MOLECULAR AND CELLULAR BIOLOGY, Nov. 1993, p. 6919–6930
0270-7306/93/116919-12$02.00/0

Vol. 13, No. 11

Identification of C/EBP Basic Region Residues Involved in DNA Sequence Recognition and Half-Site Spacing Preference

PETER F. JOHNSON

ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, P.O. Box B, Frederick, Maryland 21702-1201

Received 5 May 1993/Returned for modification 1 July 1993/Accepted 11 August 1993

C/EBP and GCN4 are basic region-leucine zipper (bZIP) DNA-binding proteins that recognize the dyad-symmetric sequences ATTGCGCAAT and ATGAGTCAT, respectively. The sequence specificities of these and other bZIP proteins are determined by their α-helical basic regions, which are related at the primary sequence level. To identify amino acids that are responsible for the different DNA sequence specificities of C/EBP and GCN4, two kinds of hybrid proteins were constructed: GCN4-C/EBP chimeras fused at various positions in the basic region and substitution mutants in which GCN4 basic region amino acids were replaced by the corresponding residues from C/EBP. On the basis of the DNA-binding characteristics of these hybrid proteins, three residues that contribute significantly to the differences in C/EBP and GCN4 binding specificity were defined. These residues are clustered along one face of the basic region α helix. Two of these specificity residues were not identified as DNA-contacting amino acids in a recently reported crystal structure of a GCN4-DNA complex, suggesting that the residues used by C/EBP and GCN4 to make base contacts are not identical. A random binding site selection procedure also was used to define the optimal recognition sequences for three of the GCN4-C/EBP fusion proteins. These experiments identify an element spanning the hinge region between the basic region and leucine zipper domains that dictates optimal half-site spacing (either directly abutted for C/EBP or overlapping by one base pair for GCN4) in high-affinity binding sites for these two proteins.

Basic region-leucine zipper (bZIP) DNA-binding proteins constitute a major class of eukaryotic transcriptional regulatory proteins (28). More than 50 genes encoding these proteins have been isolated from a variety of organisms. Most bZIP proteins can be classified into one of five subfamilies based on their DNA sequence specificities (Fig. 1). The binding sites recognized by each subfamily consist of related but distinct palindromic sequences containing two 5-bp half-sites. These consensus binding sites differ in the sequence of their half-sites and/or in the spacing between half-sites, which can be either directly abutted or overlapping by one base pair. This spacing difference alone distinguishes the binding of the Fos/Jun and ATF/CREB subfamilies.

The bZIP DNA-binding domain is characterized by two major features: a helix-permissive segment containing repeated leucine residues spaced at seven-amino-acid intervals (the leucine zipper) and an associated segment of strong net positive charge (the basic region) immediately adjacent to the leucine repeats (33). The leucine repeats are part of an amphipathic α helix that mediates protein dimerization (31, 33, 34, 63), and the paired helices in a bZIP dimer lie in a parallel orientation, as in classical coiled-coil structures (15, 50). Recently, the crystal structures of a synthetic leucine zipper peptide (49) and a complete bZIP DNA-binding domain (13) have been solved, proving that the leucine zipper dimer is a coiled coil composed of parallel α helices. Kouzarides and Ziff (30) and Vinson et al. (77) noted that an array of residues within the basic region, termed the basic motif, is shared by all leucine zipper proteins (Fig. 2B). The basic motif is located in a constant register relative to the leucine repeats, and changes in spacing between the two domains, except for seven-amino-acid insertions, result in the loss of binding activity (2, 55). These results suggested that the leucine zipper and basic domains are structurally linked.

The basic region functions as the DNA contact surface in bZIP proteins. The replacement of certain residues in the basic domain eliminates DNA-binding activity but not the ability to dimerize (6, 10, 34, 42, 74), demonstrating that the leucine zipper and basic domains are functionally separable. Furthermore, domain-switching experiments prove that the basic region alone dictates DNA-binding specificity. Agre et al. (2) exchanged the leucine zipper and basic regions between C/EBP and GCN4 and found that the binding specificity of each chimeric protein correlated with the identity of the basic domain. Similar results were obtained from leucine zipper swap experiments involving the Jun, Fos, GCN4, and CREB proteins (31, 41, 57, 65).

Two models of the bZIP domain proposed detailed structures for the basic region. The scissors grip model (77) suggests that the basic region and the leucine zipper are part of a continuous α helix. The coiled-coil helices of the leucine zipper dimer diverge at the basic region, becoming recognition helices that lie within the major groove and track in opposite directions along the DNA. The helix is interrupted at a conserved asparagine residue in the basic motif that forms an N-cap structure (58) and divides the basic region into two helical domains. The two helices are separated by a bend that allows continued tracking of the basic region in the major groove. The scissors grip model explains two known properties of bZIP proteins: the restrictions on spacing between the basic and leucine zipper segments and the recognition of palindromic binding sites. A second paradigm, the induced helical fork (47), predicts a similar structure except that the basic region is a continuous helix. In addition, four basic region residues located on one face of the recognition helix are proposed to make base-specific DNA contacts. Many essential features of the scissors grip and helical fork models have now been confirmed by the X-ray crystallographic structure of GCN4 (13).

Several other studies provide support for the prediction

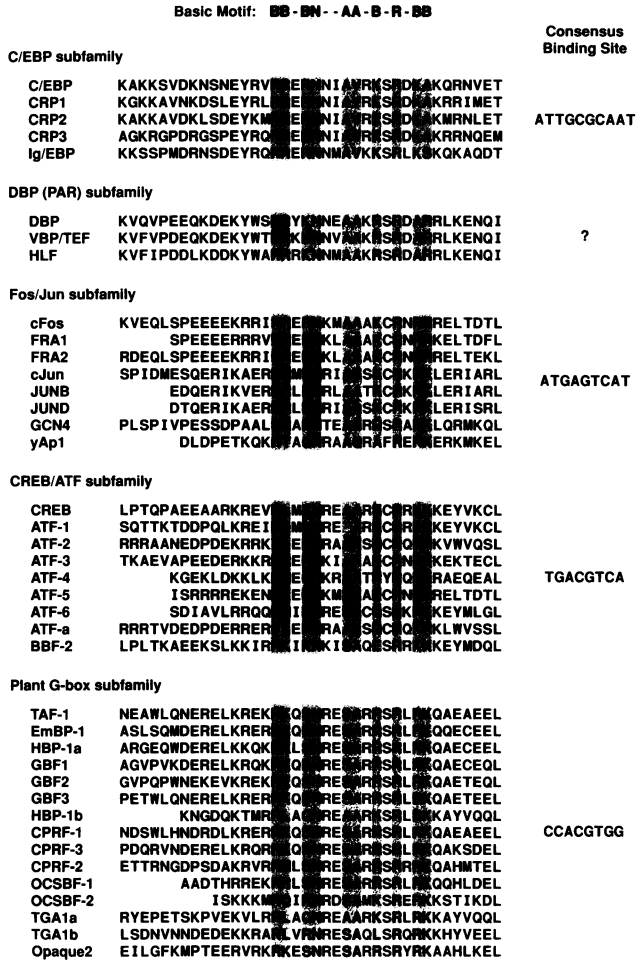


FIG. 1. Basic region amino acid sequence comparisons of proteins from several bZIP subfamilies. The sequences are aligned according to the basic motif homology and terminate at the first leucine residue within the leucine zipper. The proposed consensus recognition sequence for each subfamily is shown on the right. References for the amino acid sequences are as follows: C/EBP (32); CRP1 through CRP3 (80); Ig/EBP (59); DBP (40); VBP/TEF (26); HLF (25); c-Fos (75); FRA1 (9); FRA2 (43); c-Jun (5); JunB (62); JunD (61); GCN4 (21); yAp1 (39); CREB (24); ATF-1 through ATF-6 (18); ATF-a (14); BBF-2 (1); TAF-1 (45); EmBP-1 (17); HBP-1a (71, 72); GBF1 through GBF3 (64); HBP-1b (71); CPRF-1 through CPRF-3 (78); OCSBF-1 and OCSBF-2 (68); TGA1a and TGA1b (29); Opaque2 (19).

that the basic region is α helical. Circular dichroism measurements showed that the basic regions of GCN4 (47, 48, 73, 79) and Jun/Fos (53) contain high α-helical content when the proteins are bound to DNA. However, in the absence of the specific DNA ligand, significantly less α helicity was measured. A DNA-induced folding transition also was observed for the C/EBP basic region (48, 67). Together, these experiments indicate that the basic domain undergoes a conformational transition to a more α-helical state that is induced or stabilized by binding to DNA.

Although the α-helical nature of the DNA-binding surface was implied by numerous studies, the exact residues in this region that specify DNA sequence recognition have not been established by mutagenesis studies. The experiments described in this report were initiated to identify basic region residues that determine selective DNA recognition by the different subclasses of bZIP proteins. A series of hybrid proteins composed of basic region sequences from C/EBP and GCN4 was constructed, and the DNA-binding characteristics of these chimeras and of individual amino acid substitution mutants in GCN4 were determined. Three residues, which lie on one face of an α-helical DNA-binding surface, were found to affect the binding specificities of C/EBP and GCN4. A set of the GCN4-C/EBP basic region hybrids were also used in binding site selection experiments. These studies revealed a protein segment that specifies the optimal spacing between half-sites in the palindromic sequences that are recognized by bZIP proteins. The results are discussed with respect to the recently reported crystal structure of a GCN4-DNA complex (13).

