

Highly Conserved Residues in the bZIP Domain of Yeast GCN4 Are Not Essential for DNA Binding

WILLIAM T. PU AND KEVIN STRUHL*

*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School,
Boston, Massachusetts 02115*

Received 21 May 1991/Accepted 11 July 1991

Yeast GCN4 and the Jun oncoprotein are transcriptional activators that bind DNA via a bZIP domain consisting of a leucine zipper dimerization element and an adjacent basic region that directly contacts DNA. Two highly conserved alanines (Ala-238 and Ala-239 in GCN4) and an invariant asparagine (Asn-235) in the basic region have been proposed to play important roles in DNA sequence recognition by bZIP proteins. Surprisingly, these conserved residues can be functionally replaced in GCN4 and in a derivative containing the Jun basic region (Jun-GCN4). The ability of an amino acid to functionally substitute for Asn-235 does not correlate with its preference for assuming the N-cap position of an α helix. This finding argues against the proposal of the scissors grip model that the invariant asparagine forms an N cap that permits the basic region to bend sharply and wrap around the DNA. In contrast to a prediction of the induced fork model, the pattern of functional substitutions of the conserved alanines together with the results of uracil interference experiments suggests that Ala-238 and Ala-239 do not make base-specific DNA contacts. Finally, the Jun-GCN4 chimeric proteins appear much more active in vivo than expected from their DNA-binding properties in vitro. The mechanistic and evolutionary implications of these results are discussed.

GCN4 is a yeast protein that binds to the promoters of many amino acid biosynthetic genes and coordinately activates their transcription (2, 10). Optimal binding is observed with a 9-bp dyad, ATGA(C/G)TCAT, with the central 7 bp being most important (7, 18). GCN4 binds DNA via a bZIP structural motif shared by many eukaryotic transcription factors, including the Jun and Fos oncoproteins. This conserved bZIP motif is composed of two elements: a region rich in basic amino acids and a hydrophobic heptad repeat, the leucine zipper (33). bZIP proteins bind as dimers to dyad-symmetric DNA sequences (12). Dimerization is mediated by the leucine zipper (15, 21, 26), while the basic region directly contacts DNA (1, 30).

Alignment of the basic regions from 27 bZIP proteins reveals several residues that are highly conserved (Fig. 1). Most of these conserved residues are positively charged, with the degree of conservation and preference for lysine or arginine depending on the position. Interspersed between these basic amino acids are an invariant asparagine residue as well as two alanine residues that are found in 80% of the bZIP proteins currently known. Mutations at these conserved positions can eliminate sequence-specific DNA-binding activity (6, 14, 16, 31), whereas many of the nonconserved residues can be functionally replaced by alanine or glutamine (20).

Two models for DNA binding by bZIP proteins have been proposed (20, 33). Both models predict that the dimeric leucine zipper, a parallel coiled coil (21), symmetrically positions the diverging pair of α -helical basic regions so that they can make sequence-specific contacts with DNA. The basic regions are predicted to lie in the major groove of B-DNA, with the conserved lysine and arginine residues making electrostatic backbone contacts. These overall features of the protein-DNA complex are supported by biophysical studies which show that the GCN4 bZIP domain is

almost completely α helical when bound to high-affinity target sites (20, 34). Moreover, the spacing between the leucine zipper and basic region is critical for DNA binding, but can be altered by the insertion of an integral number of α -helical turns to both monomeric partners (22).

The models of the bZIP protein-DNA complex differ in the predicted role of the invariant asparagine residue (Asn-235 in GCN4) in the basic region. In the scissors grip model (33), the invariant asparagine is proposed to form an N-cap structure that breaks the basic region α helix. This would permit the protein to bend and continue to track along the major groove of DNA opposite that of entry, thus maximizing the protein surface available for interacting with DNA. In contrast, the induced fork model (20) proposes that the invariant asparagine is one of four highly conserved, neutral amino acids within the basic region that make base-specific contacts.

The induced helical fork model also predicts that two other conserved residues, Ala-238 and Ala-239 of GCN4, make base-specific contacts with DNA. Due to the nature of the alanine side chain, direct contacts with DNA must involve a hydrophobic group on a base, typically the 5-methyl group of thymine (13, 35). Such interactions can be studied by replacing critical thymine residues with uracil, which differs only by its lack of the 5-methyl group (5, 25, 29). Moreover, a direct interaction can be strongly inferred by missing contact probing (3); i.e., the effect of a particular uracil substitution should be reduced by mutation of the critical alanine(s) which interacts with the thymine methyl group.

