



# The Contribution of the Methyl Groups on Thymine Bases to Binding Specificity and Affinity by Alanine-Rich Mutants of the bZIP Motif

Kenneth J. Kise, Jr and Jumi A. Shin\*

*Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260, USA*

Dedicated to Professor Peter B. Dervan, a wonderful adviser and scientist—it was a pleasure and privilege to work with you, Peter!

Received 15 April 2001; accepted 11 June 2001

**Abstract**—We have used fluorescence anisotropy to measure in situ the thermodynamics of binding of alanine-rich mutants of the GCN4 basic region/leucine zipper (bZIP) to short DNA duplexes, in which thymines were replaced with uracils, in order to quantify the contributions of the C5 methyl group on thymines with alanine methyl side chains. We simplified the  $\alpha$ -helical GCN4 bZIP by alanine substitution: 4A, 11A, and 18A contain four, 11, and 18 alanine mutations in their DNA-binding basic regions, respectively. Titration of fluorescein-labeled duplexes with increasing amounts of protein yielded dissociation constants in the low-to-mid nanomolar range for all bZIP mutants in complex with the AP-1 target site (5'-TGACTCA-3'); binding to the nonspecific control duplex was >1000-fold weaker. Small changes of <1 kcal/mol in binding free energies were observed for wild-type bZIP and 4A mutant to uracil-containing AP-1, whereas 11A and 18A bound almost equally well to native AP-1 and uracil-containing AP-1. These modest changes in binding affinities may reflect the multivalent nature of protein–DNA interactions, as our highly mutated proteins still exhibit native-like behavior. These protein mutations may compensate for changes in enthalpic and entropic contributions toward DNA-binding in order to maintain binding free energies similar to that of the native protein–DNA complex.  
© 2001 Elsevier Science Ltd. All rights reserved.

## Introduction

In order to begin to probe how Nature uses the protein  $\alpha$ -helix to bind DNA, we generated minimalist, alanine-rich proteins from a core molecular recognition scaffold based on the basic region/leucine zipper (bZIP) protein GCN4,<sup>1,2</sup> a dimeric transcriptional regulatory protein that governs histidine biosynthesis in yeast under conditions of amino-acid starvation.<sup>3</sup> The full-length GCN4 monomer is 281 amino acids, and the bZIP structure comprises a dimer of  $\sim$ 60 residue monomers. Crystal structures of the bZIP domain of GCN4 bound to two different DNA sites,<sup>4–6</sup> as well as the Jun–Fos heterodimeric bZIP–DNA crystal,<sup>7</sup> show that a continuous  $\alpha$ -helix of  $\sim$ 60 amino acids provides both the basic region for sequence-specific DNA-binding function and the leucine zipper coiled coil for protein dimerization (Fig. 1). Thus, the simplicity and tractability of the bZIP

make it the ideal system for molecular design and quantitative analysis of the minimal protein determinants for binding specificity and affinity.

Of the naturally occurring amino acids, alanine possesses the highest propensity for forming and stabilizing  $\alpha$ -helical protein structures.<sup>8,9</sup> Interestingly, the bZIP basic region is disordered until binding to DNA.<sup>10–14</sup> Nature may employ this folding transition in the basic region to 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 preorganized Ala-based protein scaffolds. We therefore constructed four bZIP mutants containing increasing numbers of alanines to explore the relationship between a pre-organized, helical structure and sequence-specific, high-affinity DNA-binding function (Fig. 2). These bacterially expressed Ala-based mutants are short ( $\sim$ 100 amino acids) and hydrophobic. The leucine zipper hails from C/EBP (residues 312–338) and the basic region comprises Ala-based derivatives of GCN4 (residues

\*Corresponding author. Tel.: +1-412-624-4927; fax: +1-412-624-4255; e-mail: jumi@pitt.edu

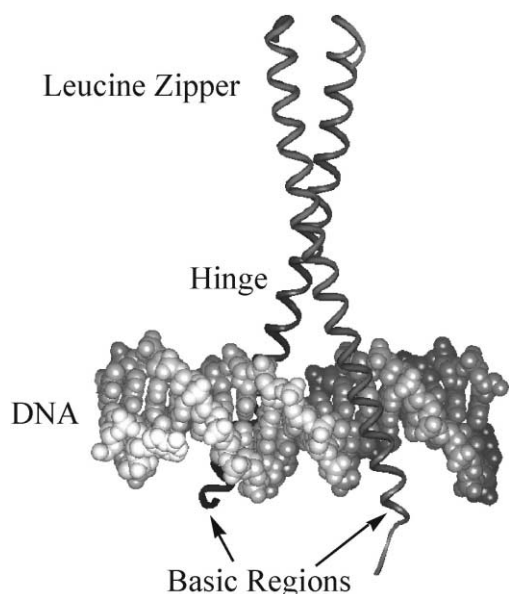
226–252).<sup>1</sup> wt bZIP (wild-type) is the ‘native’ variant comprising the GCN4 basic region fused to the C/EBP leucine zipper at the same junction used by Agre et al.; their fusion was demonstrated to mimic the DNA-binding function of native GCN4 bZIP.<sup>15</sup>

