydrophobic (Ala-mutated basic regions, leucine zipper domains).

The GCN4 bZIP-DNA crystal structures show that only four highly conserved amino acids[21] in each basic region monomer make direct contacts to bases in the DNA major groove: Asn^{235}, Ala^{238}, Ala^{239}, and Arg^{243}.[10–12] **4A** and **11A** contain four and eleven Ala substitutions, respectively; both specific interactions with DNA bases and nonspecific Coulombic interactions with the phosphodiester backbone are maintained.[10–12] Additionally, **11A** is mutated in the hinge region, which is important for spacing basic region monomers properly on the DNA site and can affect DNA-binding function.[22] 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.[10] Therefore, of the 27 residues in **18A**'s basic region, 24 are now alanine. Notably, all of our bZIP mutants retained native DNA-binding function as shown by DNase I footprinting.[8] Therefore, small α-helical proteins can be designed to target specific DNA sites.

We used fluorescence anisotropy spectroscopy for measurement of dissociation constants of our bZIP derivatives binding oligonucleotide duplexes (Figure 2). Fluorescence anisotropy measures the tumbling motion of molecules containing a fluorophore; the anisotropic tumbling of our short DNA duplexes increases measurably upon protein binding, as the 5′-fluorescein-labeled duplex is $\sim$ 13 kD and bZIP monomer is $\sim$ 11–12 kD. Not only is the mass of the labeled DNA dramatically changed upon protein binding, but also its rod-like shape; upon complexation, the bZIP radiates perpendicularly from the DNA like a tire spoke, thereby causing more drastic alteration in anisotropic behavior than binding by a

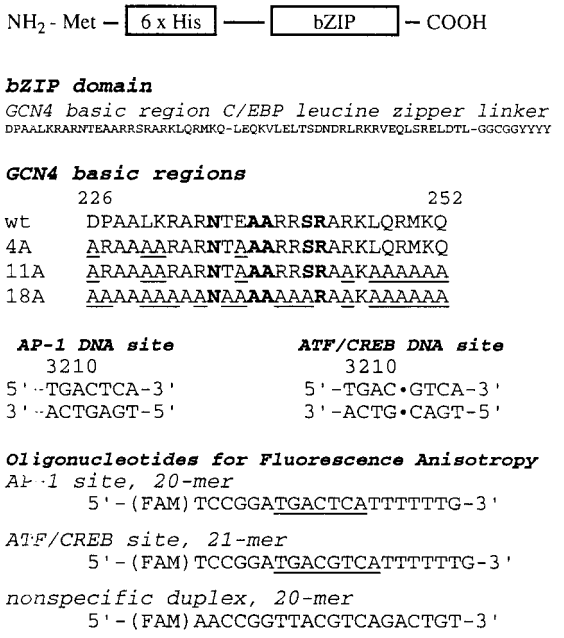

FIGURE 2 Top: Schematic of expressed protein. bZIP proteins were cloned into expression vector pTrcHis B (Invitrogen), which contains a six-histidine tag for protein purification. The bZIP is at the carboxyl termini of the expressed proteins, which is the same positioning of the bZIP domain in native GCN4. The bZIP domains comprise the basic region mutants of GCN4 (residues 226–254), leucine zipper from C/EBP (residues 310–338) plus approximately 35 residues from the pTrcHis B expression vector. Middle: Sequences of the bZIP domains. Sequence of the bZIP domain of the wild-type protein comprises the GCN4 basic region, C/EBP leucine zipper, plus a linker for chemical derivatization. The sequences for alanine mutants **4A, 11A**, and **18A** are shown below wt bZIP; these proteins are the same as wt bZIP, except for the mutated basic regions. Alanine substitutions are underlined and highly conserved bZIP residues are in bold. 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.[10–12] Bottom: Sequences of the AP-1 and ATF/CREB DNA sites. Numbering begins at the central CG base pair. Filled circle denotes division between abutting half sites in ATF/CREB. Sequences of the oligonucleotide duplexes used in fluorescence anisotropy titrations. "FAM" is fluorescein phosphoramidite, and the AP-1 and ATF/CREB sites are underlined.

globular protein of the same mass. True thermodynamic/equilibrium binding vs stoichiometric binding can be achieved due to the sensitivity of fluorescence for detection of fluorescein.[23] Therefore, protein concentration can be maintained in excess throughout the titration (> 25-fold excess protein over DNA in our work).

We have previously shown that these mutants retain α-helical structure and sequence-specific DNA-binding function.[8] In this work, we used fluorescence anisotropy spectroscopy to quantify the binding affinities of wt bZIP, **4A**, **11A**, and **18A** to fluorescein-labeled 20 or 21 base-pair DNA duplexes containing the AP-1 (5′-TGACTCA-3′) or ATF/CREB (5′-TGACGTCA-3′) target sites, as a function of added protein. The dissociation constants of the protein–DNA complexes were measured under equilibrium conditions. K_d values reveal strong, sequence-specific binding in the low nanomolar range. All four proteins displayed similar binding affinities to both the AP-1 and ATF/CREB sites; likewise, the native GCN4 bZIP does not discriminate between AP-1 and ATF/CREB.[18] Here we show that these highly simplified, Ala-rich, α-helical bZIP mutants mimic the high-affinity, sequence-specific DNA-binding function of native GCN4. This minimalist scaffold provides a strategy for design and synthesis of small proteins with desired DNA-binding capabilities.