MATERIALS AND METHODS

Construction and expression of mutant proteins. Fusions between the rat *c/ebp* gene (32) and the yeast *GCN4* gene (21) were constructed by the polymerase chain reaction (PCR) method described by Yon and Fried (81). The PCRs to create C-G fusion genes included dimethyl sulfoxide at a final concentration of 5% to promote denaturation of a guanine-cytidine-rich region in the *c/ebp* gene. Linker oligonucleotides specifying the fusion points contained 18 bases of homology to the appropriate parental gene on either side of the junction position. All of the junctions were located between adjacent codons. For G-C fusions, the exterior PCR primers were GCN4-5′ (5′-CCCACTCCTGTTCTAGAAGA TGC-3′) and C/EBP-3′ (5′-GACGGCAAGCTTGCCTCAC GCGCAGTTGCCCATGG-3′). For C-G fusions, the exterior primers were C/EBP-5′ (5′-ACGCCGCCTCTAGAACCCG TGCCCAGCCCTCAT-3′) and GCN4-3′ (5′-GACGGCAA GCTTAAATCAGCGTTCGCCAAC-3′).

Amino acid substitution mutations were introduced into GCN4 by the four-primer PCR mutagenesis procedure (20, 23), in which the mutagenic primers contained the relevant codon replacements from the *c/ebp* gene. GCN4-5′ and GCN4-3′ were used as exterior PCR primers for these reactions.

Products from the gene fusion and substitution mutagenesis PCRs were digested with *Xba*I and *Hin*dIII and inserted into the GCN4 expression plasmid pT5-GCN4 (2), which had been digested with the same two enzymes. This resulted in the replacement of carboxy-terminal GCN4 sequences between the *Xba*I site overlapping codons 167 to 169 and the *Hin*dIII site in the pT5 polylinker downstream from the GCN4 termination codon. All of the constructs were sequenced to ensure that no unintended mutations were introduced during PCR amplification. The plasmids were then introduced into the host strain *Escherichia coli* BL21 (DE3)pLysS (70) for protein overexpression. Expression was induced with isopropyl-β-D-thiogalactopyranoside (IPTG), and protein extracts were prepared and heat treated as previously described (34, 80). For the GCN4 substitution mutants, the following modified procedure was used to prepare protein extracts. Cultures (15 ml) were grown in Superbroth and induced with IPTG as described previously (34). The cell pellets were resuspended in 0.8 ml of 1 M KCl–50 mM Tris-HCl (pH 8.0)–1 mM EDTA containing 1 mM dithiothreitol (DTT) and 0.2 mM phenylmethanesulfonyl fluoride. After freezing on dry ice, the lysates were thawed and subjected to centrifugation at 30,000 rpm for 45 min in a Beckman tabletop ultracentrifuge, using a TLA100 rotor. The supernatants containing the soluble protein frac-

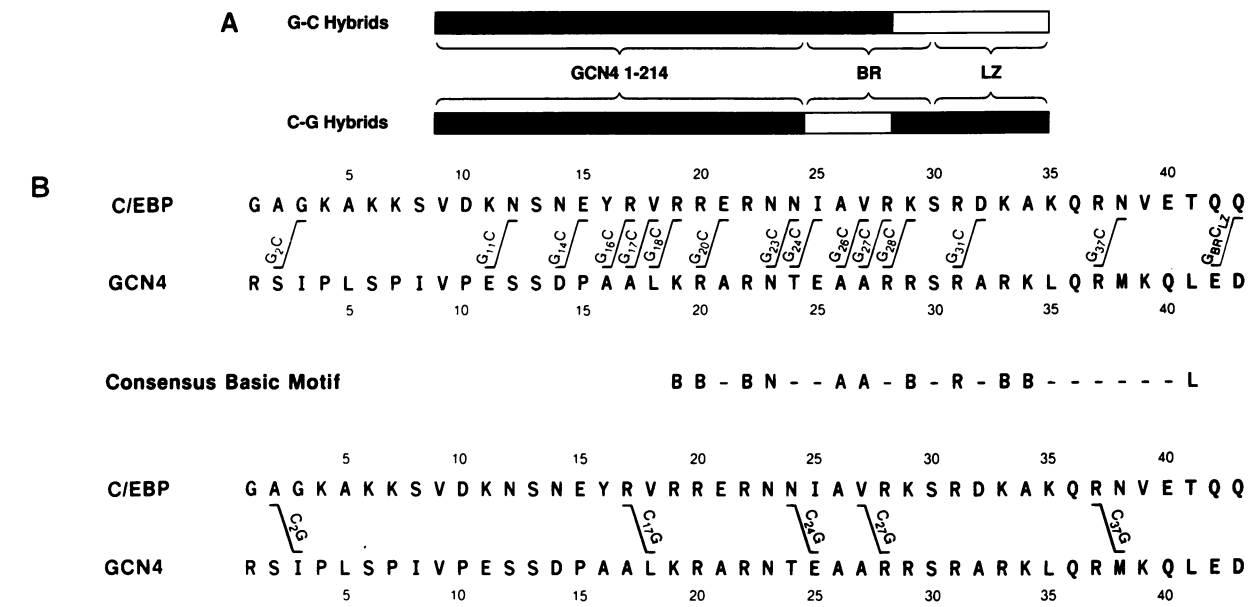


FIG. 2. (A) Gross structure of the GCN4-C/EBP hybrid proteins. Each protein contains the first 214 amino acids of GCN4 fused to a chimeric DNA-binding domain. The GCN4 amino-terminal region was included in all of the constructs because it promotes efficient expression of the recombinant proteins in *E. coli*. BR, basic region; LZ, leucine zipper; ■, GCN4 sequence; □, C/EBP sequence. (B) Junction positions within the fusion proteins. A universal amino acid numbering system for bZIP proteins was created to simplify the fusion protein nomenclature. Position 1 represents the approximate amino-terminal boundary of the extended basic region that is found in C/EBP and its relatives (80). The first conserved amino acid of the basic motif corresponds to position 19 in this system. Two series of hybrid proteins were constructed: G-C fusions (top line), which contain amino-terminal GCN4 sequences and switch to C/EBP at various points in the basic region, and G-C fusions (bottom line), which have the reciprocal architecture. $G_{BR}C_{LZ}$ is identical to the chimeric protein G_1C_1 described by Agre et al. (2). B denotes either arginine (R) or lysine (K) in the consensus basic motif.

tion were adjusted to 10% glycerol by adding 0.25 volume of 50% glycerol and then aliquoted and frozen at −70°C. Extracts prepared by this method were enriched for the expressed protein and were used without prior heat treatment.

DNA-binding assays. The C/EBP footprint probe derived from the rat albumin promoter has been described elsewhere (80). A GCN4 binding site probe from the yeast *HIS4* gene was prepared as follows. An *Xho*I fragment containing tandem GCN4 binding sites was released from the plasmid pHYC3-169 (22) and ligated into the *Sal*I site of the pEMBL18 cloning vector. The resultant plasmid (p18His4) was digested with *Bam*HI, treated with phosphatase, ^{32}P end labeled with T4 polynucleotide kinase, and digested with *Hin*dIII to release the *HIS4* fragment. ^{32}P-end-labeled fragments containing optimal C/EBP and GCN4 binding sites (see discussion of the binding site selection experiments below) were derived from clones G_2C-1 and $G_{BR}C_{LZ}$-1, respectively (see Fig. 6). These Bluescript-based (Stratagene) plasmids were digested with *Xho*I, dephosphorylated, labeled with T4 polynucleotide kinase, and digested with *Sac*I to release the DNA fragment containing the binding site.

DNase I footprint assays were performed essentially as described previously (27). Each binding reaction mixture contained 1× TM (25 mM Tris-HCl [pH 7.9], 6.25 mM $MgCl_2$, 10% glycerol, 0.5 mM EDTA, 0.5 mM DTT), 0.1% Nonidet P-40, 100 μg of bovine serum albumin (BSA) per ml, 0.2 μg of poly(dI-dC), 0.5 mM DTT, approximately 0.01 pmol of DNA probe, and protein extract as indicated. The binding reaction mixtures used for Table 1 (fusion protein assays) contained 75 mM KCl, and those used for Fig. 4 and 6 (substitution mutant assays) contained 125 mM KCl. The reaction mixtures were incubated for 15 to 20 min on ice and then subjected to partial DNase I digestion. After the addition of stop buffer containing 300 μg of proteinase K per ml, the samples were incubated at 55°C for 2 h and ethanol precipitated, and the cleavage products were separated by electrophoresis on 8% polyacrylamide sequencing gels.

Binding site selection experiments. Binding site selections were carried out by the procedure of Pollock and Treisman (54). The initial binding reaction mixtures (25 μl) contained 1× TM, 0.5 mM DTT, 0.1% Nonidet P-40, 150 mM KCl, 100 μg of BSA per ml, 0.1 μg of poly(dI-dC), approximately 0.1 pmol of the random sequence oligonucleotide (as described by Mavrothalassitis et al. [38]), approximately 25 ng of enriched (heat-treated) bacterial fusion protein, and 1 μl of anti-GCN4 antiserum. The GCN4-specific antiserum was raised in rabbits immunized with the synthetic amino-terminal GCN4 peptide, CysMetSerGluTyrGlnProSerLeuPheAlaLeuAsn (the amino-terminal Cys residue was included for covalent coupling to a carrier protein). The binding reaction mixtures were incubated for 30 min on ice and then added to 10 μl (packed volume) of protein A-Sepharose beads (Pharmacia). The samples were vortexed for 1.5 h at 4°C and centrifuged for 20 s in a microcentrifuge, and the supernatants were removed by aspiration. The beads were washed twice with ice-cold wash buffer (1× TM, 0.2% Nonidet P-40, 150 mM KCl), and the bound DNA was then eluted in 0.5 M ammonium acetate–5 mM EDTA–0.5% sodium dodecyl sulfate, extracted with phenol-chloroform, precipitated with ethanol, and dissolved in 10 μl of 10 mM Tris-HCl (pH 8.0)–1 mM EDTA (TE). A 3-μl aliquot of DNA was amplified by 20 cycles of PCR in 25-μl reaction mixtures containing 1× PCR buffer (Perkin-Elmer Cetus), 0.2 μg of each PCR primer (38), 50 μM each dATP, dGTP, and dTTP, 20 μM dCTP, 5 μCi of

[α-^{32}P]dCTP, 100 μg of BSA per ml, and 0.25 μl (1.25 U) of *Taq* DNA polymerase (Perkin-Elmer Cetus). The PCR products were separated on 10% polyacrylamide–Tris-borate-EDTA gels, and the 50-bp fragment was excised and eluted. The DNA samples were ethanol precipitated and dissolved in 30 μl of TE. A 2-μl aliquot was used for the next cycle of binding enrichment and amplification. After five such cycles, each DNA pool was digested with *Bam*HI and *Eco*RI and ligated into Bluescript which had been digested with the same two enzymes. Several clones from each of the three binding-enriched populations were selected for double-stranded DNA sequence analysis.

