Distribution of labor among bZIP segments in the control of DNA affinity and specificity

Steven J Metallo and Alanna Schepartz*

Department of Chemistry, Yale University, New Haven, CT 06511-8118, USA

Background: The basic region–leucine zipper (bZIP) family of proteins use an atypically simple motif for DNA recognition, yet family members discriminate differently between target sites that differ only in half-site spacing. Two such sites are the cAMP-response element (CRE) and the AP-1 target site. Fos/Jun prefers the AP-1 site (ATGACTCAT), while CRE-BP1 prefers CRE (ATGACGTCAT), and GCN4 binds both sites with equal affinity. We therefore asked what determines the relative specificity for CRE and AP-1 sites in bZIP proteins.

Results: Here we show that CRE/AP-1 specificity in CRE-BP1 is encoded within the spacer and basic segments of the bZIP element. Of these two regions, the basic segment is the more important. This specificity is in part achieved at the expense of affinity.

Conclusions: The small size and simplicity of the bZIP recognition helix was already unusual; our findings show that the information that determines the target site specificity of members of the bZIP family of proteins is even more condensed than expected. These results suggest that it may be possible to design surprisingly small proteins that bind DNA with high sequence specificity, although it may be more difficult to achieve high-affinity binding in small proteins.

Chemistry & Biology November 1994, **1**:143–151

Key words: CREB, GCN4, half-site spacing, protein–nucleic acid interactions

Introduction

The basic region–leucine zipper (bZIP) class of eukaryotic transcription factors employs an atypically simple structural motif for the sequence-specific recognition of duplex DNA [1]. The DNA binding activity of these proteins is localized within 60 contiguous amino acids [2] termed the bZIP element (Fig. 1) [3]. Each bZIP element contains two segments that have distinct roles in DNA recognition [4]. A 'basic segment' of ~20 amino acids near the amino terminus of the element contains residues that contact DNA directly [3,5,6], while a 'zipper segment' of ~25 amino acids uses a heptad repeat of leucine residues and interspersed salt bridges to orchestrate the formation of a coiled-coil dimer from two protein monomers [4,7–10]. The basic and zipper segments are connected by a six-residue spacer segment whose sequence varies in the different members of the bZIP family [11]. Data from circular dichroism [12–16] and nuclear magnetic resonance (NMR) [7,17,18] studies show that in the absence of DNA the zipper segment can form a coiled-coil dimer, which is found in concentration-dependent equilibrium with the monomer, whereas the basic segment has little defined structure [12–14,17,18]. Helicity is induced within the basic segment upon DNA binding [12,13,15,16,19,20], resulting in a fully helical structure for the dimer, as observed in two different bZIP•DNA complexes [3,5]. The bZip proteins are distinct from other known DNA-binding motifs in that the 'recognition helix' is isolated and appears not to be stabilized by tertiary interactions with any other region of the protein [3,5].

Although structurally simple, bZIP proteins recognize a diverse set of inverted half sites within a 9–10 base-pair sequence, a length of DNA no smaller than those recognized by protein motifs containing more complex DNA-binding domains. In addition to their ability to discriminate between target sites that differ in half-site sequence (base-pair specificity), bZIP proteins share with proteins in the steroid hormone superfamily [21] and the class-3 zinc-binding domains [22] related to GAL4 [23,24] the ability to discriminate between target sites of identical half-site sequence but different half-site spacing (half-site spacing specificity). For example, bZIP proteins related to the oncogene products Fos and Jun (the AP-1 family) prefer the pseudosymmetric nine-base-pair AP-1 target site (ATGACTCAT) which comprises two ATGA half-sites arranged in an inverted pair and separated by a single dC:dG base pair. Conversely, bZIP proteins related to CREB and to ATF-2 (the CREB/ATF family) prefer the dyad-symmetric cAMP response element (CRE) target site (ATGACGTCAT) in which the same inverted pair of half-sites is separated by two base pairs (dC:dG and dG:dC) [25]. Within the context of a B-form DNA helix, the additional dG:dC base pair in the CRE target site displaces the two ATGA contact surfaces by an axial translation of 3.25 Å and a twist angle of 34.5°. This geometric operation moves the base and phosphate groups of one half site by ~4 Å and 7 Å, respectively [5]. Despite the presumed structural differences between the CRE and AP-1 target sites, the yeast bZIP protein GCN4 [26,27] operates in a mode distinct from the AP-1 or CREB/ATF families and binds both sites with comparable affinity [28].

Recently we described a series of synthetic molecules based on GCN4 whose DNA-binding properties

*Corresponding author.

 ISSN 1074–5521

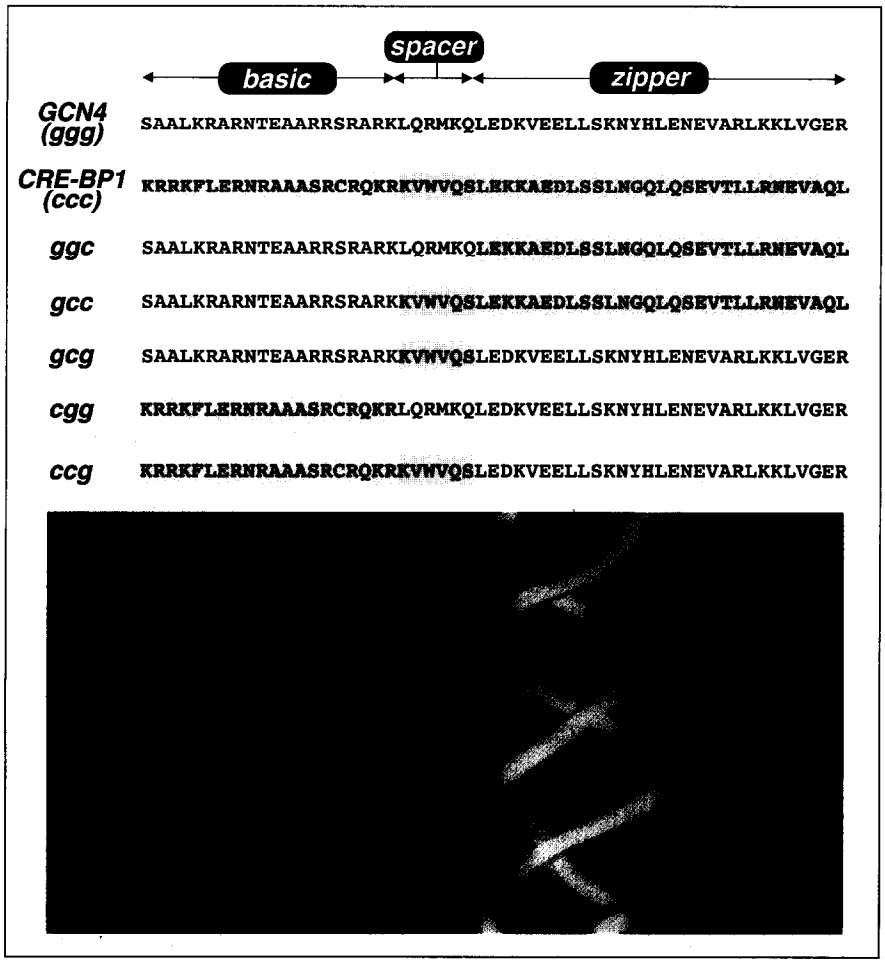

Fig. 1. **(a)** bZIP peptides and chimeras used in this study. ggg contains an amino-terminal serine followed by residues 228–281 of GCN4 [35]; ccc contains residues 354–408 of CRE-BP1 [31]. The basic segment of GCN4 is colored blue, the spacer segment red, and the zipper green. **(b)** Backbone ribbon trace of the GCN4 bZIP bound to the CRE target site [5]. Colors coded as in (a).