MATERIALS AND METHODS

Protein Construction

Protocols for DNA oligonucleotide synthesis, gene construction and cloning, protein overexpression, and purification have been described in detail.[7] A brief summary of these procedures follows: Genes for expression of bZIP proteins were constructed by mutually primed synthesis,[24,25] followed by polymerase chain reaction with terminal primers for gene amplification[26–28] and purification by nondenaturing polyacrylamide gel electrophoresis.[29] Duplex DNA was then cloned into protein expression vectors pRSET B and pTrcHis B (Invitrogen); both vectors express proteins with a six-histidine tag for purification purposes. Recombinant plasmids were transformed into *Escherchia coli* strain BL21(DE3) (Stratagene) by electroporation (Bio-Rad). Cloned inserts were sequenced by dideoxy DNA sequencing (T7 Sequenase kit, USB). Bacterial expression of bZIP proteins was performed in Luria-Bertani (LB) medium containing 50 μg/mL ampicillin[29]; induction was initiated with IPTG added to a final concentration of 1 m*M*. Cells were collected by centrifugation and lysed by sonication. These 6xHis-tagged proteins were purified first on TALON cobalt metal-ion affinity resin (Clontech), followed by further purification by size-exclusion chromatography (Superdex 75 HR 10/30 column, Pharmacia) on a Beckman System Gold high performance liquid chromatography (HPLC). Protein purification was monitored by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and Western immunoblot assay. Purified protein stocks were stored in 4*M* urea or guanidine with 1 m*M* α-toluenesolfonyl fluoride (PMSF) and 1 μg/mL pepstatin at −80°C.

Temperature-Leap Renaturation of Proteins

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.[30] Protein solutions with diluted urea were prepared by addition of an appropriate amount of the purified stocks containing 4*M* urea to the relevant buffer at 4°C. The resulting solution was incubated at 4°C for $>$ 2 h, followed by rapid heating to 37°C for 1 h. Temperature-leap was performed for each fluorescence anisotropy data point.

DNase I Footprinting

Plasmid pUC19 was linearized by digestion with *Hin*d III. Linearized plasmid pUC19 was 3′-endlabeled with [α^{32}P]-dATP and DNA polymerase I, Klenow fragment.[31] Labeled linearized plasmid pUC19 was digested with *Ssp* I, and the resulting radiolabeled $\sim$ 650 base pair DNA fragment was isolated by gel electrophoresis. DNase I footprinting reaction mixtures (10 μL total volume) contained ^{32}P-endlabeled DNA fragment ($\sim$ 20,000 cpm), bZIP dimer protein, TKMC buffer (20 m*M* Tris, pH 7.5, 4 m*M* KCl, 2 m*M* $MgCl_2$, 1 m*M* $CaCl_2$), 0.1 mg/mL tRNA (Sigma Chemical, Type XX), 0.17 μg/mL DNase I (Boehringer Mannheim, grade I), and 100–800 m*M* urea (originating from the protein stocks). All components for each reaction, except DNase I, were first combined and the temperature-leap tactic was performed (reactions were incubated at 4°C for 2; 37°C, 1 h). 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 DNase stop solution (3*M* NH_4OAc, 250 m*M* EDTA), followed by phenol/chloroform extraction, ethanol precipitation, drying, and resuspension in formamide loading buffer.[29] Chemical sequencing G reaction was performed as described by Maxam and Gilbert.[31] 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.

DNA Melting Curves

In order to test whether 800 m*M* urea denatures our short fluorescein-labeled duplexes, we monitored the melting curves of the 21-mer fluorescein-labeled ATF/CREB duplex in the presence and absence of 800 m*M* urea. One cuvette contained 116.8 n*M* DNA duplex in fluorescence anisotropy buffer [4.3 m*M* Na_2HPO_4, 1.4 m*M* KH_2PO_4, pH 7.4, 150 m*M* NaCl, 2.7 m*M* KCl, 1 m*M* EDTA, 1 m*M* dithiothreitol (DTT), 20% glycerol, 0.4 mg/mL acetylated bovine serum albumin BSA]; the second cuvette contained the same reagents as the first cuvette plus 800 m*M* urea. DNA melting was monitored at 260 nm with a Cary 3 uv/vis spectrometer (Varian). The cuvettes were incubated at 75°C for 5 min to equilibrate the DNA solutions. Temperature was decreased to 35°C at 1°C/min, and increased from 35 to 75°C at 0.5°C/min. Melts were reversible. Melting temperatures were obtained by the curve-fitting procedure from Marky and Breslauer.[32]

Fluorescence Anisotropy

Fluorescein-labeled AP-1 (20-mer) and ATF/CREB (21-mer) oligonucleotides were synthesized at the Department of Pediatrics, University of Pittsburgh School of Medicine, on an Applied Biosystems DNA Synthesizer 3948 or at the DNA Synthesis Facility, University of Pittsburgh, on a PE Biosystems Expedite 8909. The 6-carboxyfluorescein label (6-FAM, ABI or Glen Research) was incorporated at the 5′-end of the labeled oligonucleotides using phosphoramidite chemistry. All oligonucleotides were purified by SDS-PAGE. AP-1 DNA was hybridized by heating 200 pmols of AP-1/FAM oligonucleotide, and 240 pmols of the unlabeled complementary oligonucleotide in annealing buffer (10 m*M* Tris, pH 7.6, 50 m*M* NaCl, 1 m*M* EDTA) at 80°C for 10 min followed by slow cooling to room temperature over 2 h. The same procedure was followed to hybridize ATF/FAM and nonspecific/FAM with their complementary oligonucleotides.

Fluorescence measurements were taken on an SLM 8000 fluorimeter arranged in the L format (488 nm excitation; 520 nm emission; integration time, 1 s; band pass, 4 nm). Individual measurements of each component of anisotropy ($I_{v,v}$, $I_{v,h}$, $I_{h,v}$, and $I_{h,h}$, where v and h denote the vertical and horizontal polarization components of fluorescein excitation and emission, respectively) were taken. Titrations were performed in a 0.5 mL quartz fluorimetry cell (Starna). The amount of 0.2–4 μL of a 20–30 μM stock protein solution was added per data point, and pipetted up and down ten times to mix in a total volume of 0.5 mL. The fluorimetry cell contained 250 p*M* specific DNA duplex or 1 n*M* nonspecific DNA duplex in buffer (4.3 m*M* Na_2HPO_4, 1.4 m*M* KH_2PO_4, pH 7.4, 150 m*M* NaCl, 2.7 m*M* KCl, 1 m*M* EDTA, 800 m*M* urea, 20% glycerol, 0.4 mg/mL acetylated BSA, 1 m*M* DTT, and 100 μM in base pairs of calf thymus DNA).