RESULTS

A chimeric protein strategy to identify DNA specificity determinants. The experimental approach was to identify basic region sequences that effect changes in DNA-binding specificity when exchanged between two bZIP proteins with different sequence recognition properties. It was expected that such gain-of-function mutants would yield more information about the roles of specific amino acid residues than would mutations that cause loss of, or reductions in, DNA-binding activity. The C/EBP and GCN4 proteins were used in these experiments because they possess distinct DNA-binding specificities:

C/EBP A T T G C $\mathbf{G_1C_2A_3A_4T_5}$

T A A C G C G T T A

GCN4 A T G A$\mathbf{G_1T_2C_3A_4T_5}$

T A C T C A G T A

The C/EBP site is a directly abutted 10-bp dyad (77), whereas GCN4 binds to a 9-bp imperfect palindrome in which the central G · C base pair is shared by both half-sites (46, 66). However, the C/EBP and GCN4 half-site sequences (boldface) are related, containing identical nucleotides at the first, fourth, and fifth positions. The GCN4 binding site is indistinguishable from the tetradecanoyl phorbol acetate-response element, or AP1, sequence recognized by the mammalian Jun/Fos proteins (3, 35, 69). C/EBP likewise belongs to a subfamily of bZIP proteins that exhibit identical or highly related DNA-binding properties (7, 59, 80) (Fig. 1).

Moving-boundary fusions between C/EBP and GCN4. The initial experiments were carried out with a series of hybrids between C/EBP and GCN4 that were fused at different locations in the basic region (Fig. 2A). It was anticipated that transitions in binding site preference between certain pairs of fusion proteins would occur, revealing protein segments or individual amino acids that function as DNA specificity determinants. The C/EBP and GCN4 DNA-binding domains were aligned by using the conserved basic motif and leucine zipper landmarks (Fig. 2B), and a series of fusion points were chosen. A universal amino acid numbering system was established for the bZIP basic region to simplify the fusion protein nomenclature (Fig. 2B); the conserved asparagine residue corresponds to position 23 in this system.

An oligonucleotide-directed PCR method (81) was used to create precise fusions between the *GCN4* and *c/ebp* genes. A series of hybrid proteins (G-C fusions; Fig. 2B) that contain amino-terminal GCN4 sequences and then switch to C/EBP sequences at various points within the basic region was generated. All of these proteins contain the leucine zipper and carboxyl terminus of C/EBP. Each hybrid was expressed in *E. coli*, and protein extracts were prepared. Heat treatment (70°C) was used to achieve a substantial purification of the expressed proteins. The proteins were adjusted to equivalent concentrations and tested for DNA-binding specificity in DNase I footprint assays. A segment of the rat albumin gene promoter containing a C/EBP binding site (DEI [8, 37]) and a fragment of the yeast *HIS4* promoter bearing two tandem GCN4 sites (22) were used as footprint probes. Under the binding conditions used, the parental proteins interacted only with their cognate sites (data not shown).

TABLE 1. DNA-binding specificities and relative activities of the G-C and C-G hybrid proteins

G-C recombinant	DNA-binding specificity[a]		C-G recombinant	DNA-binding specificity	
	GCN4	C/EBP		GCN4	C/EBP
2	−	+++	2	++	−
11	−	+++			
14	−	+++			
16	−	++			
17	−	++	17	+++	−/+
18	−	++			
20	−	++			
23	−?	+++			
24	+	++	24	+++	−/+
26	+	+			
27	++	−?	27	+	++
31	++	−?			
37	++	−	37	−	+++
$G_{BR}C_{LZ}$	+++	−			

[a] Average of several DNase I footprinting experiments. −?, possible weak binding.

The results of several independent binding experiments using the recombinant proteins were averaged and are summarized in Table 1. Every fusion protein exhibits at least a low level of sequence-specific DNA-binding activity. As the fusion point is moved through the basic region in the amino-to-carboxyl direction, a transition from C/EBP to GCN4 binding specificity is observed. Unexpectedly, two proteins that define the transition from C/EBP to GCN4 specificity ($G_{24}C$ and $G_{26}C$) interact weakly with both binding sites. Recombinants joined further to the carboxy-terminal side lose C/EBP specificity and gradually acquire a GCN4-like character, until their binding properties approach that of wild-type GCN4 when the fusion position reaches the basic region-leucine zipper boundary.

Five reciprocal (C-G) hybrid proteins were also constructed and assayed by DNase I footprinting. These proteins show a binding specificity pattern (Table 1) that is similar to that observed in the G-C series. Namely, a binding transition occurs at the point where the hybrids are joined downstream of the conserved asparagine residue, and this transition is defined by a hybrid protein ($C_{27}G$) that exhibits dual binding specificity. These similarities indicate that C/EBP and GCN4 use equivalent segments of the basic region to contact DNA.

In two cases, significant changes in binding specificity occurred among the G-C hybrids when the fusion boundary was shifted by one amino acid (the $G_{23}C$-$G_{24}C$ and $G_{26}C$-$G_{27}C$ pairs). Since the two proteins in each pair differ by only a single residue, these two amino acids are inferred to have major effects upon DNA-binding specificity. Thus, the acquisition of GCN4 binding between the hybrids $G_{23}C$ and

$G_{24}C$ suggests that residue 24 interacts with bases in the GCN4 recognition site, while the loss of C/EBP activity between $G_{26}C$ and $G_{27}C$ indicates that valine at position 27 plays a role in C/EBP binding. The importance of Val-27 is emphasized by the fact that in both the G-C and C-G series of recombinants, the ability to bind the C/EBP site is correlated with the presence of valine in this position (Table 1).

Amino acid substitution mutants. To define individual residues that affect DNA recognition, GCN4 mutants that contain C/EBP residues at their corresponding positions in the GCN4 basic region were constructed. Because residues 24 and 27 are implicated as specificity determinants, these two positions were mutated both individually (T-24→N and A-27→V) and in combination (24/27). The mutant proteins were tested in DNase I footprint assays (Fig. 3) in which the probes were optimal C/EBP and GCN4 binding sites obtained from the random selection procedure described below. All three GCN4 mutants protected the C/EBP binding site, albeit less efficiently than the C/EBP protein, whereas wild-type GCN4 did not form a detectable complex with the C/EBP probe. Each variant also retained binding activity for the GCN4 site, although 24/27 bound more weakly than either single substitution. These results demonstrate that positions 24 and 27 function in DNA sequence recognition. However, additional residues must be involved in distinguishing the DNA-binding properties of C/EBP and GCN4, because altering these two positions only partially eliminates binding to the GCN4 site.

Positions 24 and 27 occupy adjacent spokes of the presumed α helix formed by the basic region (spokes 2 and 5, respectively [Fig. 4]). On the assumption that any additional specificity determinants would reside in nearby helical locations, positions 16 and 34 were selected for substitution mutagenesis (mutants A-16→Y and K-34→A). These residues occur on spokes 1 and 5, respectively (Fig. 4). Figure 3 shows that exchanging the amino acid at position 34 has no detectable effect on the binding specificity of GCN4. The A-16→Y mutation may cause a very slight increase in binding to the C/EBP site. However, the effects of the A-16→Y substitution are much more apparent in combination with mutations at positions 24 and/or 27, as mutants 16/24, 16/27, and 16/24/27 bind the C/EBP site more avidly than do their antecedents that lack the residue 16 substitution. The contribution of position 16 is particularly evident when one compares the binding of 24/27 and 16/24/27 to the C/EBP site (Fig. 3B, lanes 6 and 10). Binding to this site is substantially enhanced in the triple mutant, whereas binding to the GCN4 site is moderately reduced (Fig. 3A, lanes 6 and 10). Thus, substitution of Tyr at position 16 significantly increases the ratio of C/EBP- to GCN4-specific binding of the 24/27 mutant. In contrast, the K-34→A exchange does not enhance the C/EBP-like character of proteins bearing the T-24→N and A-27→V substitutions (data not shown).

A final GCN4 mutant that carries substitutions at several basic region positions (variable residues) that are not part of the conserved basic motif was tested. This protein, termed VRM1.3, contains C/EBP amino acids at residues 17, 18, and 32. VRM1.3 was previously designed to examine whether a cluster of variable positions located on one face of the basic region α helix (Fig. 4) constitutes the DNA contact surface. Residues 18 and 32 lie on the side of the helix opposite of spokes 1, 2, and 5. VRM1.3 did not generate a detectable footprint on the C/EBP binding site (Fig. 3B, lane 12), and its interaction with the GCN4 site was equivalent to that of wild-type GCN4 (Fig. 3A, lane 12). Therefore, positions 17, 18, and 32 do not contribute to the binding specificity differences between C/EBP and GCN4.