We have previously shown that the bZIP molecular recognition scaffold is highly tolerant of multiple alanine substitutions. Our Ala-rich bZIP mutants bind as specifically<sup>1</sup> and as strongly<sup>16</sup> to the pseudopalindromic AP-1 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,<sup>17</sup> as does native GCN4 bZIP. Proteins 4A and 11A contain four and 11 Ala substitutions, respectively; both specific interactions with DNA bases and nonspecific Coulombic interactions with the DNA phosphodiester backbone are maintained.<sup>4–6</sup> Additionally, 11A is mutated in the hinge region, which is important for orienting the basic region monomers properly on DNA. The GCN4 bZIP–DNA crystal structures show that only four highly conserved amino acids in each basic region monomer make direct contacts to bases in the DNA major groove: Asn<sup>235</sup>, Ala<sup>238</sup>, Ala<sup>239</sup>, and Arg<sup>243</sup>.<sup>4–6</sup> Only these four amino acids are conserved in our 18A mutant, plus Lys<sup>246</sup>, which we had hoped would improve solubility;<sup>6</sup> note that 24 of the 27 residues in the 18A basic region are alanines.

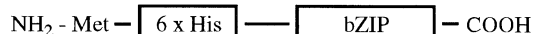
As alanine content increases in our bZIP proteins,  $\alpha$ -helicity, which we equate with structural stability, also increases.<sup>1</sup> DNase I footprinting analysis demonstrates that wt bZIP, 4A, 11A, and 18A specifically bind the AP-1 and ATF/CREB target sites.<sup>1,16</sup> Fluorescence anisotropy spectroscopy shows that all four of our bZIP

derivatives give low nanomolar dissociation constants for complexation with the AP-1 and ATF/CREB sites, and that binding to a nonspecific control duplex is >1000-fold weaker.<sup>16</sup> These  $K_d$  values compare very closely with those measured by a number of labs using a variety of techniques.<sup>4,12,14,18–22</sup> Therefore, despite the loss of numerous basic residues that make Coulombic interactions with the anionic DNA phosphodiester backbone, as seen in the crystal structures,<sup>4–6</sup> our non-polar Ala-based bZIP mutants display very similar  $K_d$  values to specific and nonspecific DNA as does native GCN4 bZIP.<sup>4,12,14,21,22</sup>

These results prompted us to examine the importance of van der Waals interactions between our hydrophobic Ala-rich proteins and DNA duplexes in which thymines were replaced with uracils. Because uracil places a hydrogen rather than methyl in the C5 position, we can compare binding affinities of our bZIP proteins to uracil-containing DNA duplexes with binding affinities to native thymine-containing duplexes. The structures of the GCN4 bZIP in complex with the AP-1<sup>5</sup> or ATF/CREB<sup>4,6</sup> sites show that the methyl side chains of Ala<sup>238</sup>



**Figure 1.** GCN4 bZIP in complex with the AP-1 DNA site, 5'-TGACTCA.<sup>5</sup> DNA is the horizontal double helix at the bottom of the figure, and the bZIP is the vertical  $\alpha$ -helical dimer. 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 major groove.



#### bZIP domain

GCN4 basic region C/EBP leucine zipper linker  
DPAALKRARNTAAARRSRARKLQRMKQ-LEQKVELTSDNDRLRKVEQLSRELDLTL-GGCGGYYY

#### GCN4 basic regions

	226	252
wt	DPAALKRARNTAAARRSRARKLQRMKQ	
4A	<u>ARAAAA</u> RARNTAAARRSRARKLQRMKQ	
11A	<u>ARAAAA</u> RARNTAAARRSRAAKAAAAAA	
18A	<u>AAAAAAAAANAAAAAA</u> RAAKAAAAAA	

#### AP-1 DNA site

3210  
5'-TGACTCA-3'  
3'-ACTGAGT-5'

#### Oligonucleotides for Fluorescence Anisotropy

AP-1 site, 20-mer

5'-(FAM)UCCGGAUGACUCAUUUUUUG-3'

nonspecific duplex, 20-mer

5'-(FAM)AACCGGUUACGUCAGACUGU-3'

**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 312–338) 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; these proteins are the same as wt, 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.<sup>4–6</sup> (Bottom) Sequence of the AP-1 DNA site. Numbering begins at the central CG base pair. Sequences of the oligonucleotide duplexes used in fluorescence anisotropy titrations. ‘FAM’ is fluorescein phosphoramidite, and the AP-1 site is underlined.