At these concentrations of FAM-labeled duplex, protein concentration was in $>$ 25-fold excess for the first titration point, and the bulk of data points were gathered at $>$ 100-fold excess protein over duplex DNA; therefore, equilibrium thermodynamic conditions were maintained throughout these experiments. The temperature-leap tactic described above was performed for every data point in order to maintain soluble protein.

Titrations were performed at 22 $\pm$ 2°C. The volume change was minimized during the titration and was kept to $<$7% of the total buffer volume within the cell. The *G* factor (the ratio of sensitivities of the monochrometer for horizontally and vertically polarized light) was calculated from the above measurements using the equation below[33]:

$$G = I_{h,v}/I_{h,h} \quad (1)$$

The *G* factor remained constant during each experiment; *G*-factor values between different experiments ranged from 0.52 to 0.56. The anisotropy, *r*, was then calculated using the equation below[33]:

$$r = (I_{v,v} - GI_{v,h})/(I_{v,v} + 2GI_{vh}) \quad (2)$$

The anisotropy values were used to calculate the fraction of bound DNA (discussed below). The fraction of bound DNA was plotted as a function of protein concentration, and the data were fit to the Langmuir equation[34] using Kaleidagraph 3.0.8.

Determination of K_d Values

The data from fluorescence anisotropy titrations were fit to a two-state binding equation to determine apparent dissociation constants for each protein–DNA complex. Equation (3) and the treatment of the calculation of dissociation constants is the same as that used by Metallo and Schepartz to determine the dissociation constants of GCN4 basic region derivatives bound to specific DNA sites[34]:

$$\theta_{app} = 1/(1 + K_d^2/[M]^2) \quad (3)$$

where K_d corresponds to the apparent monomeric dissociation constant, and *M* is the concentration of monomeric protein. Metallo and Schepartz discuss that bZIP proteins can bind to DNA by two mechanisms: the bZIP monomers can bind DNA sequentially, wherein one monomer binds, followed by the second monomer, and the bZIP dimerizes on the DNA. Alternatively, the monomers can dimerize in solution prior to binding to DNA; if dimerization is significant, then the concentration of dimer must be accounted for when calculating dissociation constants, and Eq. (3) will not be valid.[34] Because the apparent dimerization constant for the C/EBP leucine zipper is $\sim 20\ \mu M$,[16] the concentration of free protein dimer was insignificant when compared to the total protein concentrations, and therefore Eq. (3) is valid for our case. In our titrations with specific DNA target sites, [*M*] ranged from low-to-middle nanomolar. The **4A** titrations went as high as 900 n*M* protein monomer, whereas the other monomeric protein concentrations went no higher than 20 n*M*. The nonspecific DNA titrations went as high as 10 μM protein monomer; even at these micromolar concentrations of protein, saturation binding of DNA was not achieved. Therefore, dissociation constants for nonspecific binding are more approximate than those for specific binding; minimum K_d values for nonspecific binding are listed in Table I. Curve fits of the fraction of DNA bound vs protein monomer concentration yielded the apparent monomeric dissociation constants of the protein–DNA complexes. Only data sets fit to Eq. (3) with *R* values > 0.980 are reported.

Table I Dissociation Constants for GCN4 bZIP Derivatives Bound to the AP-1, ATF/CREB, or Nonspecific DNA Sites

	K_d ($10^{-9}M$)		
bZIP	AP-1	ATF/CREB	Nonspecific[a]
wt	9.1 ± 1.2	14 ± 1.5	$> 1\ \mu$M
4A	78 ± 5.7	64 ± 5.0	$> 1\ \mu$M
11A	4.8 ± 0.65	5.8 ± 0.53	$> 10\ \mu$M
18A	15 ± 1.3	7.8 ± 1.3	$> 10\ \mu$M

[a] Saturation protein binding was not achieved in any of the nonspecific duplex DNA titrations.

RESULTS

DNase I Footprinting Analysis

DNase I footprinting analysis presented in Figure 3 shows that all of our bZIP proteins bind specifically to the pseudopalindromic AP-1 DNA site (5′-TGACTCA-3′), which is the *in vivo* target site of native GCN4 in yeast, and to the palindromic ATF/CREB site (5′-TGACGTCA), which is recognized by the cAMP-response element binding factor family.[35] Our bZIP proteins comprise GCN4 basic region derivatives and the C/EBP leucine zipper fused at the same junction where Agre et al. made similar hybrids; their GCN4-C/EBP fusion was demonstrated to mimic the DNA-binding function of native GCN4 bZIP.[36] Likewise, our wt bZIP footprints strongly at both the AP-1 and ATF/CREB sites. We use the C/EBP zipper because in related experiments, we covalently link our proteins to solid support by diazotization through tyrosine's *p*-cresol side chain; histidine's imidazole side chain can, however, interfere with the diazotization reaction.[19] The GCN4 zipper contains a histidine, whereas the C/EBP zipper does not. Like wt bZIP, all of the Ala mutants give clear footprints; even at high concentrations, protein is not merely coating DNA nonspecifically (Figure 3). We used a $\sim$ 650 base-pair restriction fragment in the footprinting experiments, and we note that the mutants consistently footprinted only at the AP-1 and ATF/CREB sites, and that their DNA-binding pattern mimicked that of wt bZIP. Therefore, despite elimination and modification of numerous protein–DNA interactions by alanine mutagenesis, these Ala-based bZIP mutants still retain the sequence-specific DNA-binding function of native GCN4.