suggested a mechanism by which bZIP proteins might achieve half-site spacing specificity [20,29]. The metallopeptide dimers we constructed [30] contain two copies of a peptide comprising the basic and spacer segments of GCN4 assembled into a dimer with one of three differentially substituted *bis*(terpyridyl)iron(II) complexes (Fig. 2). Although both GCN4 and disulfide dimer peptides derived from GCN4 [12,19] bind the CRE and AP-1 sites with comparable affinity [5,28], the metallopeptide $[G_{29}T_S]_2Fe$ binds the CRE target site with 4.1 kcal mol^{-1} greater binding free energy (ΔG°_{obs}) than that with which it binds the AP-1 target site [20,29]. In other words, the *bis*(terpyridyl)iron(II) dimerization domain modifies the half-site spacing selectivity of the GCN4 basic and spacer segment peptide without altering its sequence.

The observation that a synthetic dimerization domain can alter the half-site spacing preference of a peptide containing the basic and spacer segments of GCN4 led us to question whether a natural dimerization domain could perform the same function: that is, whether the coiled coil formed by a CREB/ATF family member could impose half-site spacing selectivity on its adjacent basic-spacer segment peptide. To identify which segment of the bZIP element encodes CRE/AP-1 specificity, we have now generated a series of hybrid peptides by replacing portions of the GCN4 bZIP element with their counterparts from a member of the CREB/ATF family of bZIP proteins, CRE-BP1 [31] (Fig. 1). The boundaries between the basic, spacer and zipper segments were defined on the basis of sequence conservation; the spacer

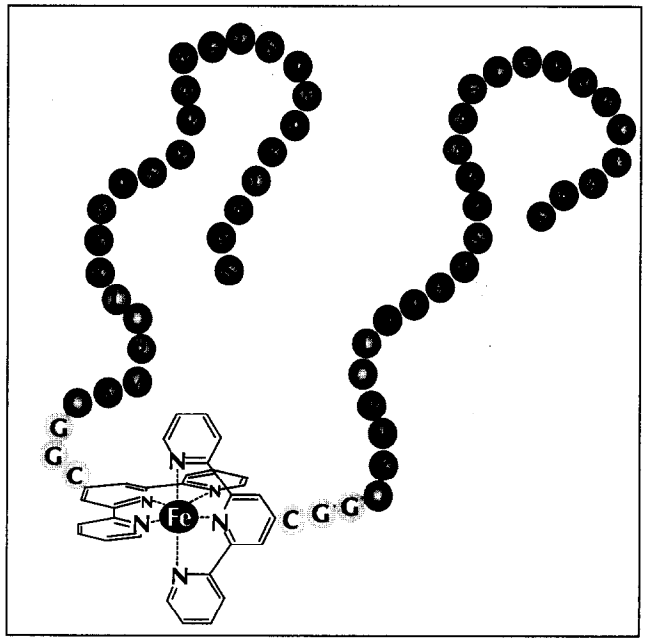

Fig. 2. Structure of $[G_{29}T_S]_2Fe$ [29]. Color code as in Fig. 1.