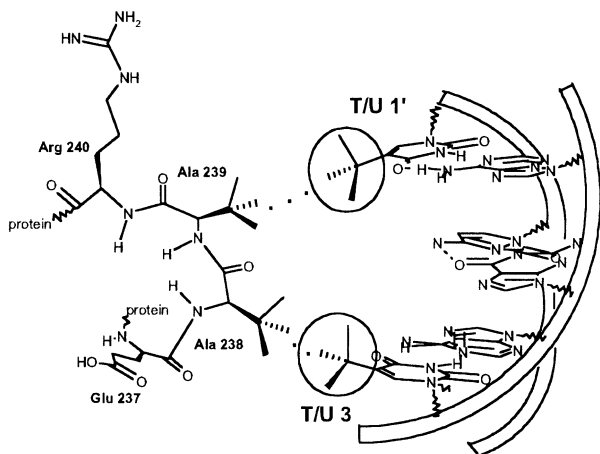
and Ala<sup>239</sup> are within van der Waals contact distance of the C5 methyl groups of thymines T3 and T1' (numbering shown in Fig. 2), respectively (Fig. 3). The ATF/CREB structure shows that the Ala<sup>238</sup> side chain is 4.0 Å from the C5 methyl of T3, and the Ala<sup>239</sup> side chain is 3.5 Å from the C5 methyl of T1'.<sup>4,6</sup> The Ser<sup>242</sup> side chain also makes a more distant van der Waals contact with the C5 methyl on T3, as well as a nonspecific phosphodiester interaction. Our mutants can maintain these nonpolar interactions, for we retain Ala<sup>238</sup> and Ala<sup>239</sup> in all our proteins; Ser<sup>242</sup> is replaced with Ala in the 18A mutant, but the Ala methyl should similarly maintain the native van der Waals interaction. Therefore, our Ala-rich mutants, while simplifying the bZIP scaffold, can maintain native van der Waals contacts; possibly, these mutants may also replace native Coulombic and hydrogen-bonded contacts with new van der Waals interactions between the Ala methyl side chain and nonpolar groups on the DNA.

We hypothesized that removal of the van der Waals contacts afforded by the thymine methyl groups to the alanine methyl side chains would affect protein–DNA binding affinities. We therefore used fluorescence anisotropy spectroscopy to characterize quantitatively the binding affinities of wt bZIP, 4A, 11A, and 18A to fluorescein-labeled 20-mer duplexes with thymine → uracil replacements (Fig. 2); these duplexes contain either the AP-1 site (5'-TGACTCA-3') or a nonspecific control sequence (contains equal amounts of the four bases in a nonhairpinning sequence). Fluorescence anisotropy, which measures the tumbling motion of fluorophore-labeled molecules, is an excellent quantitative, reproducible, solution-based methodology providing true equilibrium binding profiles (we maintained > 100-fold excess protein over DNA throughout the titrations).  $K_d$  values to the uracil-containing AP-1 duplex (u-AP-1) were strong for all proteins, in the low-to-mid nanomolar range. Binding to the uracil-containing nonspecific control duplex (u-NS) was measured for wt

bZIP and 18A: both proteins bound u-NS DNA > 1000-fold more weakly. We can contrast these data to fluorescence anisotropy measurements on thymine-containing native DNA duplexes: in these experiments, we also measure low nanomolar dissociation constants and > 1000-fold weaker binding to the nonspecific duplex.<sup>16</sup> In particular, wt bZIP binds 5-fold more weakly to u-AP-1 duplexes than to native thymine-containing AP-1 (t-AP-1), whereas 4A, 11A, and 18A bind somewhat better or equally well to u-AP-1. Changes in binding affinity are not dramatic, and this study may underscore the multivalent nature of protein–DNA interactions, as our protein mutations that severely decrease ability for Coulombic interactions, yet increase ability for van der Waals contacts, still yield functioning proteins exhibiting native-like behavior. Additionally, proteins can compensate for mutations that can affect enthalpic and/or entropic contributions toward DNA-binding in order to maintain binding free energies similar to that of the native protein–DNA complex,<sup>23</sup> and therefore, compensatory changes in our proteins may alleviate the losses of van der Waals interactions with thymine C5 methyl groups.

## Results and Discussion

Table 1 shows the dissociation constants for all four bZIP mutants binding to the u-AP-1 target site. The binding affinities of all four proteins to u-AP-1 are comparable, in the low nanomolar range. In comparison, binding to the u-NS control sequence was > 1000-fold weaker; in titrations with wt bZIP and 18A, saturation binding was never achieved, even at low micromolar concentrations of protein. Figure 4 shows the fitted curves for fluorescence anisotropy titrations of each of the four proteins bound to u-AP-1; the sigmoidal binding curves are indicative of a two-state cooperative model. In previous studies, we found that our Ala-rich bZIP mutants still retain helical structure and sequence-specific DNA-binding function.<sup>2</sup> More recently, we measured dissociation constants of our Ala mutants with the AP-1 and ATF/CREB sites and found that these mutants bind DNA with affinities comparable to the native GCN4 bZIP.<sup>16</sup> These results inspired us to examine the importance of van der Waals interactions in these protein–DNA complexes. Substitution of thymines with uracils removes specific nonpolar interactions between the thymine C5 methyl and protein