Our basic region mutant with the highest Ala content, **18A**, retains only the four highly conserved amino acids from the native GCN4 basic region (Asn^{235}, Ala^{238}, Ala^{239}, and Arg^{243}) plus Lys^{246} due to concerns about solubility of hydrophobic proteins. Lys^{246} lies in the hinge region that separates the leucine zipper from the basic region and does not contact DNA; the refined crystal structure of the GCN4 bZIP with ATF/CREB DNA shows that Lys^{246}

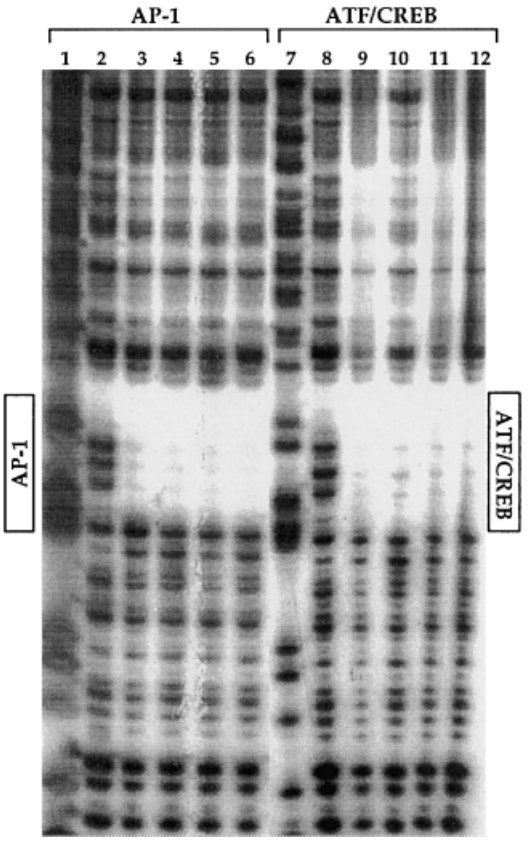

FIGURE 3 Autoradiogram of a high-resolution denaturing polyacrylamide gel of DNase I footprinting reactions on wt bZIP, **4A, 11A**, and **18A** proteins bound to the AP-1 and ATF/CREB DNA sites. Data presented for 3′ endlabeled DNA. Lanes 1–6: footprinting at AP-1 site; lanes 7–12, footprinting at ATF/CREB site. Lanes 1 and 7: chemical sequencing G reaction[31]; lanes 2 and 8, DNase I cleavage control. Lanes 3–6 and 9–12: DNase I cleavage reactions in the presence of various concentrations of protein. Lanes 3 and 9: 5 μM wt bZIP. Lanes 4 and 10: 10 μM **4A**. Lanes 5 and 11: 20 μM **11A**. Lanes 6 and 12: 100 μM **18A**. The bars drawn on the left and right sides of the autoradiogram indicate the AP-1 and ATF/CREB sites, respectively.

is involved in a water-mediated hydrogen-bonding network in the major groove and may improve solubility.[12] For us, protein solubility was not enhanced by Lys[246]. Hydrophobicity was a significant problem throughout the expression and purification stages, and during experimental manipulations.

Fluorescence Anisotropy Titrations Measure Free Energies of Protein–DNA Complexation

Fluorescence anisotropy has been used to study a variety of protein–DNA interactions.[23,37–42] It is particularly well suited to the study of protein binding to small oligonucleotide duplexes, for the nanosecond lifetimes of most fluorophores are comparable to the time for rotation of a molecule or complex of MW $\leq \sim 10^5$, the range of many protein–DNA complexes (the excited state fluorescence lifetime for fluorescein is $\sim$ 4 ns).[33] Additional advantages of this technique are that it is entirely solution-based and allows *in situ* monitoring. Moreover, titrations are very straightforward and convenient to perform, and reproducibility and precision in fluorescence anisotropy measurements make it an excellent methodology for quantitation of protein–DNA interactions.[23] We use fluorescein as the fluorophore covalently attached to DNA, as it can be detected at even picomolar levels and possesses long excitation and emission wavelengths thus minimizing scattered light effects.[43]

We conducted fluorescence anisotropy measurements to generate binding isotherms of all four bZIP proteins bound to the AP-1 20-mer and ATF/CREB 21-mer duplexes (Figure 2). In Table I are listed the dissociation constants, and Figure 4 shows the fitted curves from the titrations. The wt bZIP binds to the AP-1 and ATF/CREB sites with K_d values of 9.1 and 14 nM, respectively. These numbers correlate very well with measurements made by other groups for GCN4 and other bZIP domains.[10,18,20,34,44–47] **4A** is the weakest binder to AP-1 and ATF/CREB, with K_d values of 78 and 64 nM; thus, **4A**'s four alanine substitutions toward the amino terminus of the GCN4 basic region do not enhance DNA-binding affinity. **11A** and **18A** regain binding affinity: **11A** binds to AP-1 and ATF/CREB with dissociation constants of 4.8 and 5.8 nM, and **18A** binds at 15 and 7.8 nM, respectively.

Interestingly, the few alanine replacements in **4A** are enough to weaken its DNA-binding ability. However, more alanine substitutions, as in **11A** and **18A**, regain high-affinity DNA-binding function. These binding results closely parallel those of Takemoto and Fisher, who made Ala replacements in bHLH proteins Myc and TFEB.[48] Likewise, the K_d values of their bHLH mutant–DNA complexes varied within one order of magnitude, as do our values for bZIP mutant–DNA complexes, despite extensive mutations. In the bHLH mutants, increasing Ala mutations in the 18-residue DNA-binding basic region conferred more α-helical stability to the disordered native basic region, which, like bZIP proteins, becomes helical and stable upon sequence-specific DNA binding. Similarly, we observed increasing protein α-helicity as Ala content increases in our bZIP mutants: in previously published CD experiments, we measured the α-helicity of wt bZIP to be 27%, 4A to be 38%, 11A to be