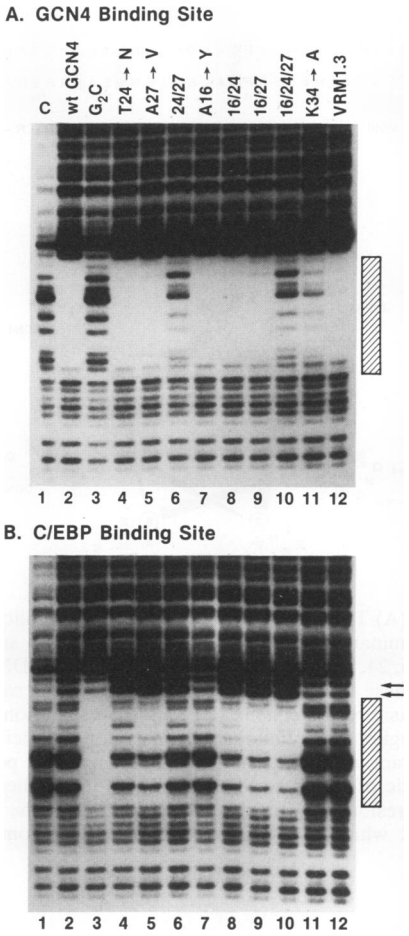


FIG. 3. DNase I footprint assays of GCN4 amino acid substitution mutants. (A) The randomly selected sequence $G_{BR}C_{LZ}$-1 (see Fig. 7), an optimal GCN4 binding site, was used as the footprint probe. The mutant protein used in each DNA-binding reaction is indicated above the lane. Extracts were adjusted so that the concentration of the expressed protein was approximately 100 ng/μl; 5 μl of each extract was used for footprint analysis. Lane C, control *E. coli* extract. wt, wild type. (B) The experiment is identical to that in panel A except that a randomly selected C/EBP binding site, G_2C-1 (see Fig. 6), was used as the probe. The hatched boxes show the extent of the protected region on each DNA fragment, and the arrows denote two DNase I-hypersensitive sites that are indicative of specific protein binding to the C/EBP site.

To compare the relative affinities of the GCN4 substitution mutants for the C/EBP binding site, footprint titration experiments were performed in which increasing amounts of protein were added to the binding reaction mixtures (Fig. 5). These data confirm that substitutions at residues 16, 24, and 27 confer C/EBP-like specificity to GCN4 in the relative order 27 > 24 > 16. Again, the A-16→Y mutation is found to accentuate the effects of substitutions at positions 24 and 27, and the K-34→A and VRM1.3 mutants are not observed to interact with the C/EBP site even at high protein concentrations.

Binding site selection experiments. While the above experiments identify amino acids that affect DNA-binding specificity, they do not indicate which bases in the binding site are

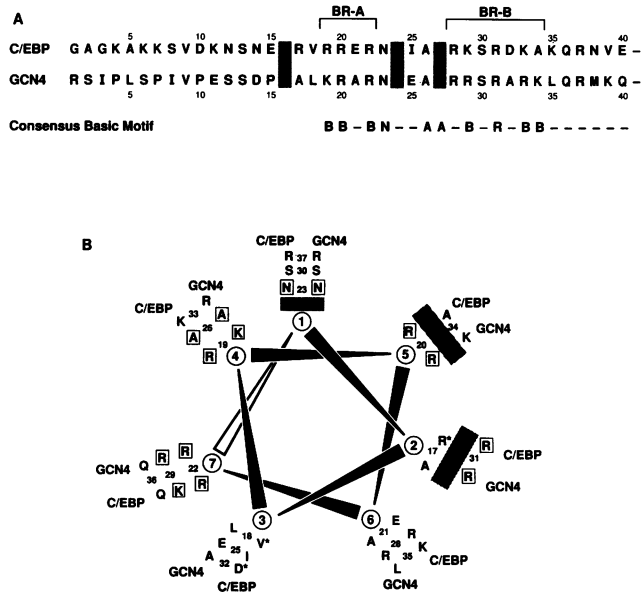


FIG. 4. (A) The three basic region residues implicated as specificity determinants define an α-helical face. The amino acids in positions 16, 24, and 27 were identified as probable DNA-contacting residues. These three positions (shaded) lie on the same face of an idealized basic region α helix (B); this representation assumes that the basic region is continuously helical. For simplicity, the helical wheel diagram uses the periodicity of 3.5 residues per turn that is characteristic of coiled-coil helices. The boxes indicate conserved basic motif residues. Asterisks signify positions in the GCN4 mutant VRM1.3 at which the corresponding residue from C/EBP was substituted.

contacted by these residues. By using DNA specificity mutants in experiments to select optimal DNA-binding sites, it should be possible to define the precise nucleotide contacts made by individual amino acids or basic region segments. This approach was applied initially to three C/EBP-GCN4 fusion proteins to begin mapping amino acid-nucleotide contacts within the bZIP domain.

A number of methods that permit the identification of optimal target sequences for DNA-binding proteins have been devised (4, 38, 54). These procedures involve repeated cycles of protein-binding enrichment from a pool of random DNA sequences and subsequent PCR amplification of the selected DNA. After several cycles, only the high-affinity binding sites for the protein remain, which are then cloned and sequenced. In this study, a selection method in which protein-DNA complexes are separated from free DNA by immunoprecipitation (54) was used. Because all of the recombinant proteins include the GCN4 amino-terminal region, an antiserum directed against the GCN4 amino terminus was used as the precipitating antibody.

Pilot immunoprecipitation experiments using the G_2C hybrid, which contains an intact C/EBP basic region, demonstrated that this protein bound a radiolabeled C/EBP site but not a GCN4 site (data not shown). Conversely, the protein $G_{BR}C_{LZ}$ (joined at the basic region-leucine zipper boundary) bound the GCN4 probe preferentially. Five cycles of binding selection and amplification were subsequently carried out with these two proteins and the $G_{27}C$ chimera that is fused 14 residues upstream of the leucine zipper boundary (Fig. 6). The starting DNA pool for the selection cycles was a 50-bp oligonucleotide containing an internal 10-bp segment of random-sequence DNA (38).

The sequences of several binding-selected oligonucleotides for each of the three hybrid proteins are presented in Fig. 6. The G_2C chimera bound sequences that include the

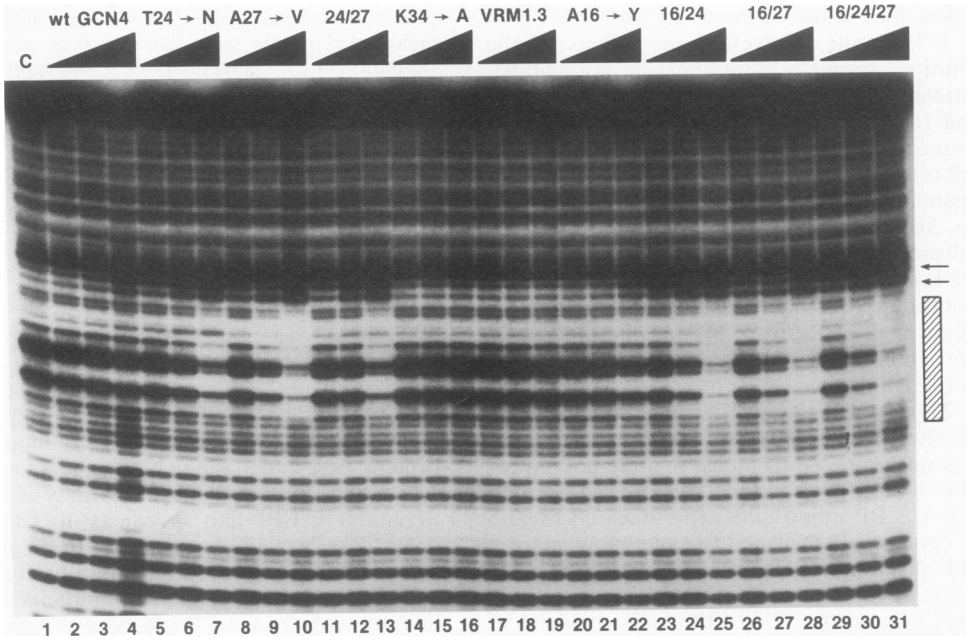
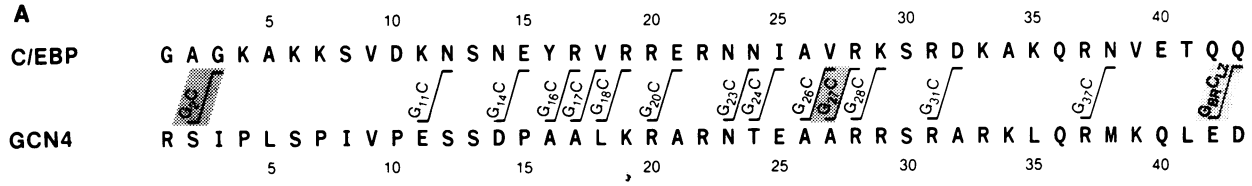


FIG. 5. Relative affinities of the GCN4 substitution mutants for a C/EBP binding site. The GCN4 mutants shown in Fig. 3 were assayed in footprint titration experiments. For each mutant, 0.75, 2.25, or 7.5 μl of the protein extract (100 ng/μl) was added to binding reaction mixtures containing the optimal C/EBP site (G_2C-1). Arrows identify the two DNase I-hypersensitive sites described in the legend to Fig. 3. Note their increased intensity in lane 22, indicating a weak interaction with the A-16→Y protein. Lane C, control *E. coli* extract. wt, wild type.