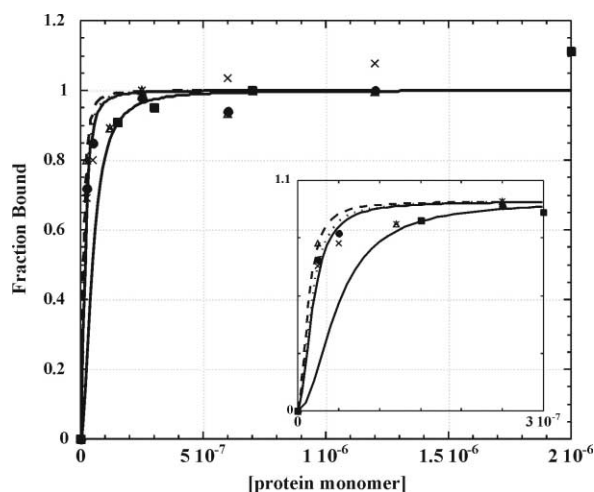
**Figure 3.** Schematic focusing on the van der Waals interactions between Ala<sup>238</sup> and Ala<sup>239</sup> of the GCN4 bZIP and the C5 methyl groups on T3 and T1' of AP-1 DNA; these methyls are circled to indicate that they are missing in the uracil-substituted DNA duplexes. van der Waals contacts are indicated by dashed lines. Adapted from ref 5.

**Table 1.** Dissociation constants for GCN4 bZIP derivatives bound to u-AP-1, t-AP-1, and u-NS sites

bZIP	$K_d$ ( $10^{-9}$ M)		
	u-AP-1	t-AP-1 <sup>a</sup>	u-NS <sup>b</sup>
wt	48.9 ± 15.7	9.12 ± 1.23	> 1 μM
4A	18.5 ± 2.14	77.8 ± 5.66	Not determined
11A	12.8 ± 2.08	4.79 ± 0.654	Not determined
18A	16.5 ± 1.34	15.2 ± 1.33	> 1 μM

<sup>a</sup>Data from ref 16.

<sup>b</sup>Saturation protein binding was not achieved in any of the nonspecific duplex DNA titrations.



**Figure 4.** Fluorescence anisotropy titrations. Binding of all four mutant bZIP proteins to u-AP-1 duplex. wt bZIP (■, solid line), 4A (× solid line), 11A (△, dashed line), and 18A (●, dashed line). Inset shows enlargement of data collected at low protein concentrations (0–300 nM range).

groups, and importantly, uracil substitutions do not affect DNA structure. Therefore, we can make meaningful, quantitative comparisons of the binding energetics of our mutants bound to native DNA or uracil-substituted DNA.

We can compare the dissociation constants for the bZIP mutants binding to u-AP-1 to those values observed for binding to native thymine-containing AP-1 (t-AP-1).<sup>16</sup> Both the u-AP-1 and t-AP-1 duplexes are of the same sequence, except for T→U substitutions; the duplexes are both 20-mers labeled with fluorescein. wt bZIP binding to u-AP-1 is 5-fold weaker than binding to t-AP-1. In the absence of the thymine C5 methyl group, the methyl side chains of Ala<sup>238</sup> and Ala<sup>239</sup> are unlikely to make van der Waals interactions with the DNA major groove. This binding is weaker by ~0.7 kcal/mol, not a substantial difference considering that there are two hydrophobic contacts, and given that Ala<sup>238</sup> and Ala<sup>239</sup> are highly conserved among bZIP proteins,<sup>24</sup> we may have expected a larger contribution from these van der Waals contacts. Moreover, the GCN4 bZIP/DNA structures show that the Ala<sup>238</sup> and Ala<sup>239</sup> side chains are within van der Waals contact distance of the C5 methyl groups of thymines T3 and T1'.<sup>4–6</sup> The GCN4 bZIP-ATF/CREB structure shows that Ala<sup>238</sup> is 4.0 Å from the T3 methyl, and that Ala<sup>239</sup> is 3.5 Å from the T1' methyl.<sup>4</sup> In contrast to wt bZIP which binds native t-AP-1 more strongly, 4A binds four-fold stronger to u-AP-1 than to t-AP-1, corresponding to ~0.6 kcal/mol, again a very modest change. Proteins 11A and 18A bind to u-AP-1 and t-AP-1 with essentially the same, low-nanomolar dissociation constants.