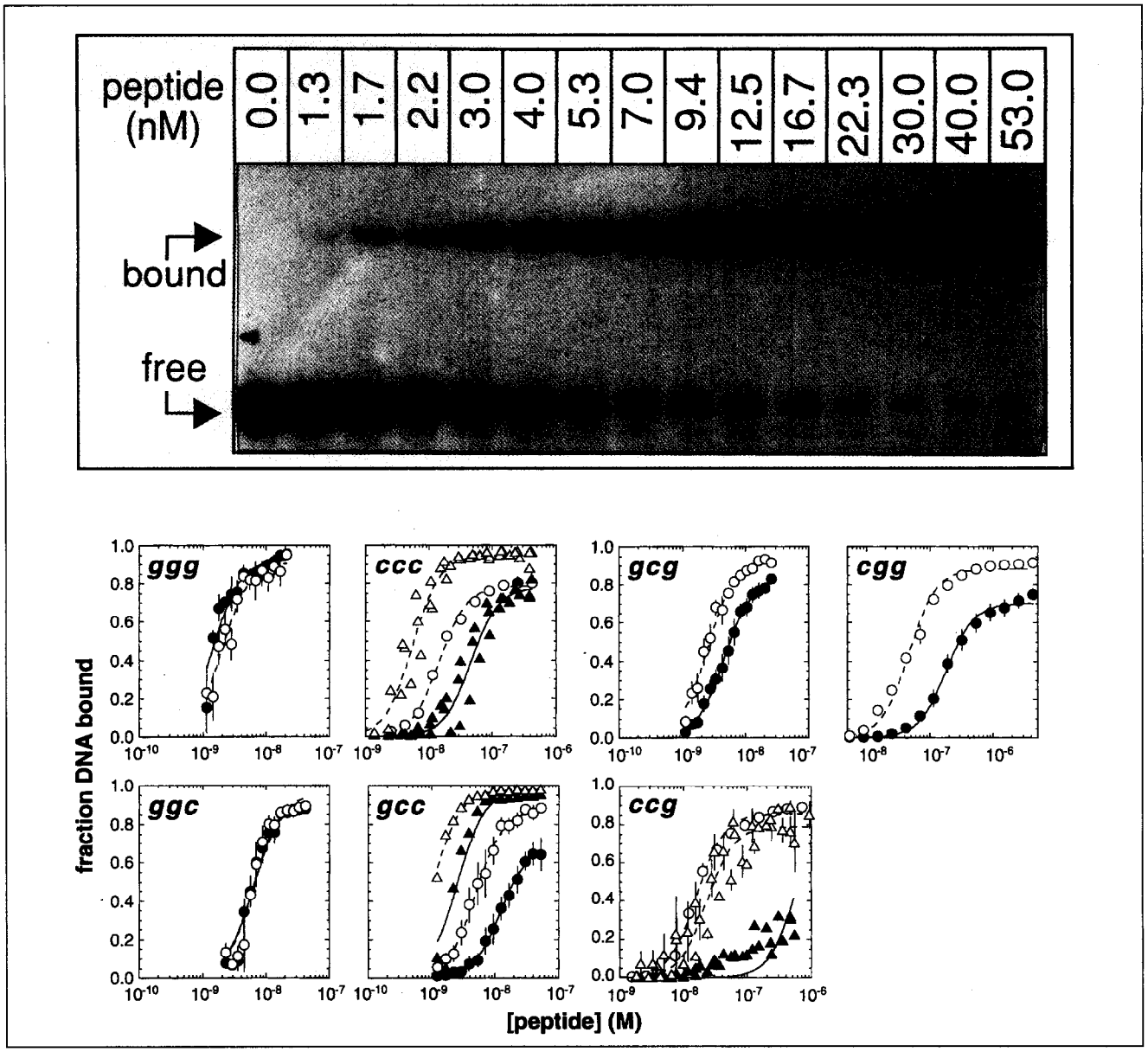

Fig. 3. Equilibrium binding of bZIP homodimers to DNA. **(a)** Autoradiogram of ^{32}P end-labeled CRE_{24} binding to increasing amounts of the gcg homodimer at 25 °C. The concentration of peptide monomer is shown above each lane. **(b)** Binding of peptide homodimers ggg, ccc, ggc, gcc, gcg, cgg, and ccg to CRE_{24} (open symbols) and $AP\text{-}1_{23}$ (closed symbols) target sites at 25 °C (circles) and at 4 °C (triangles). Solid and dotted lines represent the best fit of the data to the equation $\Theta = (S) * \frac{1}{1+\frac{K_{app}}{[A_{total}]^2}}$, where K_{app} and S are adjustable parameters. Note that the x-axis is not constant throughout the Figure.

segment is chosen as the six non-conserved residues separating the last conserved basic residue of the basic segment and the first conserved leucine residue of the zipper segment. The relative affinities of these hybrid peptides for the CRE and AP-1 target sites indicate that the coiled coil is not important in determining the half-site spacing preference of CRE-BP1. Instead, residues within the basic and spacer segments act to control half-site spacing selectivity. Our finding that the determinants of half-site spacing specificity, like the determinants of base-pair specificity, are encoded primarily within the basic and spacer segments show that, short though the bZIP recognition helix is, the information encoded within it is even more concentrated than expected.

Results

Seven peptides containing the basic, spacer, and zipper segments of either yeast GCN4 or human CRE-BP1 (Fig. 1) were prepared [32]. To determine the contribution of each segment to DNA affinity and specificity, we measured the equilibrium dissociation constants of the complexes formed between these peptide homodimers and either CRE_{24} or $AP\text{-}1_{23}$ using gel electrophoresis methods [33,34]. A typical binding experiment is illustrated in Fig. 3a, and binding isotherms for each of the seven peptides are shown in Fig. 3b. In each case, an equation that describes formation of a 2:1 peptide:DNA complex provides a good fit to the fraction of DNA bound plotted as a function of total peptide

Table 1. Equilibrium dissociation constants of bZIP peptide:DNA complexes.

	K_{app} (M^2)[a]		ΔG°_{obs} (kcal mol^{-1})		R[c]	$\Delta\Delta G^\circ_{obs}$ (kcal mol^{-1})
	AP-1	CRE	AP-1	CRE		
ggg	$(2.1 \pm 0.4) \times 10^{-18}$	$(4.1 \pm 0.6) \times 10^{-18}$	−24.1[b]	−23.7	0.5	−0.4[d]
ccc	N.D.	$(1.2 \pm 0.1) \times 10^{-15}$	−17.2[e]	−20.3		>2.3[e]
ccc	$(2.1 \pm 0.3) \times 10^{-15}$[f]	$(3.0 \pm 0.4) \times 10^{-17}$[f]	−18.6[f]	−20.9[f]	69[f]	2.3[f]
ggc	$(3.8 \pm 0.6) \times 10^{-17}$	$(4.1 \pm 0.7) \times 10^{-17}$	−22.4	−22.3	0.9	−0.1
gcc	$(1.5 \pm 0.9) \times 10^{-16}$	$(2.4 \pm 0.2) \times 10^{-17}$	−21.6	−22.7	6.3	1.1
gcc	$(6.5 \pm 1.3) \times 10^{-18}$[f]	$(1.1 \pm 0.1) \times 10^{-18}$[f]	−21.8[f]	−22.8[f]	5.9[f]	1.0[f]
gcg	$(2.1 \pm 0.1) \times 10^{-17}$	$(6.1 \pm 0.5) \times 10^{-18}$	−22.7	−23.5	3.4	0.8
cgg	$(3.3 \pm 0.3) \times 10^{-14}$	$(2.2 \pm 0.3) \times 10^{-15}$	−18.4	−20.0	15	1.6
ccg	N. D.	$(2.5 \pm 0.2) \times 10^{-16}$	−18.7[e]	−21.2		>2.5[e]
ccg	$> 1 \times 10^{-14}$[f]	$(5.0 \pm 0.9) \times 10^{-16}$[f]	<−16.8[f]	−19.4[f]	>20[f]	>2.5[f]

[a]determined as described in Materials and methods. Unless noted otherwise, all values refer to data obtained at 25 °C. All values reported represent the means of at least three determinations (except for gcc at 4 °C) ± SEM. [b]ΔG°_{obs} is equal to $-RT\ln(1/K_{app})$ where T is 298 (or 277) K and R is 0.001987 kcal mol^{-1} K^{-1}. [c]R is equal to K_{app} (AP-1)/K_{app} (CRE). [d]$\Delta\Delta G^\circ_{obs}$ is equal to $-RT\ln(1/R)$. [e]estimated based on $\Delta\Delta G^\circ_{obs}$ determined at 4 °C. We believe this estimate is reasonable because the $\Delta\Delta G^\circ_{obs}$ values determined for the gcc peptide at 25 °C and 4 °C are comparable. [f]determined at 4 °C.

monomer concentration. Calculated values for the apparent equilibrium dissociation constant (K_{app}) are shown in Table 1. With the exception of ccc and ccg, all peptide homodimers bind to the two DNA target sites tested. Dissociation constants fall between 2.1×10^{-18} M^2 and 3.3×10^{-14} M^2.

bZIP peptides ggg and ccc display the half-site spacing specificities of full-length GCN4 and CRE-BP1, respectively

Our results show that the ggg peptide binds with similar affinity to the CRE and AP-1 target sites. The two-fold preference of ggg for the AP-1 target site is consistent with results obtained with full-length GCN4 and GCN4 bZIP element peptides [19,28,29]. The dissociation constants measured for the ggg•CRE and ggg•AP-1 complexes are slightly lower (~10-fold) than those reported for the DNA complexes of a GCN4 bZIP element peptide containing two additional residues at the amino terminus [13] and are comparable to those measured for the DNA complexes of the full-length protein [13,35]. The specificity of the ccc peptide for the CRE target site is consistent with qualitative measurements on ATF-2 [36], which is virtually identical to CRE-BP1 [31] within the bZIP element. These results confirm that the DNA binding properties of the peptides studied here are relevant to the DNA-binding properties of the proteins from which they are derived.