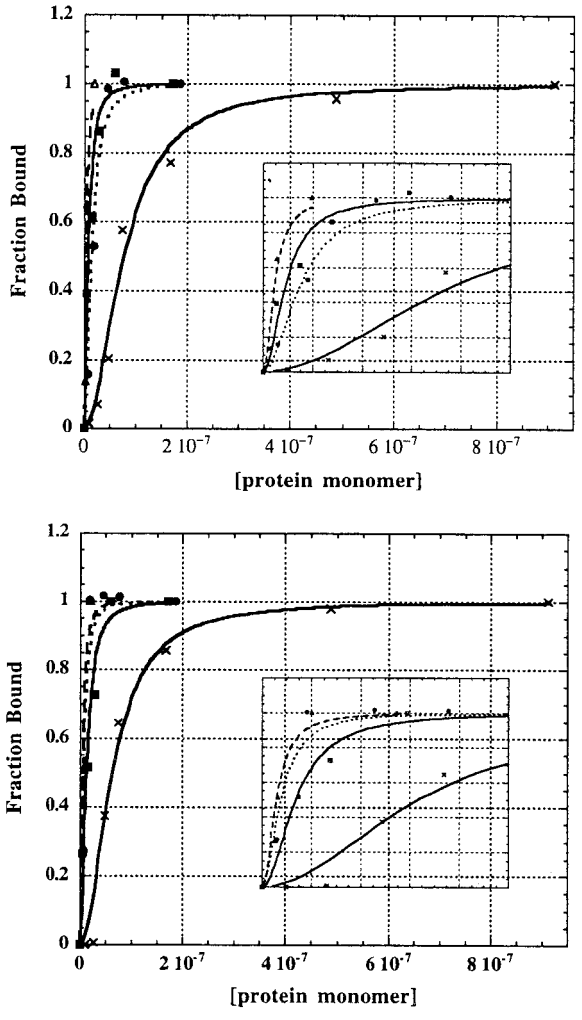

FIGURE 4 Fluorescence anisotropy titrations. Top: Binding of all four mutant bZIP proteins to AP-1 duplex. Bottom: Binding of all four mutant bZIP proteins to ATF/CREB duplex. wt bZIP (■, solid line), **4A** (×, solid line), **11A** (△, dashed line), and **18A** (●, dashed line). Insets show enlargement of data collected at low protein concentrations (0–100 n*M* range).

59%, and 18A to be 71%.[8] At 71% helicity, 18A is fully helical, as there are approximately 30 unstructured amino acids from the expression vector in the 100-residue 18A mutant. DNA-binding affinities do not strictly correlate with Ala content, however; in TFEB mutants containing four or six Ala substitutions, DNA-binding affinity decreased approximately twofold. But TFEB mutants with nine, ten, eleven, and twelve Ala replacements regained DNA-binding affinity; the most Ala-rich mutant A12 bound DNA fivefold more tightly than native TFEB.[48] This binding affinity trend is similar to that for our bZIP mutants, in which **4A** showed weaker DNA-binding than native GCN4 bZIP, but **11A** and **18A** regained native binding ability. Therefore, increasing structural preorganization and increasing DNA-binding affinity do not correlate perfectly; possibly, differences in hydration and entropy of these Ala-mutated basic regions influence the relationship between preorganized structure and DNA-binding function. The results of Takemoto and Fisher and our work demonstrate that although changes in structural preorganization and DNA-binding function do not exactly track together, increasing α-helicity in basic regions by alanine mutagenesis can generate preorganized basic regions capable of high-affinity, sequence-specific DNA binding.

We considered the possibility that nonspecific nonpolar interactions from alanines or fluorescein contribute to binding. We therefore measured the dissociation constants of all four bZIP mutants to a nonspecific 20 base-pair duplex labeled at the 5′-end with fluorescein (Figure 2). Both wt bZIP and **4A** bind nonspecific duplex with K_d values of > 1 μM, and **11A** and **18A** bind at > 10 μM (Table I); the titrations were limited by protein insolubility at high micromolar concentrations, and therefore saturation binding was not achieved. These apparent monomeric dissociation constants demonstrate at least three orders of magnitude difference in binding to specific vs nonspecific DNA. Our values for nonspecific DNA binding by GCN4 bZIP peptides correlate well with those measured by Hollenbeck and Oakley using electrophoretic gel mobility shift assay (EMSA).[20]

In previous work, we measured protein–DNA dissociation constants by EMSA.[8] At that time, we had not developed a reliable strategy for maintaining protein solubility. We now routinely employ the temperature-leap tactic (T-leap) to assist correct protein folding (see Materials and Methods); by allowing the proteins to fold at very low temperatures before warming, aggregation is minimized. This technique has been used previously on enzymes.[30,49,50] We have since found that in combination with the T-leap, we can reduce the urea concentration to 800 m*M*; this concentration of urea does not denature our short DNA duplexes. Melting curves of the ATF/CREB 21-mer duplex in the buffer used in anisotropy titrations showed reversible melting behavior, with melting temperatures of 51.4 and 57.0°C with and without 800 m*M* urea, respectively. K_ds measured by fluorescence anisotropy show the opposite trend from those obtained by EMSA. Binding affinities remain similar as Ala content increases in the anisotropy studies; in EMSA, however, wt bZIP formed a stronger complex with AP-1 than did **18A** by three orders of magnitude.[8] Moreover, dissociation constants measured by anisotropy are in the nanomolar range, whereas those

measured by EMSA were micromolar. Our previous EMSA studies suffered from protein solubility problems and use of 90 m*M* to 1.2*M* urea.

DISCUSSION

We are examining mutants of the GCN4 bZIP to seek the minimal determinants for a dimer of protein α-helices to recognize specific DNA major groove sites. To this end, we have created bZIPs of preorganized helical structure with favorable energetics for binding to DNA.[7,8] Using NMR spectroscopy on the GCN4 bZIP/AP-1 complex, Palmer and co-workers found that the GCN4 basic region populates an ensemble of highly dynamic transient structures with substantial helical character; however, restriction of the conformational space accessible to the basic region may significantly reduce the entropic cost associated with helix formation consequent to DNA binding.[51] Padmanabhan et al. found that Ala-rich 17-mers containing four lysines assumed compact, helical structures upon binding DNA, and that the energetic penalty for folding was worth 1–2 orders of magnitude in binding affinities.[52]

We anticipated that Ala-rich peptides would provide a reproducible scaffold from which we can design and meaningfully compare molecular recognition elements, and that the DNA structure would remain preserved as well. There will undoubtedly be some cooperative interactions between protein functional groups, and mutations can affect structure, perhaps in subtle fashion; however, our Ala-based proteins maintain α-helical structure characteristic of the bZIP motif, and therefore the local and global secondary and tertiary structures are retained. Mossing and Record note a similar concern in their studies on the *lac* repressor in complex with mutant operator sites.[53] In their work, DNA sequence is the variable, the reasoning being that DNA structure is typically less variable than that for protein. Our work focuses on protein sequence as the variable, but similarly, we expect helical structure to remain static in our Ala-based proteins.