B

Fusion Protein		Selected Sequence	Half-site Sequence	Dyad Type
G_2C	1.	CCTGATTGTGCAATAGGC	C/EBP	A
	2.	CCTGCAATGATTACAGGC	C/EBP	I
	3.	CCTGCAATAGCACAAGGC	C/EBP	A
	4.	CCTGTGGTGTAATCAGGC	C/EBP*	A
	5.	CCTGACGCAATACTAGGC	GCN4:C/EBP	A
	6.	CCTGTGCAATAACCAGGC	C/EBP	A
	7.	CCTGACGCAATTCTAGGC	GCN4:C/EBP	A
	8.	CCTGGTGTTGTGTAAGGC	C/EBP	A
	9.	CCTGCAATTGCGTAAGGC	C/EBP	A
	10.	CCTGCAATCATTACAGGC	C/EBP	I
	11.	CCTGTCGCAATATTAGGC	C/EBP	A,I
	12.	CCTGTGGTTGCACAAGGC	C/EBP	A

Fusion Protein		Selected Sequence	Half-site Sequence	Dyad Type
$G_{BR}C_{LZ}$	1.	CCTGTGACTCATCCAGGC	GCN4	O
	2.	CCTGACTACAAACTAGGC	GCN4	O
	3.	CCTGTGAATCACCCAGGC	GCN4	O
	4.	CCTGTGATTCAACGAGGC	GCN4	O
	5.	CCTGAGTCAGTTGCAGGC	GCN4	O
	6.	CCTGTGACTAATCTAGGC	GCN4	O
	7.	CCTGTGACGACTTAAGGC	GCN4	A
	8.	CCTGCAGTCATAAAAGGC	GCN4	O
	9.	CCTGTGAGTCAACAAGGC	GCN4	O
	10.	CCTGGCAGTCATACAGGC	GCN4	O
	11.	CCTGGAGTCATATGAGGC	GCN4	O
	12.	CCTGTAGTCACTGTAGGC	GCN4	O

Fusion Protein		Selected Sequence	Half-site Sequence	Dyad Type
$G_{27}C$	1.	CCTGTGACGTCACAAGGC	GCN4	A
	2.	CCTGATGACGTAATAGGC	GCN4:C/EBP*	A
	3.	CCTGTGACATCATAAGGC	GCN4	A
	4.	CCTGTGACGTCATAAGGC	GCN4	A
	5.	CCTGTGACGTCACCAGGC	GCN4	A
	6.	CCTGATGACGTCACAGGC	GCN4	A
	7.	CCTGATGACGTAATAGGC	GCN4:C/EBP*	A
	8.	CCTGTGACGTCATAAGGC	GCN4	A
	9.	CCTGTGACGTCATTAGGC	GCN4	A
	10.	CCTGATGACGTCATAGGC	GCN4	A
	11.	CCTGTTACGTAATCAGGC	C/EBP*	A

FIG. 6. Sequences of several randomly selected binding sites for three GCN4-C/EBP fusion proteins. (A) The hybrid proteins used in the selection protocol. (B) Optimal sites were selected from a starting oligonucleotide pool that contained a random 10-bp internal segment. Sequences flanking the 10-bp random segment, which frequently contribute to the recognition site, are underlined. Palindrome half-sites and their orientations are indicated by the arrows. The half-site sequence and half-site spacing classifications of each sequence are shown in the two right-hand columns. C/EBP* signifies the presence of a GTAAT half-site in C/EBP-selected sequences, as opposed to the more frequently occurring GCAAT half-site. A, abutted; O, overlapping; I, inverted.

previously proposed C/EBP consensus half-site GCAAT (77) or the related sequence GTAAT. Consistent with many naturally occurring C/EBP binding sites (60), these sequences can diverge considerably from a perfect palindrome. However, an A is frequently present at nucleotide 4 in the more divergent half-site, and in most cases the palindromes contain directly abutted half-sites. Curiously, a few of the palindromes contain half-sites in the tail-to-tail rather than head-to-head orientation. This type of sequence also was occasionally selected by immunoprecipitation of DNA-protein complexes with a Fos-specific antibody (54). The appearance of these inverted dyads may reflect the formation of ring structures in the immunoprecipitation reaction consisting of two proteins, each of which interacts with one of the two half-sites and which are linked together by the bivalent antibody molecule. The inverted C/EBP dyads must be examined in direct binding assays to determine whether they function as true binding sites for C/EBP dimers.

$G_{BR}C_{LZ}$ selected sequences that represent good matches to the canonical GCN4 site—palindromes whose half-sites overlap at a central G · C base pair and match the GTCAT/C consensus. Similar sequences were previously identified in biochemical (38, 46) and genetic (66) selections for GCN4 binding sites. Because the $G_{BR}C_{LZ}$-selected sequences contain overlapping half-sites, the ability to recognize this type of palindrome cannot be a function of the leucine zipper domain (since $G_{BR}C_{LZ}$ contains the C/EBP leucine zipper and C/EBP preferentially binds to abutted dyads). Therefore, the protein segment that determines half-site spacing must be located upstream of the leucine zipper.

From an inspection of the binding sites selected by the $G_{27}C$ chimera, it was possible to further localize the sequences responsible for palindrome preference. Similar to the sites selected by $G_{BR}C_{LZ}$, the $G_{27}C$ sequences contain GCN4-like half-sites; however, their spacing is directly abutted. Thus, shifting the GCN4-C/EBP fusion point 14 amino acids upstream from the leucine zipper boundary alters the dyad preference but not the half-site sequence recognition properties of the protein. This specificity difference defines an element, located between residues 27 and 41 of the basic region, that controls palindrome spacing. While the minimal sequence that constitutes this spacing element has not yet been determined, the 14-amino-acid segment in which it resides includes the linker (77) or hinge region between the leucine zipper and basic region in which the two paired helices of the leucine zipper dimer are predicted to bifurcate. A possible explanation for the effects on half-site spacing by the hinge region is presented below.

DISCUSSION

Basic region residues affecting DNA-binding specificity. The analysis of C/EBP-GCN4 basic-region chimeras has identified three amino acids that influence DNA sequence recognition and thus may interact directly with binding site nucleotides (Fig. 7). The strategy of transferring DNA-binding specificity between C/EBP and GCN4 should reveal

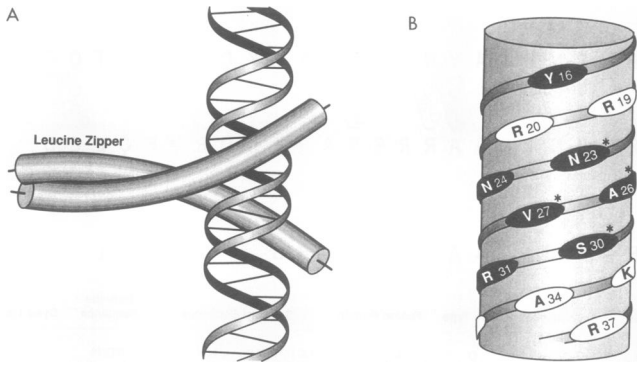


FIG. 7. (A) Generalized depiction of a bZIP dimer bound to DNA. The structure is based on the scissors grip (77) and helical fork (47) models and on the crystal structure of GCN4 (13). (B) Cylindrical projection of the basic region segment that contacts DNA. The amino acid sequence of C/EBP is shown; amino acids are numbered according to the system described in the legend to Fig. 2. Shaded symbols are positions that correspond to base-contacting residues in the GCN4 protein as determined by X-ray crystallography (13). Residues marked with asterisks were predicted by O'Neil et al. (47) to make base-specific contacts. Solid symbols represent positions identified in the present study as amino acids that distinguish C/EBP and GCN4 binding specificity.

residues that contribute to the differences in their binding properties. However, any residues that provide homologous functions in the two proteins, such as contacting the guanine nucleotide in the first position of both the C/EBP and GCN4 consensus half-sites, will be transparent to this kind of analysis. Therefore, the specificity residues identified here should not be interpreted as a complete map of DNA-binding determinants in the bZIP basic region. While it seems most likely that the three specificity residues contact DNA directly, the possibility that substitutions at these positions affect the folded structure of the basic region and thereby alter the DNA contact surface indirectly cannot be ruled out.

Position 27. Of the three specificity residues, Val-27 is clearly the most potent determinant of C/EBP binding character. In most bZIP proteins, position 27 is occupied by alanine (Fig. 1) and was therefore categorized as an invariant residue in the basic motif (30). However, because the basic motif consensus was derived primarily from proteins of the Jun/Fos/GCN4 subfamily, any of the basic motif amino acids could, in principle, contribute sequence-specific contacts. Moreover, of the known bZIP proteins, only those with C/EBP DNA-binding properties feature valine in this position (Fig. 1). Pu and Struhl (56) have reported that position 27 (Ala-239) in GCN4 tolerates substitutions by serine and valine without loss of in vivo activity and concluded that this residue does not make base-specific contacts. However, their genetic assay monitored loss of function, which, coupled with the observation that A-27→V retains affinity for GCN4 sites (Fig. 3A), may account for the absence of a phenotype for this mutation in vivo. It is also possible that Val-27 is more critical for the recognition properties of C/EBP than Ala-27 is for GCN4.