The CRE-BP1 zipper segment affects affinity for DNA but not half-site spacing specificity

Peptide ggc contains the basic and spacer segments of GCN4 fused to the zipper segment of CRE-BP1. This peptide bound with high affinity as a homodimer to both CRE_{24} and $AP\text{-}1_{23}$ and exhibited no CRE/AP-1 selectivity. This result indicates that ggg cannot be converted into a CRE-specific peptide homodimer by replacing its coiled coil with one derived from CRE-BP1. Although our results with metallopeptides illustrated that the structure of the dimerization domain could alter the specificity of attached basic-segment peptides [20,29], ccc employs a different mechanism to achieve half-site spacing specificity.

Although the CRE-BP1 zipper segment does not impose CRE target site selectivity on the ggc homodimer, comparison of K_{app} for the CRE and AP-1 complexes of ggg and ggc reveals the effect of the CRE-BP1 zipper segment on CRE and AP-1 target site affinity. The measured ΔG°_{obs} for formation of the CRE and AP-1 complexes of ggc are less favorable by 1.4 and 1.7 kcal mol^{-1}, respectively, than those for the corresponding ggg complexes. This decrease in affinity could be due to a decrease in the thermodynamic stability of the coiled-coil dimer or to changes in conformation that alter the recognition interface of the bZIP element. The measured T_m values for all three peptides containing the CRE-BP1 zipper segment, however, are 20 °C (ggc) or less than 4 °C (ccc and gcc) at a peptide concentration of 220 μM, at least 40 °C lower than those containing the GCN4 zipper segment (Table 2). This suggests that at least part of the loss in affinity is due to the lower thermodynamic stability of the ccc coiled coil relative to the ggg coiled coil.

Residues within the spacer segment affect specificity; the effect on affinity is context-dependent

Substitution of the six spacer segment residues of GCN4 with their CRE-BP1 counterparts has a moderate effect on the preference of the resultant homodimer for the CRE target site. The gcg peptide is 1.2 kcal mol^{-1} more CRE-selective than the ggg peptide; the gcc peptide is 1.2 kcal mol^{-1} more selective than the ggc peptide; and the ccg peptide is ~0.9 kcal mol^{-1} more selective than the cgg peptide. Thus, the increase in specificity afforded by the CRE-BP1 spacer segment corresponds to between 0.9 and 1.2 kcal mol^{-1} and is relatively independent of the origins of the basic and zipper segment residues. This increase represents less than half of the total difference in selectivity between ccc and ggg (2.7 kcal mol^{-1}). In each

Table 2. Melting temperatures (T_m) for bZIP peptides and chimeras[a].

Peptide	T_m (°C)
ccc	<4
gcc	<4
ggc	20±5
gcg	66±7
cgg	68±7
ccg	65±7

[a]determined as described in Materials and methods.

case, increased discrimination results from a selection against the AP-1 target site. We note that the increase in specificity exhibited by the metallopeptide $[G_{29}T_S]_2Fe$ also results from a selection against the AP-1 target site [20]. These results demonstrate that the CRE-BP1 spacer segment sequence is important in determining the specificity of CRE-BP1 for the CRE target site.

Although the effect of the CRE-BP1 spacer segment on specificity is relatively independent of the origins of the basic and zipper segment residues, its effect on affinity is not. The CRE-BP1 spacer segment has a negligible effect on affinity for the CRE target site when fused to the GCN4 basic segment; this effect corresponds to a decrease in affinity of 0.2 kcal mol^{-1} when ggg and gcg are compared and an increase in affinity of 0.4 kcal mol^{-1} when ggc and gcc are compared. However, the same spacer segment has a significant stabilizing effect ($\Delta\Delta G^\circ = -1.2$ kcal mol^{-1}) on affinity for the CRE target site when fused to the CRE-BP1 basic segment (compare cgg and ccg). Thus the effect of the CRE-BP1 spacer segment is context-dependent; it is small in the context of a GCN4 basic segment and large in the context of a CRE-BP1 basic segment. This result suggests that there is communication between the basic and spacer segments that may be interpreted in light of recent findings regarding the conformations of the CRE and AP-1 target sites in complex with the bZIP peptides studied here; this will be discussed later.

Residues within the basic segment are important in half-site spacing specificity

Comparison of the relative affinities of the ggg and cgg peptide homodimers for the CRE and the AP-1 target sites illustrates the dramatic effect on affinity and specificity encoded within the CRE-BP1 basic segment. Substitution of the CRE-BP1 basic segment for the GCN4 basic segment lowers affinity for the AP-1 target site by 5.7 kcal mol^{-1} and lowers affinity for the CRE target site by 3.7 kcal mol^{-1}. The result is a net increase in differential binding energy $\Delta\Delta G^\circ_{obs}$ of 1.6 kcal mol^{-1}, a value which approaches that exhibited by ccc itself. The effect of the CRE-BP1 basic segment is also seen in the context of a CRE-BP1 spacer segment and a GCN4 zipper segment: substitution of the basic segment in gcg with that of CRE-BP1 lowers affinity for the CRE target site by 2.3 kcal mol^{-1} to generate a peptide (ccg) whose differential binding energy equals that of ccc. Nine of the twelve basic and spacer segment residues shown to contact the CRE target site in the GCN4•CRE complex are conserved in CRE-BP1. Therefore, either CRE-BP1 is able to convert these small differences into a highly specific recognition interface, or residues elsewhere within the basic segment control specificity [37], or the CRE-BP1 spacer segment orients conserved basic segment residues for specific recognition of the CRE target site.

Discussion

The experiments described here were stimulated by the observation that a 'metallopeptide' containing two copies of the GCN4 basic-spacer segment peptide joined to a *bis*(terpyridyl)iron(II) complex binds selectively ($\Delta\Delta G^\circ_{obs}$ > 4 kcal mol^{-1}) to the CRE target site [20,29], even though GCN4 binds the CRE and AP-1 target sites with equal affinity. This observation demonstrates that it is possible to alter the half-site spacing specificity of a basic segment peptide through changes in dimerization domain structure; we set out to determine whether this mechanism is exploited by a CREB/ATF protein. It is not. The half-site specificity of CRE-BP1 is independent of the origins of the zipper segment. CRE/AP-1 selectivity, however, is encoded by residues within the basic and spacer segments, with the distribution of labor favoring the basic segment.