To address how the conserved residues in the basic region contribute to the function of bZIP proteins, we used degenerate oligonucleotides to individually randomize the codons for Asn-235, Ala-238, and Ala-239 of GCN4. The residue corresponding to Asn-235 was also randomized in Jun-GCN4, a chimeric protein in which the basic region of Jun replaces the GCN4 basic region. Mutant proteins were assayed for GCN4 function in vivo and then tested for

* Corresponding author.

GCN4	D P A A L K R A R N T E A A R R S R A R K L Q R M K Q L
acr1	R K R K E F L E R N R V A A S K F R K R K K E Y I K K I
atf-1	L K R E I R L M K N R E A A R E C R R K K K E Y V K C L
atf-2	E K R R K V L E R N R A A A S R C R Q K R K V W V Q S L
atf-3	E R K K R R R E R N K I A A A K C R N K K K E K T E C L
atf-4	D K K L K K M E Q N K R A A T R Y R Q K K R A E Q E A L
atf-5	R E K E N P K E R N K M A A A K C R N R R R E L T D T L
atf-6	L R R Q Q R M I K N R E S A C Q S R K K K K E Y M L G L
c-ebp	N E Y R V R R E R N N I A V R K S R D K A K Q R N V E T
cpc-1	D V V A M K R A R N T L A A R K S R E R K A Q R L E E L
creb	R K R E V R L M K N R E A A R E C R R K K K E Y V K C L
crem	R K R E L R L M K N R E A A R E C R R K K K E Y V K C L
cys-3	A A E E D K R K R N T A A S A R F R I K K K Q R E Q A L
ebp1	L K R E R R K Q S N R E S A R R S R L R K Q Q E C E E L
fos-c	E K R R I R R E R N K M A A A K C R N R R R E L T D T L
fra-1	E R R R V R R E R N K L A A A K C R N R R K E L T D T L
hbp1	E L K A Q K R L S N R E S A R R S R L R K Q A E C E E L
jun-c	I K A E R K R M R N R I A A S K C R K R K L E R I A R L
junb	I K V E R K R L R N R L A A T K C R K R K L E R I A R L
jund	I K A E R K R L R N R I A A S K C R K R K L E R I S R L
lap	D E Y K M R R E R N N I A V R K S R D K A K M R N L E T
opaque2	E R V R K R K E S N R E S A R R S R Y R K A A H L K E L
pap1	E P S S K R K A Q N R A A Q R A F R K R K E D H L K A L
tga-1	E K V L R R L A Q N R E A A R K S R L R K K A Y V Q Q L
tgalb	E K K R A R L V R N R E S A Q L S R Q R K K H Y V E E L
treb5	E K A L R R K L K N R V A A Q T A R D R K K A R M S E L
yap1	E T K Q K R T A Q N R A A Q R A F R E R K E R K M K E L
<hr/>	
CONSERVED	<u>R</u> <u>R N R</u> <u>A A</u> <u>S R</u> <u>R K</u> <u>L</u>
	k k k C k r

FIG. 1. Conserved residues in the basic regions of bZIP proteins. Amino acid sequences of the basic regions of 27 bZIP proteins are listed, with conserved residues indicated in boldface. Shown below is the consensus; capital letters indicate the predominant basic residue, lowercase letters indicate minor basic residues, and invariant residues are underlined. The conserved leucine at the end of the displayed sequences is the first leucine of the leucine zipper.

DNA-binding activity in vitro. Using a novel uracil interference assay (details to be published elsewhere), we also examined the importance of the 5-methyl group of thymine for DNA binding by GCN4 derivatives mutated at the conserved alanine residues. The results indicate that these highly conserved residues are not essential for GCN4 function. Moreover, they suggest that Asn-235 does not function by forming an N-cap structure and that Ala-238 and Ala-239 are not involved in standard base-specific contacts.