In the GCN4–DNA complex, is binding specificity and affinity dictated by the highly conserved Asn^{235}, Ala^{238}, Ala^{239}, and Arg^{243}? How essential are Coulombic interactions? Jones et al. analyzed numerous high-resolution structures and found that sequence-specific protein–DNA interfaces depend strongly on polar interactions, including hydrogen bonds and bridging water molecules: Arg is the most common residue found at protein–DNA interfaces.[54] Interestingly, we note that Arg^{227} in **4A** and **11A** is a cloning artifact; the native sequence has Pro and **18A** has Ala at position 227 (Figure 2). Although crystal structures show that residue 227 does not contact DNA,[10–12] Coulombic interactions can operate at longer distances, and Arg^{227} may enhance the binding affinities of **4A** and **11A** to DNA. In view of this, the strong binding of **18A** to AP-1 and ATF/CREB is even more noteworthy and reinforces that in the bZIP system, high affinity and specificity can be maintained despite loss of numerous polar interactions, few residues are necessary for DNA-binding function, and that extensive mutation is well tolerated.

The polyanionic DNA biopolymer is hydrophilic along the phosphodiester backbone exterior, but the base-pair interior has significant nonpolar character. The GCN4 bZIP crystal structures show methyl side chains from Ala^{238} and Ala^{239} making van der Waals interactions with thymine C_5 methyl groups in the major groove, and the Ser^{242} side chain making a more distant hydrophobic contact with the C_5 methyl on T_3, as well as a nonspecific phosphodiester interaction.[10–12] **18A** substitutes Ser^{242} with Ala, which is still capable of maintaining the van der Waals contact with thymine but loses the hydrogen bond to the DNA backbone. Our results are consistent with that of Struhl and co-workers, who mutated Ala^{238} to serine and cysteine, and Ala^{239} to serine; the methylene units on these side chains can mimic the Ala methyl group. Notably, all three mutants behaved like native GCN4.[55] Struhl's group also showed that substitution of Asn^{235}, Ala^{238}, and Ala^{239} with large hydrophobic residues (Trp, Tyr, Val) did not significantly affect GCN4's DNA-binding abilities.[56] Our Ala-based proteins, while simplifying the bZIP scaffold, can also maintain van der Waals contacts with the major groove. The *lac* repressor-operator system also depends on nonpolar contacts; sequence-specific DNA recognition requires that $\sim$ 50% of its interactions be Coulombic, whereas binding to nonspecific, relatively featureless, DNA is mostly Coulombic.[53] Even in short peptides of less than 20 residues comprising mostly alanines and lysines, nonspecific binding is driven by Coulombic interactions.[52] These results, as well as our results presented here, are in direct contrast with the detailed analysis of high-resolution structures by Jones et al. described above.[54]

Despite numerous alanine replacements in the basic region, our bZIP mutants maintain low-nanomolar dissociation constants for binding to the AP-1 and ATF/CREB sites. Enthalpy and entropy, which usually move in opposite directions, can compensate in order to maintain fairly stable overall free energies of binding.[57,58] Compensation occurs frequently in biological systems, for enthalpy and entropy values can

vary widely in dependent fashion, even though ΔG values cluster in a relatively narrow range.[57] For example, we performed fluorescence anisotropy titrations on our bZIP mutants binding to AP-1 duplexes in which thymines were replaced with uracil: hence, removal of the highly conserved van der Waals contacts between thymine C_5 methyls and Ala^{238} and Ala^{239}.[59] Modest changes of less than one order of magnitude in binding affinities were measured; the multivalent nature of protein–DNA interactions may allow enthalpy/entropy compensation to occur easily.

A primary source of favorable ΔS in protein–DNA complexation is release of water from nonpolar surfaces.[57] Castro and Anderson generated single alanine mutations in bovine pancreatic trypsin inhibitor (BPTI), with little change in enzyme inhibition; they attribute this result to compensatory changes in enthalpy and entropy, in particular, desolvation effects at the binding interface.[60] In our system, water release from DNA should remain the same, but water removal from polar wt bZIP should be more costly than that from nonpolar **18A**. Along with the side-chain/base interactions between protein and DNA, the folding and stability of the GCN4 basic region also factors into binding free energies. The entropic penalty should not be as severe for the preformed Ala-rich helix as for the disordered native basic region. A potential source of unfavorable enthalpy change is from strain, often arising from energetically costly DNA distortions; crystal structures show that GCN4 binds to AP-1 with no distortion,[11] but the GCN4-ATF/CREB structure shows that the central base pairs exhibit A-form character and a 20° bend.[10] Berger et al. used microcalorimetry to show that although ΔG is the same for GCN4 binding to either AP-1 or ATF/CREB, and GCN4-DNA complexation is accompanied by favorable ΔH and unfavorable ΔS at all temperatures studied, ΔH was more favorable for the undistorted AP-1 site.[61]

Does a code exist for protein–DNA recognition? It appears unlikely that recognition patterns involving proteins will be as straightforward as base-pairing patterns for nucleic acids, where Watson–Crick rules predominate. Another complicating factor is that we cannot expect a straightforward linear additivity of binding interactions in such an adaptive system as the protein scaffold.[62] For example, the free energy of interaction of the *lac* repressor with mutant operators cannot be explained by linearly additive contributions from independent interactions with individual base pairs.[53,63] Even in the case of much smaller scaffolds, there is nonadditivity of individual interactions: for minor-groove binding synthetic polyamides comprising imidazoles and pyrroles, the energetic cost for binding a double mismatch DNA site is not twice that for a single mismatch site.[64] Additionally, changes can affect binding unpredictably: a pyrrole-to-imidazole change in the hairpin polyamide ligand increases binding affinity for the match site, while reducing specificity for the match vs. single mismatch sites.[64,65] Although no simple code exists that would allow one to predict the structure and energetics of protein–DNA interactions, thermodynamic studies suggest that less obvious structure–function relationships can exist.[53,63–65] Dill reasons that it may be inappropriate to sum individual free energies to describe a larger complex, but the success of additivity may depend on uniformity of environment—i.e., each substituent has the same neighbors.[62] Our Ala-rich bZIP mutants provide α-helical structure and a more homogeneous nearest-neighbor environment that may be be advantageous for quantitative dissection of protein–DNA recognition.