Other research groups have observed similarly modest changes in binding of specific DNA sites by GCN4 bZIP derivatives, wherein substitutions were made at positions 238 and 239 on the bZIP or T→U substitutions on the DNA.<sup>24–26</sup> Pu and Struhl mutated the highly conserved Ala<sup>238</sup> and Ala<sup>239</sup>, as well as the

invariant Asn<sup>235</sup>, of GCN4 and Jun bZIP derivatives and examined the effects on binding to the AP-1 DNA site.<sup>26</sup> Mutations were tested by phenotype analysis of transcription and uracil interference assay; we note that these assays are not quantitative. Although Asn<sup>235</sup> was found to be important for recognition of the AP-1 site, fairly conservative mutations at Ala<sup>238</sup> and Ala<sup>239</sup> did not affect protein function: Ala<sup>238</sup>Ser, Ala<sup>238</sup>Cys, and Ala<sup>239</sup>Ser mutants behaved like the native bZIP. The authors suggest that Ala<sup>238</sup> and Ala<sup>239</sup> may be highly conserved evolutionarily for structural compatibility with neighboring residues, not for intrinsic functional importance, for they are not critical to recognition of AP-1 in these experiments.<sup>26</sup> In subsequent work, even substitution of Asn<sup>235</sup> with Trp showed only a modest decrease in binding to specific DNA sites, and Kim et al. suggest that flexibility of the protein–DNA interface accommodates this mutation by allowing proper orientation of the indole ring.<sup>25</sup>

The relatively disordered structure of the basic region prior to binding DNA may also contribute to the flexibility of the protein–DNA interface.<sup>25</sup> Kim et al. also examined GCN4 mutants now with amino-acid replacements possessing larger side chains: Ala<sup>238</sup>Tyr and Ala<sup>239</sup>Val.<sup>25</sup> These larger functionalities were expected to promote steric clash between the protein and DNA, but phenotype analysis and uracil interference assay showed no effect on DNA binding by the Tyr derivative, and only somewhat decreased binding by the Val mutant. In comparison, Johnson used DNase I footprinting on GCN4-C/EBP bZIP hybrids and found that Ala<sup>239</sup> was important for discrimination of GCN4 and C/EBP binding behavior: GCN4 recognizes the 5'-TGAC half site, whereas C/EBP, which possesses a valine in the 239 position, binds 5'-TTGC.<sup>24</sup> Kim et al. did not, however, examine binding of their GCN4 Ala<sup>239</sup>Val mutant to the 5'-TTGC site.<sup>25</sup>

Although subtle, van der Waals interactions to thymine C5 methyl groups can contribute significantly to binding specificity. By making T→U substitutions in operator sites, Albright and Matthews showed that for Cro repressor and λ-repressor, binding specificity is subtly achieved largely through differences in van der Waals contacts to specific thymine C5 methyl groups; they find that binding affinity, however, depends more on hydrogen bonds rather than nonpolar interactions.<sup>27</sup> For example, in the O<sub>R</sub>1 site, 5'TATCACCGC (increasing numbering from left to right), the authors find a strong van der Waals interaction worth 0.8 kcal/mol between the C5 methyl on T3 and Asn<sup>231</sup> of Cro. The corresponding residue in λ-repressor, Gly<sup>48</sup>, is too far from the T3 methyl to make a significant interaction (5.7 Å). Chen et al. demonstrated that C5 methyl groups on four thymines were critical for interaction of the diphtheria toxin repressor (DtxR) with the *tox* operator.<sup>28</sup> Ser<sup>37</sup> and Pro<sup>39</sup> of DtxR were shown to interact with four thymine methyls in the protein–DNA crystal structure; gel mobility shift assays with T→U substitutions in the operator site demonstrated that the Ser<sup>37</sup>/Pro<sup>39</sup> contacts to the four methyl groups contributed a total of 3.4 kcal/mol in binding energy, and these van der Waals

interactions were important for both binding specificity and affinity. Thus, in the DtxR–*tox* operator complex, each van der Waals interaction is worth  $\sim 0.8$  kcal/mol binding energy. Using DNA sites with T→U substitutions, Duggan et al. observed an energetic penalty of 1–2 kcal/mol for removal of a single thymine methyl group in their studies on Tus protein, which is responsible for termination of DNA replication *Escherichia coli*, binding to the *TerB* DNA site.<sup>29</sup> These values agree well with studies on Cro and  $\lambda$ -repressors, for which individual van der Waals contacts with thymine methyl groups were found to be worth 0.3–2 kcal/mol in binding free energy.<sup>27</sup> In a detailed structural analysis of numerous protein–DNA and protein–RNA complexes, Jones et al. found that van der Waals interactions were significantly more common than hydrogen-bond contacts in specifically bound complexes.<sup>30</sup> Similarly, Mossing and Record showed that in specific binding between *lac* repressor and its operator site, approximately half of the interactions were van der Waals, the other half, Coulombic; in nonspecific *lac* repressor–DNA binding, most of the interactions were Coulombic in nature.<sup>31</sup>

van der Waals interactions between protein and thymine methyl groups can be significant<sup>27–31</sup> or not detectable,<sup>25,26</sup> and the contributions these interactions make can be very subtle. In our work, modest or no change in binding free energies is observed when native van der Waals interactions between GCN4 Ala<sup>238</sup> and Ala<sup>239</sup> and thymine methyl groups are removed. For our Ala-rich proteins binding to DNA, favorable contributions to binding free energies would arise from the increased helical stability and preorganization of the basic region and less costly desolvation of our nonpolar proteins. Unfavorable contributions would stem from decreased Coulombic and hydrogen-bonded contacts between protein and DNA. These changes can compensate to give binding energetics in mutant protein–DNA complexes similar to that for the native complex.