MATERIALS AND METHODS

Construction of mutants. The DNA molecules used in this study were derived from YCp88-Sc4400, a *ura3* centromeric vector that permits expression of GCN4 in yeast cells from the *ded1* promoter and in vitro from the bacteriophage SP6 promoter. YCp88-Sc4400 was constructed from YCp88-GCN4 (11) as follows (Fig. 2A): the *XmnI-EcoRI* fragment was replaced with the corresponding fragment from pTZ18 containing the *f1* origin; a *BglIII* site was generated at position 1227 by oligonucleotide-directed mutagenesis; and the region encoding the native GCN4 DNA-binding domain was replaced by a synthetic version generated by Chris Brandt that contains new restriction sites and preferred *Escherichia coli* codons without changing the encoded protein. Although the resulting molecule lacks the 3' untranslated region of GCN4, there is no detectable difference in the ability of YCp88-Sc4400 and YCp88-GCN4 to complement a *gcn4* strain. To generate libraries in which Asn-235, Ala-238, and Ala-239 were randomized individually, degenerate oligonucleotides (Fig. 2B) were converted to double-stranded DNA by mutually primed synthesis (19), digested with the indicated restriction enzymes, and inserted at the matching restriction sites in the GCN4 coding region. The quality of

each library was assessed by sequencing 6 to 12 recombinant DNA molecules.

Phenotypic analysis. Libraries encoding the mutant proteins were introduced into KY1365 (relevant genotype *ura3-52 gcn4-Δ1 his3-Δ86,189*), a derivative of KY803 (11) constructed by Dimitris Tzamaras. The *his3* promoter region contains both *his3-189*, an allele which generates the optimal GCN4 binding site, and *Δ86*, which replaces sequences between -103 and -447 by an *EcoRI* linker, thus removing all promoter elements upstream of the GCN4 binding site (7). *Ura*⁺ transformants were selected, and the resulting strains were assayed for GCN4 function by growth in 5 or 40 mM aminotriazole (AT) (11). For each library, plasmid DNAs were recovered from approximately 10 individual yeast strains by transformation of *E. coli*, and the sequence of the basic region was determined. These DNAs, as well as those derived by sequencing unselected molecules from the libraries, were reintroduced into KY1365 to confirm their growth phenotypes. The resulting strains were grown to saturation in glucose minimal medium containing Casamino Acids (Difco) and lacking uracil. A 10-μl amount of a 1:100 dilution (approximately 1,000 cells) of each culture was spotted on plates containing minimal medium supplemented with tryptophan, leucine, adenine, lysine, and either 5 or 40 mM AT.

DNA-binding experiments. ³⁵S-labelled GCN4 proteins were synthesized in wheat germ extracts programmed with mRNA obtained by transcribing the relevant DNA templates with SP6 RNA polymerase (10). Roughly equivalent amounts of the various ³⁵S-proteins (estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis) were incubated with 8 nM ³²P-labelled 435-bp *EcoRI-HindIII* DNA fragment containing the optimal binding site (27). The resulting protein-DNA complexes were separated by acryl-