The bZIP can tolerate a high level of mutation; note that our heavily mutagenized proteins retain native DNA-binding specificity and affinity. Pu and Struhl found that seven residues in the hinge confer the correct orientation of the basic region segments for strong DNA binding—the identity of the residues was less important than the number of residues.[55] Similarly, **4A** and **11A**, which only differ in that **11A** has alanine replacements in the hinge, specifically bind the AP-1 and ATF/CREB sites, which differ by one centrally located base pair, with comparable affinities. Mutations may introduce unfavorable interactions that distort the recognition surface, and therefore protein (and DNA) can adapt to minimize loss of specific contacts; hence, GCN4 adjusts to bind strongly to both AP-1 and ATF/CREB. Perhaps the protein scaffold has evolved to accommodate mutations that occur during evolution.[66]

Our proteins were designed to simplify the minimalist bZIP structure. Amino acid substitutions can serve to decrease the conformational entropy of a protein structure by entropic stabilization.[66] Thus stabilization of the individual components in a complex can enhance both the rate of assembly and stability of the resultant complex.[67] The bZIP-DNA crystal structures show that only four amino acids per monomer make base-specific contacts with the major groove[10–13]: these amino acids can be considered "responsible" for bZIP function, i.e., sequence-specific recognition of the DNA major groove. Of the twenty proteinogenic amino acids, alanine is sufficient to maintain the minimalist bZIP backbone α-helix. Those amino acids necessary for maintenance of the α-helix can be considered responsible for bZIP structure. Therefore, Ala-rich α-helices with judiciously placed residues

conferring DNA-binding function may comprise the minimal protein determinants for high-affinity, sequence-specific recognition of the DNA major groove.

These results suggest that short, predictable peptides can make design, synthesis, and characterization of biomolecular assemblies a more tractable problem. The GCN4 basic region is not a stable peptide by itself. Kim and co-workers,[68,69] Morii and co-workers,[70] and our own group (J. A. Shin, unpublished results) synthesized GCN4 basic region derivatives dimerized through a Cys–Cys disulfide linkage in lieu of the leucine zipper, and found these basic region dimers to have low affinity for DNA and no capability for sequence-specific binding at room temperature. The leucine zipper serves a dual purpose of dimerizing element and stabilizing the adjacent disordered basic region; therefore, the basic region *and* leucine zipper—i.e., the full bZIP—are required for native GCN4 DNA-binding function. Our work demonstrates that with Ala replacements in the basic region, we may be able to dispense with the leucine zipper altogether. Therefore, extraordinarily small, simplified peptides may be designed and synthesized to bind DNA with high sequence specificity and binding affinity.

We are grateful to Jim Noll for technical expertise, Bruce Armitage for helpful discussion and critical reading of the manuscript, Andrea Jaquins-Gerstl for fluorescein-labeled DNA, Stuart Kushon for assistance with the DNA melting and curve fit, Jeff Brodsky for helpful discussion, and Mike Cascio for use of the CD. This work was supported by a grant from the National Science Foundation (CAREER MCB-9733410) to JAS and the University of Pittsburgh.

REFERENCES

1. Dawson, P. E.; Kent, S. B. H. J Am Chem Soc 1993, 115, 7263–7266.
2. Grove, A.; Mutter, M.; Rivier, J. E.; Montal, M. J Am Chem Soc 1993, 115, 5919–5924.
3. Marqusee, S.; Baldwin, R. L. Proc Natl Acad Sci USA 1987, 84, 8898–8902.
4. Sasaki, T.; Kaiser, E. J Am Chem Soc 1989, 111.
5. Wharton, R. P.; Ptashne, M. Nature 1985, 316, 601–605.
6. Ghadiri, M. R.; Soares, C.; Choi, C. J Am Chem Soc 1992, 114, 4000–4002.
7. Lajmi, A. R.; Wallace, T. R.; Shin, J. A. Protein Exp Purif 2000, 18, 394–403.
8. Lajmi, A. R.; Lovrencic, M. E.; Wallace, T. R.; Thomlinson, R. R.; Shin, J. A. J Am Chem Soc 2000, 122, 5638–5639.
9. Hill, D. E.; Hope, I. A.; Macke, J. P.; Struhl, K. Science 1986, 234, 451–457.
10. König, P.; Richmond, T. J. J Mol Biol 1993, 233, 139–154.
11. Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. Cell 1992, 71, 1223–1237.
12. Keller, W.; König, P.; Richmond, T. J. J Mol Biol 1995, 254, 657–667.
13. Glover, J. N. M.; Harrison, S. C. Nature 1995, 373, 257–261.
14. O'Neil, K. T.; DeGrado, W. F. Science 1990, 250, 646–651.
15. Luque, I.; Mayorga, O. L.; Freire, E. Biochemistry 1996, 35, 13681–13688.
16. O'Neil, K. T.; Shuman, J. D.; Ampe, C.; DeGrado, W. F. Biochemistry 1991, 30, 9030–9034.
17. Saudek, V.; Pasley, H. S.; Gibson, T.; Gausepohl, H.; Frank, R.; Pastore, A. Biochemistry 1991, 30, 1310–1317.
18. Weiss, M. A.; Ellenberger, T.; Wobbe, C. R.; Lee, J. P.; Harrison, S. C.; Struhl, K. Nature 1990, 347, 575–578.
19. Shin, J. A. Bioorg Med Chem Lett 1997, 7, 2367–2372.
20. Hollenbeck, J. J.; Oakley, M. G. Biochemistry 2000, 39, 6380–6389.
21. Johnson, P. F. Mol Cell Biol 1993, 13, 6919–6930.
22. Sera, T.; Schultz, P. G. Proc Natl Acad Sci USA 1996, 93, 2920–2925.
23. LeTilly, V.; Royer, C. A. Biochemistry 1993, 32, 7753–7758.
24. Oliphant, A. R.; Struhl, K. Methods Enzymol 1987, 155, 568–582.
25. Oliphant, A. R.; Nussbaum, A. L.; Struhl, K. Gene 1986, 44, 177–183.
26. Casimiro, D. R.; Wright, P. E.; Dyson, H. J. Structure 1997, 5, 1407–1412.
27. Casimiro, D. R.; Toy-Palmer, A.; Blake, R. C., II; Dyson, H. J. Biochemistry 1995, 34, 6640–6648.
28. Prytulla, S.; Dyson, H. J.; Wright, P. E. FEBS Lett 1996, 399, 283–289.
29. Sambrook, J.; Fritsch, E. F.; Maniatis, T. Molecular Cloning: A Laboratory Manual, 2nd ed.; Cold Spring Harbor Press: Cold Spring Harbor, NY, 1989.
30. Xie, Y.; Wetlaufer, D. B. Protein Sci 1996, 5, 517–523.
31. Maxam, A.; Gilbert, W. Methods Enzymol 1980, 65, 499–560.
32. Marky, L. A.; Breslauer, K. J. Biopolymers 1987, 26, 1601–1620.
33. Lakowicz, J. R. Principles of Fluorescence Spectroscopy; 2nd ed.; Plenum Press: New York, 1999.
34. Metallo, S. J.; Schepartz, A. Chem Biol 1994, 1, 143–151.
35. Brindle, P. K.; Montminy, M. R. Curr Opin Gen Dev 1992, 2, 199–204.
36. Agre, P.; Johnson, P. F.; McKnight, S. L. Science 1989, 246, 922–926.
37. Heyduk, T.; Lee, J. C. Proc Natl Acad Sci USA 1990, 87, 1744–1748.