Biological systems are complex: specific contacts are not isolated or independent but typically cooperative, and additivity of interactions is not linear or straightforward. 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: that is each substituent has the same neighbors as do other substituents.<sup>32</sup> Our bZIP mutants provide an  $\alpha$ -helical structure, and the increasing Ala content in our proteins provides a more simplified, homogeneous nearest-neighbor environment. Yet free-energy prediction is complicated by the ability of enthalpy and entropy, which usually move in opposite directions, to compensate for changes in order to maintain fairly stable overall free energies of binding.<sup>23,33</sup> Compensation occurs frequently in biological systems, for enthalpy and entropy values can vary widely in dependent fashion, even though  $\Delta G$  values cluster in a relatively narrow range.<sup>23</sup> 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 DNA with no distortion,<sup>5</sup> but the GCN4-ATF/CREB structure shows that the central base pairs exhibit A-form character and a 20° bend.<sup>4</sup> Berger et al. used microcalorimetry to show that although  $\Delta G$  is the same for GCN4 binding to either AP-1 or ATF/CREB sites, and GCN4–DNA complexation was accompanied by a favorable  $\Delta H$  and an unfavorable  $\Delta S$  at all temperatures studied,  $\Delta H$  was more favorable for the undistorted AP-1 site.<sup>34</sup>

A primary source of favorable  $\Delta S$  in protein–DNA complexation is release of water from nonpolar surfaces.<sup>23</sup> Castro and Anderson generated single alanine mutations in bovine pancreatic trypsin inhibitor (BPTI) and measured these mutants' enzyme-inhibition abilities. Using calorimetry, the authors showed that alanine mutations had little effect on enzyme inhibition, and they attribute this result to compensatory changes in enthalpy and entropy, in particular, desolvation effects at the binding interface.<sup>35</sup> In our systems, water release from AP-1 DNA should remain the same, but water removal from polar wt bZIP should be markedly 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. There is an entropy penalty when the basic region folds from a loosely ordered, nascent helix to an  $\alpha$ -helix when the protein binds DNA. The entropic penalty should not be as severe for the Ala-rich mutants, because the helix is preformed. Although 18A may possess a more favorable  $\Delta S$  of binding,  $\Delta H$  may be less favorable due to loss of Coulombic interactions and hydrogen bonds between polar side chains and DNA; even water-mediated interactions, which can be as important as direct interactions with DNA,<sup>36</sup> would likely be fewer in our Ala-rich mutants.

## Conclusion

Although our mutants have lost the enthalpic benefits of Coulombic and hydrogen-bonded interactions, favorable contributions from the hydrophobic effect and immobilization of protein backbone and side chains with Ala substitutions may counter these enthalpic losses and allow the mutants to maintain native binding specificity and affinity. Enthalpy/entropy compensation confines free energies of binding within a restricted range; therefore, DNA-binding proteins can be fairly tolerant of mutations and still maintain native binding function. At the same time, however, this compensation can complicate our efforts to create proteins with widely variable binding affinities for DNA.

## Experimental

### Materials

[ $\alpha^{32}\text{P}$ ]-dATP was supplied by Amersham (Piscataway, NJ), and radioactivity was monitored on a Beckman LS

6500 scintillation counter. Water was purified through a Milli Q filtration system (Millipore, Bedford, MA). Fluorescein-labeled AP-1 (20-mer) and fluorescein-labeled nonspecific (20-mer) oligonucleotides and their complementary oligonucleotides were synthesized at the DNA Synthesis Facility, University of Pittsburgh, on a PE Biosystems Expedite 8909.

### Protein construction

Protocols for DNA oligonucleotide synthesis, gene construction and cloning, protein overexpression and purification have been described in detail.<sup>1</sup> A brief summary of these procedures follows: genes for expression of bZIP proteins were constructed by mutually primed synthesis,<sup>37,38</sup> followed by polymerase chain reaction with terminal primers for gene amplification<sup>39–41</sup> and purification by nondenaturing polyacrylamide gel electrophoresis.<sup>42</sup> Duplex DNA was then cloned into protein expression vectors pRSET B and pTrcHis B (Invitrogen, Carlsbad, CA); both vectors express proteins with a six-histidine tag for purification purposes. Recombinant plasmids were transformed into *E. coli* strain BL21(DE3) (Stratagene, La Jolla, CA) by electroporation. Cloned inserts were sequenced by dideoxy DNA sequencing (T7 Sequenase kit, USB, Cleveland, OH). Bacterial expression of bZIP proteins was performed in LB medium containing 50 µg/mL ampicillin.<sup>42</sup> Proteins were purified first on TALON cobalt metal-ion affinity resin (Clontech, Palo Alto, CA), followed by further purification by size-exclusion chromatography (SEC) or by HPLC on a C4 column (Vydac, Hesperia, CA). Protein purification was monitored by SDS-PAGE and Western immunoblot assay. Purified protein stocks were stored in 4 M urea or guanidine with 1 mM 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.<sup>43</sup> Protein solutions with diluted urea were prepared by addition of an appropriate amount of the purified stocks to that experiment's relevant buffer containing less than 4 M urea at 4 °C. The resulting solution was incubated at 4 °C for > 2 h, followed by rapid heating to 37 °C for 1 h. This renatured protein was then immediately used. Temperature-leap was performed for each fluorescence anisotropy data point.