Position 24. The replacement of Thr-24 by Asn also significantly shifts GCN4 toward C/EBP binding specificity. In an earlier study, residue 24 was changed to arginine in an idealized GCN4 bZIP peptide that contained numerous other substitutions (47). This synthetic peptide bound to a GCN4 site, albeit with somewhat lower affinity than the wild type did, indicating that a basic amino acid at residue 24 is compatible with GCN4 specificity. Position 24 is occupied by basic amino acids in many proteins of the Jun/Fos/GCN4 subfamily (Fig. 1); hence, Thr24 cannot be a general requirement for GCN4-like binding. Nevertheless, the fact that Asn-24 is present in all five proteins of the C/EBP subfamily (Fig. 1) is consistent with its role as a C/EBP specificity determinant. In the future, substitution of Thr, Arg, or Lys at position 24 in C/EBP will test the importance of these amino acids for GCN4 specificity.

The double-replacement mutant 24/27 binds with lower affinity to C/EBP sites than either single mutant does and displays reduced affinity for the GCN4 binding site. While the latter result was predictable, the decreased binding to C/EBP sites was unexpected. This may reflect a general decrease in its DNA-binding activity, perhaps due to context effects that destabilize folding interactions or create incompatibilities between neighboring amino acid side chains.

Position 16. The effect of the A-16→Y mutation is most apparent from its augmentation of the T-24→N, A-27→V, and 24/27 mutant phenotypes. Two additional observations support a role for position 16 in DNA recognition. First, a comparison of the hybrid proteins C_2G and $C_{17}G$ reveals the partial acquisition of C/EBP specificity by $C_{17}G$ (Table 1). This transition may reflect DNA contacts mediated by residue 16. Second, by examining the binding properties of a series of synthetic GCN4 peptides that were progressively shortened at their amino termini, Oakley and Dervan (44) observed a decrease in binding affinity when the amino terminus was shifted from position 14 to position 20. This result is also consistent with the idea that residue 16 contributes to GCN4 DNA-protein interactions.

Proposed amino acid-nucleotide interactions. The specificity-determining positions 16 and 27 are occupied by hydrophobic amino acids (Ala or Val) in three of the four residues in C/EBP and GCN4. Although base contacts usually involve hydrogen bonds, hydrophobic interactions are also evident from crystallographic structures of protein-DNA complexes (reviewed in references 51 and 52). The use of Ala and Val in bZIP proteins to contact specific bases has precedents in bacterial DNA-binding proteins. For example, in λ repressor, the methyl group of Ala-49 makes van der Waals interactions with two thymines in the recognition sequence (36), and in the catabolite gene activator protein (CAP), a Glu-to-Val replacement alters sequence specificity (11, 12). Because Ala contacts thymine in λ repressor, it was tempting to speculate that Ala-27 interacts with the thymine (or T · A base pair) at position 2 of the GCN4 half-site, while Val-27 contacts the cytidine (or C · G base pair) in the analogous position of the C/EBP half-site. This contact has been confirmed by the crystal structure of a GCN4-DNA complex (see below). The more amino-terminal determinants (residues 24 and 16) would most likely interact with nucleotide 3, the other position that differs between C/EBP and GCN4 half-sites (see below).

The finding that certain GCN4-C/EBP chimeras exhibit dual binding specificities was somewhat unexpected. This class of proteins probably results from the relatedness of the C/EBP and GCN4 recognition sequences. Amino acids involved in contacting the conserved nucleotides at positions 1, 4, and 5 in the half-sites are likely to be functionally interchangeable between the two proteins, whereas residues contacting nucleotides 2 and 3 probably account for C/EBP-GCN4 discrimination. As long as a minimal number of base contacts is maintained in a chimeric protein (perhaps four of five nucleotides in a half-site), binding to both C/EBP and GCN4 sites is sufficiently avid to generate a DNase I

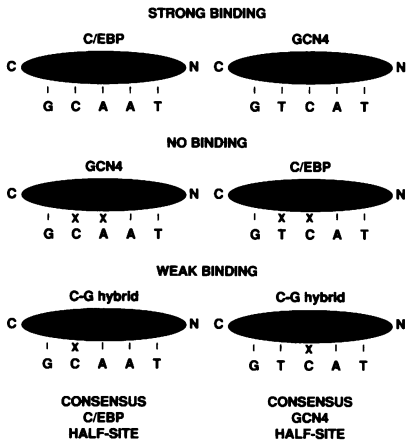


FIG. 8. Model for the dual binding specificity exhibited by certain GCN4-C/EBP recombinants. Contacts between the proteins and the DNA half-sites are depicted schematically. The bifunctional hybrids are proposed to make contacts with four of the five half-sites nucleotides in both the C/EBP and GCN4 recognition sequences, resulting in measurable binding to both sites. Because three nucleotides are common to both the C/EBP and GCN4 consensus half-sites, amino acids that contact these bases are assumed to be functionally interchangeable between the two proteins.

footprint (Fig. 8). Such a situation appears to occur in G-C hybrids whose junction points fall between residues 24 and 27 (Table 1). Positions 16 and 24 could combine to generate partial GCN4 specificity in these chimeras, while the presence of valine in position 27 allows recognition of C/EBP binding sites. These proposed base contacts also explain the dual recognition properties of the GCN4 amino acid substitution mutants (Fig. 3).

Comparison with the GCN4-DNA crystal structure. A 2.9-Å (0.29-nm) crystal structure of the GCN4 DNA-binding domain complexed with its binding site was recently determined (13). The physical structure obtained supports the predictions that the basic region and leucine zipper form a continuous α helix and that DNA recognition involves a helical basic region surface (47, 77). The two paired helices of the dimer are oriented orthagonally with respect to the DNA molecule, forming a T-shaped structure. The helices diverge gradually as the subunits enter the major grooves of each half-site in the recognition sequence. Interactions with DNA are provided by phosphate backbone contacts made predominantly by basic amino acid side chains and by base-specific contacts that involve a group of five amino acids. These five residues (depicted in Fig. 7) correspond to positions 23, 26, 27, 30, and 31 of the basic region. Arg-31, an absolutely conserved residue, contacts guanine at nucleotide 1 of the half-site, Ser-30 and Ala-26 interact with thymine in the A · T base pair occupying position 4, Ala-27 specifies thymine at nucleotide 2, and Asn-23 contacts both the cytidine in position 3 and the thymine of the A · T base pair at position 4.

Surprisingly, this quintet of DNA-contacting amino acids forms part of the basic motif that is shared by all members of the bZIP superfamily. Therefore, while the nucleotide contacts determined in the crystal structure can account for DNA recognition by GCN4, they fail to provide an explanation for the unique DNA-binding specificities of other bZIP proteins (13). The specificity residues identified in the present study help to resolve this paradox. Mutagenesis of GCN4 shows that replacement of Ala-27 by Val has a strong effect in promoting recognition of a C/EBP binding site (Fig. 3 and 5). This result conforms well with the observation that Ala-27 in GCN4 contacts nucleotide 2 (thymidine) and suggests that Val-27 interacts with the cytidine occupying the corresponding position in C/EBP sites. In fact, among the protein subfamilies compiled in Fig. 1, there is a perfect correlation between Ala-27 and thymidine at nucleotide 2 and between Val-27 and cytidine at this position. Consequently, rather than being a conserved residue, position 27 contains variations that serve to distinguish the binding specificities of different classes of bZIP proteins.

The mutationally defined contributions of amino acids 16 and 24 are more difficult to reconcile with the physical structure of GCN4. As proposed above, these two residues would most logically contact nucleotide 3, the other major sequence variation between the C/EBP and GCN4 half-sites. However, the GCN4 crystal structure shows the invariant asparagine 23 contacting cytidine in position 3 of the AP1 site, yet this base-specific interaction does not account for the fact that C/EBP and G-box proteins recognize sites containing A and G, respectively, at nucleotide 3 (Fig. 1). Therefore, contacts by amino acids other than (or in addition to) the conserved asparagine must account for the recognition of A or G. From analysis of the GCN4 substitution mutants, positions 16 and 24 are likely candidates for specifying A at nucleotide 3 in the C/EBP site. Reciprocal changes must now be introduced into the C/EBP protein to determine whether residues 16 and 24 also are required for GCN4 binding specificity. However, the GCN4-C/EBP chimeras demonstrate that at least partial GCN4 specificity is encoded by sequences that lie N terminal to residue 25 (hybrid $G_{24}C$; Table 1), even though this would not be predicted by the GCN4 crystal structure (13). Clearly, it will be important to analyze amino acid substitutions at positions 16, 24, and 27 in C/EBP (and other proteins) and to determine the structure of a protein from another bZIP subfamily to resolve the inconsistencies between the structural and genetic data.

A half-site spacing determinant in the hinge region. The C/EBP and GCN4 consensus binding sites differ by not only the sequences of their half-sites but also the spatial relationship between half-sites. The binding site selection experiments identified a protein segment that determines the preference for directly abutted half-sites by C/EBP and partially overlapping half-sites by GCN4. This element spans a 14-amino-acid sequence immediately preceding the leucine zipper that includes the hinge or linker segment (77). Preliminary sequence comparisons have not revealed a conserved motif corresponding to this half-site spacing element, although subtle similarities among bZIP proteins may become apparent once the minimal functional element has been defined.

The linker is located in the saddle of the Y-shaped bZIP dimer in which the helices bifurcate. How does the hinge region dictate half-site spacing? Presumably, either the angle of bifurcation or the exact position at which the paired helices diverge—properties determined by the amino acid sequence of the linker segment—might influence the distance between the paired basic domains. This distance would in turn determine the relative affinities for abutted and overlapping palindromes. bZIP dimers that preferentially bind to partially overlapping palindromes (i.e., the Jun/Fos/GCN4 subfamily) are predicted to contain more closely spaced basic regions than dimers that favor directly abutted dyads. One could propose that C/EBP prefers abutted dyads because it contains Thr instead of Leu at the first leucine

repeat position (residue 41), causing the helices to diverge further from the basic region and positioning the two basic regions farther apart. However, the protein sequences from several bZIP subfamilies (Fig. 1) show no correlation between leucine at residue 41 and preference for abutted dyads, indicating that the position of the first leucine in the zipper does not per se determine half-site spacing.