38. Royer, C. A.; Ropp, T.; Scarlata, S. F. Biophys Chem 1992, 43, 197–211.
39. Wittmayer, P. K.; Raines, R. T. Biochemistry 1996, 35, 1076–1083.
40. Tanha, J.; Lee, J. S. Nucleic Acids Res 1997, 25, 1442–1449.
41. Walker, G. T.; Linn, C. P.; Nadeau, J. G. Nucleic Acids Res 1996, 24, 348–353.
42. Perez-Howard, G. M.; Weil, P. A.; Beechem, J. M. Biochemistry 1995, 34, 8005–8017.
43. Heyduk, T.; Ma, Y.; Tang, H.; Ebright, R. H. Methods Enzymol 1996, 274, 492–503.
44. Park, C.; Campbell, J. L.; Goddard, W. A., III. J Am Chem Soc 1995, 117, 6287–6291.
45. Morii, T.; Yamane, J.; Aizawa, Y.; Makino, K.; Sugiura, Y. J Am Chem Soc 1996, 118, 10011–10017.
46. Foulds, G. J.; Etzkorn, F. A. Nucleic Acids Res 1998, 26, 4304–4305.
47. Sellers, J. W.; Vincent, A. C.; Struhl, K. Mol Cell Biol 1990, 10, 5077–5086.
48. Takemoto, C.; Fisher, D. E. Gene Express 1995, 4, 311–317.
49. Danner, M.; Seckler, R. Protein Sci 1993, 2, 1869–1881.
50. Betts, S. D.; King, J. Protein Sci 1998, 7, 1516–1523.
51. Bracken, C.; Carr, P. A.; Cavanagh, J.; Palmer, A. G., III. J Mol Biol 1999, 285.
52. Padmanabhan, S.; Zhang, W.; Capp, M. W.; Anderson, C. F.; Record, J., M. T. Biochemistry 1997, 36, 5193–5206.
53. Mossing, M. C.; Record, M. T. J Mol Biol 1985, 186, 295–305.
54. Jones, S.; van Heyningen, P.; Berman, H. M.; Thornton, J. M. J Mol Biol 1999, 287, 877–896.
55. Pu, W. T.; Struhl, K. Mol Cell Biol 1991, 11, 4918–4926.
56. Kim, J.; Tzamarias, D.; Ellenberger, T.; Harrison, S. C.; Struhl, K. Proc Natl Acad Sci USA 1993, 90, 4513–4517.
57. Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A. Structure 2000, 8, 1015–1023.
58. Jen-Jacobson, L.; Engler, L. E.; Ames, J. T.; Kurpiewski, M. R.; Grigorescu, A. Supramol Chem 2000, 12, 143–160.
59. Kise, K. J., Jr.; Shin, J. A. Bioorg Med Chem 2001, 9, 2485–2491.
60. Castro, M. J. M.; Anderson, S. Biochemistry 1996, 35, 11435–11446.
61. Berger, C.; Jelesarov, I.; Bosshard, H. R. Biochemistry 1996, 35, 14984–14991.
62. Dill, K. A. J Biol Chem 1997, 272, 701–704.
63. Frank, D. E.; Saecker, R. M.; Bond, J. P.; Capp, M. W.; Tsodikov, O. V.; Melcher, S. E.; Levandoski, M. M.; Record, M. T. J Mol Biol 1997, 267, 1186–1206.
64. Pilch, D. S.; Poklar, N.; Baird, E. E.; Dervan, P. B.; Breslauer, K. J. Biochemistry 1999, 38, 2143–2151.
65. Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. Proc Natl Acad Sci USA 1996, 93, 8306–8311.
66. Matthews, B. W. Biochemistry 1989, 28, 6–9.
67. Dürr, E.; Jelesarov, I.; Bosshard, H. R. Biochemistry 1999, 38, 870–880.
68. Talanian, R. V.; McKnight, C. J.; Kim, P. S. Science 1990, 249, 769–771.
69. Talanian, R. V.; McKnight, C. J.; Rutkowski, R.; Kim, P. S. Biochemistry 1992, 31, 6871–6875.
70. Okahata, Y.; Niikura, K.; Sugiura, Y.; Sawada, M.; Morii, T. Biochemistry 1998, 37, 5666–5672.