Others have also observed the contribution of spacer segment residues to CRE/AP-1 selectivity. Johnson [38] synthesized and characterized a series of molecules containing sequences from either GCN4 or C/EBP (C/EBP, like CRE-BP1, prefers a dyad-symmetric 10-bp target site). Qualitative data from electrophoretic mobility shift experiments revealed that residues within the C/EBP spacer segment lowered the affinity of the chimeric peptide for the AP-1 target site. Kim *et al.* [39] showed that mutation of three residues in the spacer segment of GCN4 (Leu247, Arg249, Met250) to the corresponding residues in CREB (Lys304, Tyr306, and Val307) generated a protein that bound preferentially by a factor of two to the CRE target site. Here we show that replacing all six residues of the GCN4 spacer segment with those of CRE-BP1 generates a molecule (gcg) that binds preferentially by a factor of four to the CRE target site. In this system, selectivity results from a diminished affinity for the AP-1 target site: when compared with the ggg peptide, the gcg peptide binds ~0.2 kcal mol^{-1} less well to the CRE target site but 1.4 kcal mol^{-1} less well to the AP-1 target site. In terms of a distribution of labor, the thermodynamic preference of the ccc homodimer peptide for the CRE target site ($\Delta\Delta G^\circ_{obs}$ = 2.3 kcal mol^{-1}) is equal to the sum of the specificities contributed by each of its three composite segments: the thermodynamic preferences of the cgg, gcg, and ggc for the CRE target site are 1.6, 0.8 and –0.1 kcal mol^{-1}, respectively. Therefore, our results are consistent with those of other workers who observed small changes in CRE/AP-1 selectivity that correlate with the sequence changes within the spacer

segment. We also find, however, that the basic segment is important in determining CRE/AP-1 specificity.

Specificity is achieved at the expense of affinity

The enhanced specificity exhibited by peptides containing the CRE-BP1 basic segment is accompanied by a considerable loss in binding free energy compared to their less selective counterparts that contain the GCN4 basic segment. The affinity of the ggg homodimer for the CRE target site is 4.2 kcal mol^{-1} higher than the affinity of the ccc homodimer for the same site, and 3.7 kcal mol^{-1} higher than that of the cgg homodimer. These observations may be interpreted in the context of recent studies on the conformations of the CRE and AP-1 target sites, free and in complex with each of the peptides shown in Fig. 1 [37]. These results, obtained via helical phasing analysis [40], demonstrate that the dyad-symmetric CRE target site is bent towards the major groove by ~13° in the absence of any protein [37]. Binding of the ggg peptide to the CRE target site maintains the magnitude and direction of this intrinsic bend. Binding of the ccc peptide to the CRE target site, however, is accompanied by curvature towards the minor groove which negates the intrinsic bend and in effect 'straightens' the DNA. Analogous phasing experiments performed with the other peptides in Fig. 1 demonstrate that DNA bending towards the minor groove is uniformly absent in all bZIP•CRE complexes containing the basic segment of GCN4 and present in complexes containing the basic segment of CRE-BP1 ([37] and unpublished data). This correlation parallels the results reported here that peptides containing the basic segment of CRE-BP1 have affinities for the CRE target site that are systematically lower than those of corresponding peptides containing the basic segment of GCN4. Thus, the observed drop in CRE affinity may reflect energetic costs of DNA distortion which are not adequately compensated for by favorable protein•DNA contacts.

The effect of the CRE-BP1 basic segment on the conformations of the CRE and AP-1 target sites [37] may also explain the context-dependent effect of the CRE-BP1 spacer segment. The CRE-BP1 spacer segment increases affinity for the CRE target site when fused to the CRE-BP1 basic segment but has little effect when fused to the GCN4 basic segment. Put another way, the CRE-BP1 spacer segment increases affinity for the CRE target site only when that site is 'straightened' towards the minor groove [37]. It is possible that the induced bending of the CRE target site observed with peptides containing the CRE-BP1 basic segment [37] brings certain nucleic acid recognition elements into position to contact protein recognition elements within the CRE-BP1 spacer segment. This contact would increase the affinity of ccg for the CRE target site relative to cgg. The free energy gained by this contact in the ccc complex may compensate in part for the lower thermodynamic stability of the CRE-BP1 zipper segment.

bZIP proteins: induced fit or induced structure?

The DNA-binding motif employed by bZIP proteins is distinctive in its structural simplicity. Recognition is achieved through interactions of a single, extended, solvent-exposed helix with the DNA major groove. Whereas other α-helical DNA-binding motifs use globular domains to form a scaffold for the recognition helix, the bZIP recognition helix acts alone. No tertiary constraints appear to be imposed on the helix [3,5], and the protein takes up a helical conformation only in the presence of DNA [12–16,19,41]. On this basis, bZIP proteins might be expected to be more 'flexible' or 'adaptable' than their globular counterparts, and correspondingly better equipped to accommodate subtle, sequence-dependent variations in duplex structure [3,42,43]. The bZIP protein GCN4 illustrates this in what is, perhaps, a classic example of 'induced fit' [44]: the helical structure it assumes in the presence of the CRE and AP-1 target sites appears to require no major changes in DNA structure [37].

The simplicity of the bZIP motif belies the complexity of its binding, however, and extension of this induced fit model to all members of the bZIP family is unwarranted. GCN4 has proven to be the exception rather than the rule among bZIP proteins; many members are unable to present a recognition surface suitable for their target site and the DNA must be distorted to ensure a proper fit [37,45–47]. Thus, the absence of complicated tertiary structure, a characteristic believed to be shared by members of the bZIP family, does not always lead to induced fit of the peptide to its DNA target. In the case of CRE-BP1, selectivity for the CRE target site is achieved through differential lowering of CRE and AP-1 target site affinities for the CRE-BP1 basic segment relative to that for the GCN4 basic segment. Distortions of the CRE target site upon formation of the ccc•CRE and cgg•CRE complexes are correlated with this lowering in affinity. This indicates that the basic segment of CRE-BP1 fails to present a helical recognition interface that complements the structure presented by the free CRE target site. It therefore falls to the DNA to adapt to the set of helical conformations accessible to the CRE-BP1 basic segment.

Significance

The control of transcription in eukaryotes that occurs in response to extracellular signals or during development requires the accurate assembly of multi-protein complexes at promoters and enhancer sites. Sequence-specific transcription factors are central in this process; they recognize a defined sequence of DNA and select which gene will be activated. For members of the bZIP transcription factor family, this recognition event requires discrimination between two DNA sequences that differ only in the presence or absence of a single base pair, a type of specificity termed half-site spacing selectivity. We found previously that the half-site spacing selectivity of a bZIP protein could

be altered without changing the identity of the DNA contact residues, simply by swapping the natural protein dimerization domain — the coiled coil — for a transition metal complex. This led us to question whether the determinants of bZIP half-site spacing selectivity were encoded within the natural dimerization domain, that is, within the coiled coil.