The spacing element imposes a preference, but not a rigid requirement, for a particular geometry of half-sites. Flexibility in the binding specificity of GCN4 was reported in a previous study, in which both overlapping 9-bp sequences and abutted 10-bp sequences were found to be efficient binding sites in vitro (66). However, the asymmetric 9-bp sequence was clearly the optimal GCN4 target in a genetic screen demanding transcriptional activation function in vivo (66). C/EBP exhibits a similar plasticity in its capacity to bind the sequences ATTGCGCAAT (abutted) and ATTG GCAAT (overlapping) in vitro (76), despite the fact that the binding enrichment experiments yielded only sequences with the abutted configuration (Fig. 6). These flexible dyad recognition properties probably account for the ability to create GCN4 mutants that bind to symmetrical C/EBP sites, even though they contain the GCN4 half-site spacing element.

It is worth noting that the sequences selected by the $G_{27}C$ fusion protein (directly abutted pairs of GCN4 half-sites) exactly match the consensus recognition sequence for the ATF/CREB subfamily of bZIP proteins. This observation suggests a pathway for a direct evolutionary link between these two regulatory protein subfamilies. A mutation in the linker region of a progenitor bZIP protein could have created a variant with novel sequence recognition properties, merely by changing the half-site spacing preference of the new variant. It has been proposed that a factor related to GCN4 was the ancestral bZIP protein, which subsequently gave rise to the Jun/Fos and ATF/CREB families in higher eukaryotes (16). Genetic evidence for an ATF/CREB-like protein in *Saccharomyces cerevisiae* (66) suggests that this branch point may have taken place at an early stage of evolution. Regardless of the chronology of this divergence, it seems plausible that the other known bZIP subfamilies, which all appear to recognize abutted dyads, evolved from an ATF/CREB-related ancestor as a consequence of basic region mutations that created new half-site sequence specificities.

ACKNOWLEDGMENTS

I thank Alan Hinnebusch for providing the *HIS4* promoter plasmid, John Burch and Simon Williams for valuable insights and discussion, Carrie Cantwell and Clara Choi for assistance with DNA sequencing, Marilyn Powers for oligonucleotides, Terry Copeland and Pat Wesdock for antisera, Hilda Marusiodis for preparing the manuscript, Anne Arthur for editing, and Marge Strobel, Barbara Graves, Simon Williams, Charles Vinson, John Burch, and Esta Sterneck for critical comments on the manuscript.

This research was sponsored by the National Cancer Institute under contract NO1-CO-74101 with ABL.

ADDENDUM

Suckow et al. (70a) recently reported similar amino acid substitution studies to identify specificity residues in C/EBP, GCN4, and TAF-1. These investigators also observe that positions 24 and 27 of the basic region are critical for the specificity differences between C/EBP and GCN4.

REFERENCES

1. **Abel, T., R. Bhatt, and T. Maniatis.** 1992. A *Drosophila* CREB/ATF transcriptional activator binds to both fat body- and liver-specific regulatory elements. Genes Dev. **6**:466–480.
2. **Agre, P., P. F. Johnson, and S. L. McKnight.** 1989. Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP. Science **246**:922–925.
3. **Angel, P., M. Imagawa, R. Chiu, B. Stein, R. J. Imbra, H. J. Rahmsdorf, C. Jonat, P. Herrlich, and M. Karin.** 1987. Phorbol ester-inducible genes contain a common *cis* element recognized by a TPA-modulated *trans*-acting factor. Cell **49**:729–739.
4. **Blackwell, T. K., and H. Weintraub.** 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science **250**:1104–1110.
5. **Bohmann, D., T. J. Bos, A. Admon, T. Nishimura, P. K. Vogt, and R. Tjian.** 1987. Human proto-oncogene c-*jun* encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. Science **238**:1386–1392.
6. **Bohmann, D., and R. Tjian.** 1989. Biochemical analysis of transcriptional activation by Jun: differential activity of c- and v-Jun. Cell **59**:709–717.
7. **Cao, Z., R. M. Umek, and S. L. McKnight.** 1991. Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells. Genes Dev. **5**:1538–1552.
8. **Cereghini, S., M. Ramondjean, A. Garcia Carranca, P. Herbomel, and M. Yaniv.** 1987. Factors involved in control of tissue-specific expression of albumin gene. Cell **50**:627–638.
9. **Cohen, D. R., and T. Curran.** 1988. *fra*1: a serum-inducible, cellular immediate-early gene that encodes a Fos-related region. Mol. Cell. Biol. **8**:2063–2069.
10. **Dwarki, V. J., M. Montminy, and I. M. Verma.** 1990. Both the basic region and the 'leucine zipper' domain of the cyclic AMP response element binding (CREB) protein are essential for transcriptional activation. EMBO J. **9**:225–232.
11. **Ebright, R. H., P. Cossart, B. Gicquel-Sanzey, and J. Beckwith.** 1984. Mutations that alter the DNA sequence specificity of the catabolite gene activator protein of *E. coli*. Nature (London) **311**:232–235.
12. **Ebright, R. H., P. Cossart, B. Gicquel-Sanzey, and J. Beckwith.** 1984. Molecular basis of DNA sequence recognition by the catabolite gene activator protein: detailed inferences from three mutations that alter DNA sequence specificity. Proc. Natl. Acad. Sci. USA **81**:7274–7278.
13. **Ellenberger, T. E., C. J. Brandl, K. Struhl, and S. C. Harrison.** 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.
14. **Gaire, M., B. Chatton, and C. Kedinger.** 1990. Isolation and characterization of two novel, closely related ATF cDNA clones from HeLa cells. Nucleic Acids Res. **18**:3467–3473.
15. **Gentz, R., F. J. Rauscher III, C. Abate, and T. Curran.** 1989. Parallel association of Fos and Jun leucine zippers juxtaposes DNA binding domains. Science **243**:1695–1699.
16. **Guarente, L., and O. Bermingham-McDonogh.** 1992. Conservation and evolution of transcriptional mechanisms in eukaryotes. Trends Genet. **8**:27–32.
17. **Guiltinan, M. J., W. R. Marcotte, Jr., and R. S. Quatrano.** 1990. A plant leucine zipper protein that recognizes an abscisic acid response element. Science **250**:267–271.
18. **Hai, T., F. Liu, W. J. Coukos, and M. R. Green.** 1989. Transcription factor ATF cDNA clones: an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. Genes Dev. **3**:2083–2090.
19. **Hartings, H., M. Maddaloni, N. Lazzaroni, N. Di Fonzo, M. Motto, F. Salamini, and R. Thompson.** 1989. The *O2* gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. EMBO J. **8**:2795–2801.
20. **Higuchi, R., B. Krummel, and R. Saiki.** 1988. A general method of *in vitro* preparation and specific mutagenesis of DNA frag-