### DNA hybridization

Fluorescein-labeled DNA solutions were stored at –20 °C and wrapped in foil to minimize light exposure. 200 pmol fluorescein-labeled oligonucleotide and 240 pmol (1.2 equiv) of the complementary oligonucleotide were added to annealing buffer (10 mM Tris, 1 mM EDTA, 50 mM NaCl, pH 7.4). The solution was heated at 80 °C for 10 min, and cooled slowly to room temperature (~2 h). The duplex DNA stock solution was stored at –20 °C.

### Fluorescence anisotropy

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, Atascadero, CA). 0.2–4 µL stock protein 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 pM specific DNA duplex or 1 nM nonspecific DNA duplex in buffer (4.3 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.4 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4, 150 mM NaCl, 2.7 mM KCl, 1 mM EDTA, 800 mM urea, 20% glycerol, 0.4 mg/mL acetylated BSA, 1 mM DTT, and 100 µM in base pairs calf thymus DNA).

At these concentrations of FAM-labeled duplex, protein concentration was in > 50-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. Fluorescein-containing solutions were handled under dark or low light conditions to minimize photodegradation of the fluorescein label.

Titrations were performed at 22 ± 2 °C. The volume change was minimized during the titration and was kept to < 10% of the total buffer volume within the cell. The *G* factor (the ratio of sensitivities of the monochromator for horizontally and vertically polarized light) was calculated from the above measurements using the equation below.<sup>44</sup>

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

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.<sup>44</sup>

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

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<sup>20</sup> using Kaleidagraph 3.0.8 (Synergy Software, Reading, PA).

### Determination of *K<sub>d</sub>* 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. Eq (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,<sup>20</sup>

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

where  $K_d$  corresponds to the apparent monomeric dissociation constant, and  $M$  is the concentration of monomeric protein. The apparent dimerization constant for the C/EBP bZIP is  $\sim 20 \mu\text{M}$ ,<sup>10</sup> and therefore, the concentration of protein dimer was insignificant when compared to the total protein concentration. In our titrations with specific DNA target sites,  $[M]$  ranged from low-to-high nanomolar. For the nonspecific duplex titrations, monomeric protein concentrations went as high as  $1.2 \mu\text{M}$ . Even at these very low micromolar concentrations of protein, saturation binding of nonspecific 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 1.

Curve fits of the fraction of DNA bound versus protein monomer concentration yielded the apparent monomeric dissociation constants of the protein–DNA complexes. Only data sets fit to eq (3) with  $R$  values  $> 0.980$  and  $\chi^2$  values  $< 0.020$  are reported.

The AP-1 duplex contains only one specific binding site, whereas the 20 base-pair nonspecific duplex comprises 8–9 overlapping nonspecific binding sites, for we assume that the binding site size for the GCN4 bZIP dimer is 11–12 base pairs.<sup>4–6</sup> In order to derive microscopic  $K_d$  values for nonspecific DNA, our measured macroscopic apparent dissociation constants should be corrected by a statistical factor that accounts for the increased number of potential binding sites on the nonspecific DNA duplex.<sup>45,46</sup> We report apparent monomeric  $K_d$ 's; therefore, the square of the microscopic monomeric dissociation constant is related to the square of the observed monomeric  $K_d$  by this statistical factor.<sup>14</sup> Thus, our measured apparent monomeric dissociation constants overestimate the affinity of our bZIP proteins for nonspecific DNA by  $8^{1/2}$ – $9^{1/2}$ , or  $\sim 3$ -fold.

### Acknowledgements

We are grateful to Jim Noll for technical expertise and Marc Greenberg for helpful discussion. This work was supported by a grant from the National Science Foundation (CAREER MCB-9733410) to J. A. S. and the University of Pittsburgh.