The quantitative affinity measurements presented here show that the coiled coil does not influence the half-site spacing specificity of the bZIP proteins we examined. Half-site spacing preferences are encoded within the basic and spacer segments of the bZIP element. Of these two regions, the basic segment is the more important, although both segments are required for maximal specificity. Thus, the bZIP recognition helix is even more efficient than expected in terms of DNA-binding ability per unit size of protein, and is startlingly more efficient than DNA-binding proteins that have globular DNA-binding domains. These results suggest that it may be possible to design surprisingly small proteins that bind DNA with high sequence-specificity. The challenge for the future will be to achieve both high affinity and high specificity in small proteins.

Materials and methods

Peptides

A 56-amino acid peptide containing the sequence Gly-Ser at its amino-terminus followed by the bZIP element of GCN4 (residues 228–281), labeled ggg in Fig. 1, was obtained from Professor Jon Shuman (University of Alabama, Birmingham). A peptide comprising the bZIP element of CRE-BP1 (residues 354–408, labeled ccc in Fig. 1), and five chimeric peptides ggc, gcg, gcc, cgg, and ccg were synthesized by standard solid-phase techniques [32] at the WM Keck Foundation Biotechnology Resource Laboratory in the Yale University School of Medicine. Synthetic peptides were purified by reverse-phase high performance liquid chromatography (HPLC) with a Vydac C_{18} column (4.6 x 250 mm) and characterized by electrospray (fast atom bombardment (FAB) mass spectrometry, *meta*-nitrobenzyl alcohol and trifluoroacetic acid matrix) or laser desorption (LD, α-cyano-4-hydroxycinnamic acid matrix) mass spectrometry and by amino acid analysis (AAA). All peptides analyzed correctly for amino acid content. Mass spectral data: ccc, [M^+] predicted: 6466.5, found: 6464.9 (LD); gcc, [M+H] predicted: 6206.1, found: 6205.8 (FAB); ggc, [M+H] predicted: 6263.2, found: 6263.6 (FAB); gcg, [M+] predicted: 6430.4, found: 6431.8 (LD); cgg, [M+1] predicted: 6750.9, found: 6752.1 (LD); ccg, [M+1] predicted: 6693.8, found: 6695.2 (LD). Peptide concentrations were determined by amino acid analysis in triplicate using homoserine and norleucine as internal standards. Stock solutions (2–20 μM) were dissolved in a buffer containing 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2PO_4, 1.4 mM KH_2PO_4 (pH 7.4), 1 mM EDTA, 1 mM DTT, 0.1 % NP40.

DNA fragments

Four oligonucleotides, one containing the CRE target site (CRE_{24A}: 5'-AGTGGAGATGACGTCATCTCGTGC-3', ε_{260} = 2.34 x 10^5 M^{-1} cm^{-1}) and its complementary strand CRE_{24B} (5'-GCACGAGATGACGTCATCTCCACT-3', ε_{260} = 2.27 x 10^5 M^{-1} cm^{-1}), a second containing the AP1 site ($AP1_{23A}$: 5'-AGTGGAGATGACTCATCTCGTGC-3', ε_{260} = 2.28 x 10^5 M^{-1} cm^{-1}) and its complementary strand $AP1_{23B}$ (5'-GCACGAGATGAGTCATCTCCACT-3', ε_{260} = 2.17 x 10^5 M^{-1} cm^{-1}), were synthesized using automated solid-phase methods (Applied Biosystems 380B Synthesizer) [48]. Each strand was purified by preparative denaturing (20 % polyacrylamide, 20:1 acrylamide:bis) gel electrophoresis [49]. Oligonucleotides were radiolabeled on the 5'-end and annealed to their unlabeled complement by heating an equimolar mixture of the two fragments to 85 °C (in 100 mM KCl, 10 mM phosphate (pH 7.4)) for 2 min then cooling slowly to room temperature.

Electrophoretic mobility shift assays

Binding reactions were performed in PBS binding buffer (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2PO_4, 1.4 mM KH_2PO_4 (pH 7.4), 1 mM EDTA, 1 mM DTT, 0.1 % NP40, 0.4 mg ml^{-1} acetylated BSA, and 5 % glycerol). Equilibrium dissociation constants of peptide•DNA complexes were determined using an electrophoretic mobility-shift assay [33,34]. In a typical procedure, a given peptide was diluted serially from a stock solution of known concentration into PBS binding buffer. To the peptide solution was added labeled DNA (in PBS binding buffer) to a final concentration of < 50 pM (400–600 counts per minute (cpm)). Binding reactions were incubated for 30 min at 25 °C or 4 °C, then applied to a running 16 x 18 cm nondenaturing 8 % polyacrylamide (80:1 acrylamide:bisacrylamide) gel prepared in 45 mM Tris-borate, 1 mM EDTA, pH = 8.0 (0.5 x TBE). The electrophoresis running buffer contained 22.5 mM Tris-borate, 0.5 mM EDTA, pH = 8.0. Gels were pre-equilibrated for 30 min at 300 V and were maintained at a constant temperature during electrophoresis by immersion in a circulating, temperature-controlled water bath. Samples were loaded and subjected to electrophoresis at 300 V for 1 h at 25 °C, or at 300 V for 3 h at 4 °C. The amounts of complexed and free DNA were quantified on a Betagen 603 Blot Analyzer (Betagen Inc., Waltham, MA). The fraction DNA bound (Q) was calculated as the cpm contained within the band corresponding to the bound DNA divided by the sum of the cpm present in the bands corresponding to the bound and free DNAs within each lane. Within each experiment, the sum of the cpm contained in the bound and free areas was constant to within 10 %. A series of binding time-courses were performed to confirm that all peptides studied reached equilibrium with their DNA targets within 30 min. In each case, the fraction of DNA bound in the presence of a given peptide concentration after a 30 min incubation was within experimental error of the fraction of DNA bound after a 60 min incubation (data not shown).

Calculations

It has not been established whether all bZIP proteins bind DNA as pre-assembled dimers in a single step or whether two protein monomers bind sequentially with dimerization occurring on the DNA [3]. If dimerization precedes DNA binding, the binding reaction will be described by the following scheme:

Scheme 1 $$2U \underset{K_1}{\rightleftharpoons} A_2 \overset{O}{\underset{K_2}{\rightleftharpoons}} A_2O$$