ments: study of protein and DNA interactions. Nucleic Acids Res. **16**:7351–7367.
21. **Hinnebusch, A. G.** 1984. Evidence for translational regulation of the activator of general amino acid control in yeast. Proc. Natl. Acad. Sci. USA **81**:6442–6446.
22. **Hinnebusch, A. G., G. Lucchini, and G. R. Fink.** 1985. A synthetic HIS4 regulatory element confers general amino acid control on the cytochrome c gene (CYC1) of yeast. Proc. Natl. Acad. Sci. USA **82**:498–502.
23. **Ho, S. N., H. D. Hunt, R. M. Horton, J. K. Pullen, and L. R. Pease.** 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene **77**:51–59.
24. **Hoeffler, J. P., T. E. Meyer, Y. Yun, J. L. Jameson, and J. F. Habener.** 1988. Cyclic AMP-responsive DNA-binding protein: structure based on a cloned placental cDNA. Science **242**:1430–1433.
25. **Hunger, S. P., K. Ohyashiki, K. Toyama, and M. L. Cleary.** 1992. Hlf, a novel hepatic bZIP protein, shows altered DNA-binding properties following fusion to E2A in t(17;19) acute lymphoblastic leukemia. Genes Dev. **6**:1608–1620.
26. **Iyer, S. V., D. L. Davis, S. N. Seal, and J. B. E. Burch.** 1991. Chicken vitellogenin gene-binding protein, a leucine zipper transcription factor that binds to an important control element in the chicken vitellogenin II promoter, is related to rat DBP. Mol. Cell. Biol. **11**:4863–4875.
27. **Johnson, P. F., W. H. Landschulz, B. J. Graves, and S. L. McKnight.** 1987. Identification of a rat liver nuclear protein that binds to the enhancer core element of three animal viruses. Genes Dev. **1**:133–146.
28. **Johnson, P. F., and S. L. McKnight.** 1989. Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. **58**:799–839.
29. **Katagiri, F., E. Lam, and N.-H. Chua.** 1989. Two tobacco DNA-binding proteins with homology to the nuclear factor CREB. Nature (London) **340**:727–730.
30. **Kouzarides, T., and E. Ziff.** 1988. The role of the leucine zipper in the fos-jun interaction. Nature (London) **336**:646–651.
31. **Kouzarides, T., and E. Ziff.** 1989. Leucine zippers of *fos*, *jun* and GCN4 dictate dimerization specificity and thereby control DNA binding. Nature (London) **340**:568–571.
32. **Landschulz, W. H., P. F. Johnson, E. Y. Adashi, B. J. Graves, and S. L. McKnight.** 1988. Isolation of a recombinant copy of the gene encoding C/EBP. Genes Dev. **2**:786–800.
33. **Landschulz, W. H., P. F. Johnson, and S. L. McKnight.** 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240**:1759–1764.
34. **Landschulz, W. H., P. F. Johnson, and S. L. McKnight.** 1989. The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science **243**:1681–1688.
35. **Lee, W., A. Haslinger, M. Karin, and R. Tjian.** 1987. Two factors that bind and activate the human metallothionein II_A gene *in vitro* also recognize the SV40 promoter and enhancer regions. Nature (London) **325**:368–372.
36. **Lewis, M., A. Jeffrey, J. Wang, R. Ladner, M. Ptashne, and C. O. Pabo.** 1982. Structure of the operator-binding domain of bacteriophage λ repressor: implications for DNA recognition and gene regulation. Cold Spring Harbor Symp. Quant. Biol. **47**:435–440.
37. **Lichsteiner, S., J. Wuarin, and U. Schibler.** 1987. The interplay of DNA-binding proteins on the promoter of the mouse albumin gene. Cell **51**:963–973.
38. **Mavrothalassitis, G., G. Beal, and T. S. Papas.** 1990. Defining target sequences of DNA-binding proteins by random selection and PCR: determination of the GCN4 binding sequence repertoire. DNA Cell Biol. **9**:783–788.
39. **Moye-Rowley, W. S., K. D. Harshman, and C. S. Parker.** 1989. Yeast YAP1 encodes a novel form of the jun family of transcription activator proteins. Genes Dev. **3**:283–292.
40. **Mueller, C. R., P. Maire, and U. Schibler.** 1990. DBP, a liver-enriched transcriptional activator is expressed late in ontogeny and its tissue specificity is determined posttranscriptionally. Cell **61**:279–291.
41. **Nakabeppu, Y., and D. Nathans.** 1989. The basic region of Fos mediates specific DNA binding. EMBO J. **8**:3833–3841.
42. **Neuberg, M., M. Schuermann, J. B. Hunter, and R. Muller.** 1989. Two functionally different regions in Fos are required for the sequence-specific DNA interaction of the Fos/Jun protein complex. Nature (London) **338**:589–590.
43. **Nishina, H., H. Sato, T. Suzuki, M. Sato, and H. Iba.** 1990. Isolation and characterization of *fra2*, an additional member of the *fos* gene family. Proc. Natl. Acad. Sci. USA **87**:3619–3623.
44. **Oakley, M. G., and P. B. Dervan.** 1990. Structural motif of the GCN4 binding domain characterized by affinity cleaving. Science **248**:847–850.
45. **Oeda, K., J. Salinas, and N.-H. Chua.** 1991. A tobacco bZip transcription activator (TAF-1) binds to a G-box-like motif conserved in plant genes. EMBO J. **10**:1793–1802.
46. **Oliphant, A. R., C. J. Brandl, and K. Struhl.** 1989. Defining sequence specificity DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of the yeast GCN4 protein. Mol. Cell. Biol. **9**:2944–2949.
47. **O'Neil, K. T., R. H. Hoess, and W. F. DeGrado.** 1990. Design of DNA-binding peptides based on the leucine zipper motif. Science **249**:774–778.
48. **O'Neil, K. T., J. D. Shuman, C. Ampe, and W. F. DeGrado.** 1991. DNA-induced increase in the alpha-helical content of C/EBP and GCN4. Biochemistry **30**:9030–9034.
49. **O'Shea, E. K., J. D. Klemm, P. S. Kim, and T. Alber.** 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science **254**:539–544.
50. **O'Shea, E. K., R. Rutkowski, and P. S. Kim.** 1989. Evidence that the leucine zipper is a coiled coil. Science **243**:538–542.
51. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53**:293–321.
52. **Pabo, C. O., and R. T. Sauer.** 1992. Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. **61**:1053–1095.
53. **Patel, L., C. Abate, and T. Curran.** 1990. Altered protein conformation on DNA binding by Fos and Jun. Nature (London) **347**:572–575.
54. **Pollock, R., and R. Treisman.** 1990. A sensitive method for the determination of protein-DNA binding specificities. Nucleic Acids Res. **18**:6197–6204.
55. **Pu, W. T., and K. Struhl.** 1991. The leucine zipper symmetrically positions the adjacent basic regions for specific DNA binding. Proc. Natl. Acad. Sci. USA **88**:6901–6905.
56. **Pu, W. T., and K. Struhl.** 1991. Highly conserved residues in the bZIP domain of yeast GCN4 are not essential for DNA binding. Mol. Cell. Biol. **11**:4918–4926.
57. **Ransone, L. J., P. Wamsley, K. L. Morley, and I. M. Verma.** 1990. Domain swapping reveals the modular nature of Fos, Jun, and CREB proteins. Mol. Cell. Biol. **10**:4565–4573.
58. **Richardson, J. S., and D. C. Richardson.** 1988. Amino acid preferences for specific locations at ends of α helices. Science **240**:1648–1652.
59. **Roman, C., J. S. Platero, J. Shuman, and K. Calame.** 1990. Ig/EBP-1: a ubiquitously expressed immunoglobulin enhancer binding protein that is similar to C/EBP and heterodimerizes with C/EBP. Genes Dev. **4**:1404–1415.
60. **Ryden, T. A., and K. Beemon.** 1989. Avian retroviral long terminal repeats bind CCAAT/enhancer-binding protein. Mol. Cell. Biol. **9**:1155–1164.
61. **Ryder, K., A. Lanahan, E. Perez-Albuerne, and D. Nathans.** 1989. *Jun-D*: a third member of the *Jun* gene family. Proc. Natl. Acad. Sci. USA **86**:1500–1503.
62. **Ryder, K., L. F. Lau, and D. Nathans.** 1988. A gene activated by growth factors is related to the oncogene v-*jun*. Proc. Natl. Acad. Sci. USA **85**:1487–1491.
63. **Sassone-Corsi, P., L. J. Ransone, W. W. Lamph, and I. M. Verma.** 1988. Direct interaction between fos and jun nuclear oncoproteins: role of the 'leucine zipper' domain. Nature (London) **336**:692–695.
64. **Schindler, U., A. E. Menkens, H. Beckmann, J. R. Ecker, and A. R. Cashmore.** 1992. Heterodimerization between light-regulated and ubiquitously expressed *Arabidopsis* GBF bZIP proteins. EMBO J. **11**:1261–1273.
65. **Sellers, J. W., and K. Struhl.** 1989. Changing Fos oncoprotein to

a Jun-independent DNA-binding protein with GCN4 dimerization specificity by swapping 'leucine zippers.' Nature (London) **341**:74–76.

66. **Sellers, J. W., A. C. Vincent, and K. Struhl.** 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.
67. **Shuman, J. D., C. R. Vinson, and S. L. McKnight.** 1990. Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP. Science **269**:771–774.
68. **Singh, K., E. S. Dennis, J. G. Ellis, D. J. Llewellyn, J. G. Tokuhisa, J. A. Wahleithner, and W. J. Peacock.** 1990. OCSBF-1, a maize ocs enhancer binding factor: isolation and expression during development. Plant Cell **2**:891–903.
69. **Struhl, K.** 1987. The DNA-binding domains of the jun oncoprotein and the yeast GCN4 transcriptional activator are functionally homologous. Cell **50**:841–846.
70. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185**:60–89.

70a. **Suckow, M., B. von Wilcken-Bergmann, and B. Muller-Hill.** 1993. Identification of three residues in the basic regions of the bZIP proteins GCN4, C/EBP, and TAF-1 that are involved in specific DNA binding. EMBO J. **12**:1193–1200.

71. **Tabata, T., T. Nakayama, K. Mikami, and M. Iwabuchi.** 1991. HBP-1a and HBP-1b: leucine zipper-type transcription factors of wheat. EMBO J. **10**:1459–1467.
72. **Tabata, T., H. Takase, S. Takayama, K. Mikami, A. Nakatsuka, T. Kawata, T. Nakayama, and M. Iwabuchi.** 1989. A protein that binds to a cis-acting element of wheat histone genes has a leucine zipper motif. Science **245**:965–967.
73. **Talanian, R. V., C. J. McKnight, and P. S. Kim.** 1990. Sequence-specific DNA binding by a short peptide dimer. Science **249**:769–771.
74. **Turner, R., and R. Tjian.** 1989. Leucine repeats and an adjacent DNA binding domain mediate the formation of functional cFos-cJun heterodimers. Science **243**:1689–1694.
75. **Van Straaten, F., R. Muller, T. Curran, C. Van Beveren, and I. M. Verma.** 1983. Complete nucleotide sequence of a human c-*onc* gene: deduced amino acid sequence of the human c-*fos* gene protein. Proc. Natl. Acad. Sci. USA **80**:3183–3187.
76. **Vinson, C. R. (National Cancer Institute, Bethesda, Md.).** Personal communication.
77. **Vinson, C. R., P. B. Sigler, and S. L. McKnight.** 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science **246**:911–916.
78. **Weishaar, B., G. A. Armstrong, A. Block, O. da Costa e Silva, and K. Hahlbrock.** 1991. Light-inducible and constitutively expressed DNA-binding proteins recognizing a plant promoter element with functional relevance in light responsiveness. EMBO J. **10**:1777–1786.
79. **Weiss, M. A., T. Ellenberger, C. R. Wobbe, J. P. Lee, S. C. Harrison, and K. Struhl.** 1990. Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. Nature (London) **347**:514–515.
80. **Williams, S. C., C. A. Cantwell, and P. F. Johnson.** 1991. A family of C/EBP-related proteins capable of forming covalently linked leucine zipper dimers in vitro. Genes Dev. **5**:1553–1567.
81. **Yon, J., and M. Fried.** 1989. Precise gene fusion by PCR. Nucleic Acids Res. **17**:4895.