### References and Notes

- Lajmi, A. R.; Wallace, T. R.; Shin, J. A. *Prot. Exp. Purif.* **2000**, *18*, 394.
- Lajmi, A. R.; Lovrencic, M. E.; Wallace, T. R.; Thomlinson, R. R.; Shin, J. A. *J. Am. Chem. Soc.* **2000**, *122*, 5638.
- Hill, D. E.; Hope, I. A.; Macke, J. P.; Struhl, K. *Science* **1986**, *234*, 451.
- König, P.; Richmond, T. J. *J. Mol. Biol.* **1993**, *233*, 139.
- Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223.
- Keller, W.; König, P.; Richmond, T. J. *J. Mol. Biol.* **1995**, *254*, 657.
- Glover, J. N. M.; Harrison, S. C. *Nature* **1995**, *373*, 257.
- O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250*, 646.
- Luque, I.; Mayorga, O. L.; Freire, E. *Biochemistry* **1996**, *35*, 13681.
- O'Neil, K. T.; Shuman, J. D.; Ampe, C.; DeGrado, W. F. *Biochemistry* **1991**, *30*, 9030.
- Saudek, V.; Pasley, H. S.; Gibson, T.; Gausepohl, H.; Frank, R.; Pastore, A. *Biochemistry* **1991**, *30*, 1310.
- Weiss, M. A.; Ellenberger, T.; Wobbe, C. R.; Lee, J. P.; Harrison, S. C.; Struhl, K. *Nature* **1990**, *347*, 575.
- Shin, J. A. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2367.
- Hollenbeck, J. J.; Oakley, M. G. *Biochemistry* **2000**, *39*, 6380.
- Agre, P.; Johnson, P. F.; McKnight, S. L. *Science* **1989**, *246*, 922.
- Bird, G. H.; Lajmi, A. R.; Shin, J. A. Manuscript submitted.
- Brindle, P. K.; Montminy, M. R. *Curr. Op. Gen. Dev.* **1992**, *2*, 199.
- Park, C.; Campbell, J. L.; Goddard, W. A., III. *J. Am. Chem. Soc.* **1995**, *117*, 6287.
- Morii, T.; Yamane, J.; Aizawa, Y.; Makino, K.; Sugiura, Y. *J. Am. Chem. Soc.* **1996**, *118*, 10011.
- Metallo, S. J.; Schepartz, A. *Chem. Biol.* **1994**, *1*, 143.
- Foulds, G. J.; Etzkorn, F. A. *Nucl. Acid. Res.* **1998**, *26*, 4304.
- Sellers, J. W.; Vincent, A. C.; Struhl, K. *Mol. Cell. Biol.* **1990**, *10*, 5077.
- Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A. *Structure* **2000**, *8*, 1015.
- Johnson, P. F. *Mol. Cell. Biol.* **1993**, *13*, 6919.
- Kim, J.; Tzamarias, D.; Ellenberger, T.; Harrison, S. C.; Struhl, K. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4513.
- Pu, W. T.; Struhl, K. *Mol. Cell. Biol.* **1991**, *11*, 4918.
- Albright, R. A.; Matthews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3431.
- Chen, C. S.; White, A.; Love, J.; Murphy, J. R.; Ringe, D. *Biochemistry* **2000**, *39*, 10397.
- Duggan, L. J.; Hill, T. M.; Wu, S.; Garrison, K.; Zhang, X.; Gottlieb, P. A. *J. Biol. Chem.* **1995**, *270*, 28049.
- Jones, S.; Daley, D. T. A.; Luscombe, N. M.; Berman, H. M.; Thornton, J. M. *Nucl. Acid. Res.* **2001**, *29*, 943.
- Mossing, M. C.; Record, M. T. *J. Mol. Biol.* **1985**, *186*, 295.
- Dill, K. A. *J. Biol. Chem.* **1997**, *272*, 701.
- Jen-Jacobson, L.; Engler, L. E.; Ames, J. T.; Kurpiewski, M. R.; Grigorescu, A. *Supramol. Chem.* **2000**, *12*, 143.
- Berger, C.; Jelesarov, I.; Bosshard, H. R. *Biochemistry* **1996**, *35*, 14984.
- Castro, M. J. M.; Anderson, S. *Biochemistry* **1996**, *35*, 11435.
- Janin, J. *Structure* **1999**, *7*, R277.
- Oliphant, A. R.; Struhl, K. *Meth. Enzymol.* **1987**, *155*, 568.
- Oliphant, A. R.; Nussbaum, A. L.; Struhl, K. *Gene* **1986**, *44*, 177.
- Casimiro, D. R.; Wright, P. E.; Dyson, H. J. *Structure* **1997**, *5*, 1407.
- Casimiro, D. R.; Toy-Palmer, A.; Blake, R. C., II; Dyson, H. J. *Biochemistry* **1995**, *34*, 6640.
- Prytulla, S.; Dyson, H. J.; Wright, P. E. *FEBS Lett.* **1996**, *399*, 283.
- Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 2nd ed.; Cold Spring Harbor Press: New York, 1989.
- Xie, Y.; Wetlaufer, D. B. *Prot. Sci.* **1996**, *5*, 517.
- Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 2nd ed.; Plenum Press: New York, 1999.
- McGhee, J. D.; von Hippel, P. H. *J. Mol. Biol.* **1974**, *86*, 469.
- Brown, B. M.; Sauer, R. T. *Biochemistry* **1993**, *32*, 1354.