Here, U represents unfolded bZIP monomer, A_2 represents coiled-coil homodimer, O represents duplex DNA, and A_2O represents the DNA•peptide homodimer complex. K_1 represents the dissociation constant of the peptide homodimer and K_2

represents the dissociation constant of the complex of this homodimer and DNA. For this case the following equilibrium and mass conservation equations hold:

$$K_1=\frac{[U]^2}{[A_2]} \quad (1)$$

$$K_2=\frac{[A_2][O]}{[A_2O]} \quad (2)$$

$$A_{total}=[U]+2[A_2]+2[A_2O] \quad (3)$$

Under our experimental conditions, the total peptide concentration, A_{total}, exceeds the total DNA concentration (< 50 pM) by a factor of at least 10. As a result, the term $[A_2O]$ in equation (3) is small compared with A_{total} and can be ignored. Equation 3 is simplified further by the consideration that the values of K_1 for the peptide homodimers studied here are also greater than A_{total}. The value of K_1 for GCN4p, a peptide that differs from ggg by four residues at the amino terminus, is ~5 μM under experimental conditions almost identical to those used here (50 mM potassium phosphate pH 7.0, 200 mM KCl, 25 °C) [13]; we assume that the K_1 for ggg and gcg are similar based on their similar T_m values (67 °C [14] versus 66 ± 7 °C, respectively). If so, at the highest concentration of ggg or gcg used (30 nM), the concentration of homodimer A_2 will be no greater than 2 % of A_{total} and may be ignored. For ccc, gcc and ggc, we estimate K_1 values greater than 100 μM and as a result, the concentration of these homodimers will be negligible at all protein concentrations studied. For cgg and ccg, we can assign K_1 values that are comparable to that of gcg based on the comparable T_m values of these three peptides (68 ±7 °C (cgg) versus 66 ± 7 °C (gcg) versus 65 ± 7 °C (ccg)). The concentrations of cgg and ccg required to bind DNA are considerably higher than that required of gcg, however. At the top of the cgg•AP-1 titration, the cgg concentration is 500 nM and ~15 % of the peptide is in the dimer state. At the cgg concentration required for half-maximal binding of the AP-1 target site (which coincides with the top of the cgg•CRE titration), the cgg concentration is 200 nM, corresponding to ~7 % dimer. Similarly, the ccg concentration at the top of the ccg•CRE titration is 200 nM, corresponding to ~7 % dimer. If a value for K_1 of 5 μM is defined, and the data describing formation of the cgg•CRE, cgg•AP-1, and ccg•CRE complexes are fit to an equation which expresses [U] in terms of K_1 and A_{total} [50], the values obtained for K_{app} fall within the error of the values obtained by approximating [U] with A_{total}. Thus it is possible to ignore the term $[A_2]$ in equation 3 and closely approximate $[A_{total}]$ with [U]. In this regime the fraction DNA bound (Θ) is given by:

$$\Theta=\frac{1}{1+\frac{K_{app}}{[A_{total}]^2}} \quad (4)$$

The discrimination ratio R is defined as the ratio of K_{app} for binding to the AP-1 target site divided by K_{app} for binding to the CRE target site.

An alternative binding model, recently established for the binding of LexA to operator DNA [51], involves the sequential binding of two monomers to the target DNA, as represented by scheme 2:

Scheme 2 $$U \underset{K_{m1}}{\overset{O}{\rightleftharpoons}} AO \underset{K_{m2}}{\overset{U}{\rightleftharpoons}} A_2O$$

Here, K_{m1} represents the dissociation constant for binding the first protein monomer to DNA and K_{m2} represents the dissociation constant for binding the second monomer to the complex AO. In this case the relevant equilibrium and mass action equations are:

$$K_{m1}=\frac{[O][U]}{[AO]} \quad (5)$$

$$K_{m2}=\frac{[U][AO]}{[A_2O]} \quad (6)$$

$$\text{and } A_{total}=[U]+[AO]+2[A_2O] \quad (7)$$

Once again [U] will closely approximate A_{total} under our experimental conditions and the fraction DNA bound (Θ) is given by:

$$\Theta=\frac{1}{1+\frac{K_{m1}K_{m2}}{[U]^2}}=\frac{1}{1+\frac{K_{app}}{[A_{total}]^2}} \quad (8)$$

Substitution of A_{total} for [U] yields an expression for the dependence of Θ on A_{total} that is identical to equation (4). Therefore, regardless of the pathway that the reaction follows, fitting the data to equation (4) will yield the correct K_{app}. Curve-fitting was performed using Kaleidagraph 3.0.2 (Abelbeck Software, Reading, PA). Dissociation constants estimated by nonlinear least squares analysis of the data are shown in Table 1.

Circular dichroism experiments

Circular dichroism (CD) experiments were performed on an Aviv 62DS spectrometer with a 0.1 cm pathlength cell. Samples were dissolved in 50 mM potassium phosphate pH 7.0, 200 mM KCl, and contained 220 μM peptide monomer. Temperature was increased at a rate of 1 °C per min. Spectra were baseline-corrected but were not smoothed. For each peptide, the T_m was determined by taking the first derivative of the CD signal (Θ) with respect to the reciprocal of the temperature in degrees Kelvin and finding the minimum of this function [52]. The error in the measurement of T_m was taken as the width of the dΘ/d(1/T) plot.

Acknowledgements: This work was supported by the American Cancer Society, with additional support provided by The National Foundation for Cancer Research, the David and Lucile Packard Foundation (Fellowship in Science and Engineering), Eli Lilly & Co. (Grantee in Biochemistry), the National Science Foundation (Presidential Young Investigator Award), the Camille and Henry Dreyfus Foundation (Teacher–Scholar Award), the Alfred P. Sloan Foundation (Research Fellowship) and by a 1995 Arthur C. Cope Scholar Award.

References

1. McKnight, S.L. (1991). Molecular zippers in gene regulation. *Sci. Am.* **264**, 54–64.
2. Hope, I. & Struhl, K. (1987). GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *EMBO J.* **6**, 2781–2784.
3. Ellenberger, T.E., Brandl, C.J., Struhl, K. & 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. Landschulz, W.H., Johnson, P.F. & McKnight, S.L. (1988). The

leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759–1764.
5. König, P. & Richmond, T. (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.* **230**, 139–154.
6. Blatter, E.E., Ebright, T.W. & Ebright, R.H. (1992). Identification of an amino acid–base contact in the GCN4–DNA complex by bromouracil-mediated photocrosslinking. *Nature* **359**, 650–652.
7. Oas, T.G., McIntosh, L.P., O'Shea, E.K., Dahlquist, F.W. & Kim, P.S. (1990). Secondary sructure of a leucine zipper determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **29**, 2891–2894.
8. O'Shea, E.K., Rutkowski, R. & Kim, P.S. (1989). Evidence that the leucine zipper is a coiled coil. *Science* **243**, 538–542.
9. Rasmussen, R., Benvegnu, D., O'Shea, E.K., Kim, P.S. & Alber, T. (1991). X-ray scattering indicates that the leucine zipper is a coiled coil. *Proc. Natl. Acad. Sci. USA* **88**, 561–564.
10. O'Shea, E.K., Klemm, J.D., Kim, P.S. & Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**, 539–544.
11. Pu, W.T. & Struhl, K. (1991). The leucine zipper symmetrically positions the adjacent basic regions for specific DNA binding. *Proc. Natl. Acad. Sci. USA* **88**, 6901–6905.
12. Talanian, R.V., McKnight, C.J. & Kim, P.S. (1990). Sequence-specific DNA binding by a short peptide dimer. *Science* **249**, 769–771.
13. Weiss, M.A., *et al.*, & Struhl, K. (1990). Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **347**, 575–578.
14. Weiss, M. (1990). Thermal unfolding studies of a leucine zipper domain and its specific DNA complex: implications for scissor's grip recognition. *Biochemistry* **29**, 8020–8024.
15. O'Neil, K.T., Hoess, R.H. & DeGrado, W.F. (1990). Design of DNA-binding peptides based on the leucine zipper motif. *Science* **249**, 774–778.
16. O'Neil, K.T., Shumann, J.D., Ampe, C. & DeGrado, W.F. (1991). DNA-induced increase in the α-helical content of C/EBP and GCN4. *Biochemistry* **30**, 9030–9034.
17. Saudek, V., *et al.*, & Gibson, T. (1991). The solution structure of a leucine-zipper motif peptide. *Protein Eng.* **4**, 519–529.
18. Saudek, V., *et al.*, & Pastore, A. (1991). Solution structure of the basic region from the transcriptional activator GCN4. *Biochemistry* **30**, 1310–1317.
19. Talanian, R.V., McKnight, C.J., Rutkowski, R. & Kim, P.S. (1992). Minimum length of a sequence-specific DNA binding peptide. *Biochemistry* **31**, 6871–6875.
20. Cuenoud, B. & Schepartz, A. (1993). Altered specificity of DNA-binding proteins with transition metal dimerization domains. *Science* **259**, 510–513.
21. Umesono, K. & Evans, R.M. (1989). Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell* **57**, 1139–1146.
22. Harrison, S.C. (1991). A structural taxonomy of DNA-binding domains. *Nature* **353**, 715–719.
23. Corton, J.C. & Johnston, S.A. (1989). Altering DNA-binding specificity of GAL4 requires sequences adjacent to the zinc finger. *Nature* **340**, 724–727.
24. Reece, R.J. & Ptashne, M. (1993). Determinants of binding-site specificity among yeast C6 zinc cluster proteins. *Science* **261**, 909–911.
25. Hai, T., Liu, F., Coukos, J. & Green, M.R. (1989). Transcription factor ATF clones: an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. *Genes Dev.* **3**, 2083–2090.
26. Penn, M.D., Galgoci, B. & Greer, H. (1983). Identification of AAS genes and their regulatory role in general control of amino acid biosynthesis in yeast. *Proc. Natl. Acad. Sci. USA* **80**, 2704–2708.
27. Hinnebusch, A.G. & Fink, G.R. (1983). Positive regulation in the general control of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **80**, 5374–5378.
28. Sellers, J.W., Vincent, A.C. & Struhl, K. (1990). Mutations that define the optimal half-site for binding yeast GCN4 activator protein and identify an ATF/CREB-like repressor that recognizes similar DNA sites. *Mol. Cell. Biol.* **10**, 5077–5086.
29. Cuenoud, B. & Schepartz, A. (1993). Design of a metallo-bZIP protein that discriminates between CRE and AP1 target sites: selection against AP1. *Proc. Natl. Acad. Sci. USA* **90**, 1154–1159.
30. Cuenoud, B. & Schepartz, A. (1991). A general scheme for incorporating nonnatural functionality into peptides. *Tetrahedron Lett.* **32**, 3325–3328.
31. Maekawa, T,. *et al.*, & Ishii, S. (1989). Leucine zipper structure of the protein CRE-BP1 binding to the cyclic AMP response element in brain. *EMBO J.* **8**, 2023–2028.
32. Stewart, J.M. & Young, J.D. (1984). *Solid Phase Peptide Synthesis*. Raven Press, New York.
33. Fried, M. & Crothers, D.M. (1981). Equilibria and kinetics of lac repressor-operator interactons by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**, 6505–6525.
34. Garner, M.M. & Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res.* **9**, 3047–3060.
35. Hope, I. & Struhl, K. (1985). GCN4 protein, synthesized *in vitro*, binds HIS3 regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast. *Cell* **43**, 177–188.
36. Hai, T., Liu, F., Allegretto, E.A., Karin, M. & Green, M.R. (1988). A family of immunologically related transcription factors that includes multiple forms of ATF and AP-1. *Genes Dev.* **2**, 1216–1226.
37. Paolella, D.N., Palmer, C.R. & Schepartz, A. (1994). DNA targets for certain bZIP proteins distinguished by an intrinsic bend. *Science* **264**, 1130–1133.
38. Johnson, P.F. (1993). Identification of C/EBP basic region residues involved in DNA sequence recognition and half-site spacing preferences. *Mol. Cell. Biol.* **13**, 6919–6930.
39. Kim, J., Tzamarias, D., Ellenberger, T., Struhl, K. & Harrison, S.C. (1993). Adaptability at the protein–DNA interface is an important aspect of sequence recognition by bZIP proteins. *Proc. Natl. Acad. Sci. USA* **90**, 4513–4517.
40. Zinkel, S.S. & Crothers, D.M. (1987). DNA bend direction by phase sensitive detection. *Nature* **328**, 178–181.
41. Santiago-Rivera, Z.I., Williams, J.S., Gorenstein, D.G. & Andrisani, O.M. (1993). Bacterial expression and characterization of the CREB bZip module: circular dichroism and 2D-^{1}H NMR studies. *Protein Science* **2**, 1461–1471.
42. Ellenberger, T. (1994). Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix–loop–helix DNA binding domains. *Curr. Biol.* **4**, 12–21.
43. Alber, T. (1993). How GCN4 binds DNA. *Curr. Biol.* **3**, 182–184.
44. Koshland, D.E. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA* **44**, 98–104.
45. Kerppola, T.K. & Curran, T. (1991). Fos-Jun heterodimers and Jun homodimers bend DNA in opposite orientations: implications for transcription factor cooperativity. *Cell* **66**, 317–326.
46. Kerppola, T.K. & Curran, T. (1991). DNA bending by Fos and Jun: the flexible hinge model. *Science* 254, 1210–1214.
47. Kerppola, T.K. & Curran, T. (1993). Selective DNA bending by a variety of bZIP proteins. *Mol. Cell. Biol.* **13**, 5479–5489.
48. Beaucage, S.L. & Caruthers, M.H. (1981). Deoxynucleoside phosphoramidites — a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22**, 1859–1862.
49. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1987). *Molecular Cloning: A Laboratory Manual.* (2nd edn), Cold Spring Harbor Press, New York.
50. Brown, B.M. & Sauer, R.T. (1993). Assembly of the Arc repressor-operator complex: cooperative interactions between DNA-bound dimers. *Biochemistry* **32**, 1354–1363.
51. Kim, B. & Little, J.W. (1993). Dimerization of a specific DNA-binding protein on the DNA. *Science* **255**, 203–206.
52. Cantor, C.R. & Schimmel, P.R. (1980). *Biophysical Chemistry*. Freeman, New York.

Received: **9 Sep 1994**; revisions requested: **22 Sep 1994**; revisions received: **29 Sep 1994**. Accepted: **30 Sept 1994**.