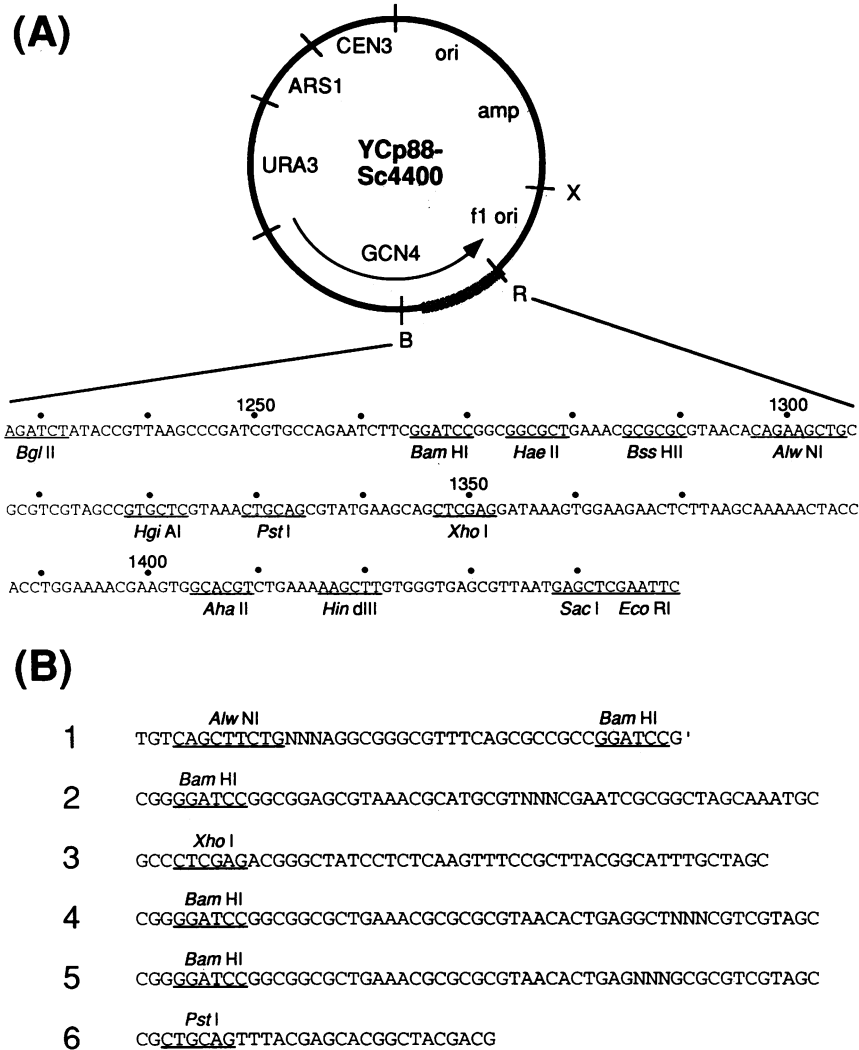


FIG. 2. DNA molecules. (A) Structure of YCp88-Sc4400 (see Materials and Methods for construction) indicating the locations of important genetic elements on the vector, GCN4 coding region (arrow indicates direction of transcription; shaded bar indicates DNA-binding domain), selected restriction sites (G, *Bgl*II; R, *Eco*RI; X, *Xmn*I), and the DNA sequence of the synthetic GCN4 DNA-binding domain, with novel restriction sites underlined and GCN4 sequences numbered as described previously (9). (B) Sequences of degenerate oligonucleotides containing individually randomized codons (N indicates positions programmed to be 25% each nucleotide) in the DNA-binding domain of GCN4 that were used to generate protein libraries as follows: 1, Asn-235 substitutions of GCN4; 2 and 3, Asn-235 substitutions of Jun-GCN4; 4 and 6, Ala-239 substitutions of GCN4; 5 and 6, Ala-238 substitutions of GCN4.

amide gel electrophoresis. Given the errors involved in estimating the amounts of GCN4 derivatives and problems due to varying quantities of bulk protein from the wheat germ extract (necessary to normalize the amount of ³⁵S-protein added for each determination), twofold differences in protein-DNA complex formation are probably not significant.

Uracil interference assay. Randomly deoxyuracil-substituted, double-stranded DNA substrates were synthesized by polymerase chain reaction (PCR). A 200-fmol amount of an oligonucleotide containing the optimal GCN4 binding site (GGGCTTTCTGCTCTGTCATCGAATTCGGATGACTCATTTTTGGATCCATGGTCATAGCTGTT), 20 pmol of each of two primers (GGGCTTTCTGCTCTGTCATC and AACAGCTATGACCATGG), one of which was 5' labelled with ³²P, 5 U of *Taq* DNA polymerase, 200 μM each of the four deoxynucleoside triphosphates, and 50 μM dUTP were

incubated in 100 μl of PCR buffer (50 mM KCl, 3.5 mM MgCl₂, 10 mM Tris-HCl [pH 8.3], 100 μg of gelatin per ml) for eight cycles; the reaction product was purified on an 8% native polyacrylamide gel. In vitro-synthesized proteins were incubated with the purified product, and DNA-protein complexes were isolated by electrophoresis through a polyacrylamide gel. DNA recovered from the complexes was digested for 1 h with 1 U of uracil-*N*-glycosylase (Perkin Elmer-Cetus) at 37°C in 60 μl of PCR buffer, ethanol precipitated, and treated with 1 M piperidine at 90°C for 30 min. The sample was lyophilized, resuspended in formamide, and run onto a 12% denaturing polyacrylamide gel.

RESULTS

Randomization of conserved residues in the basic region of GCN4. The basic regions of bZIP proteins contain several

TABLE 1. Quality of mutant libraries^a

Basic region	Residue	Size	G	A	T	C
GCN4	Asn-235	3,000	24	43	14	19
Jun-GCN4	Asn-235	100	19	5	62	14
GCN4	Ala-238	2,000	35	5	41	17
GCN4	Ala-239	2,000	4	4	66	25

^a Libraries, named by the basic region (GCN4 or Jun) and the randomized codon, are characterized by their size (total number of *E. coli* transformants) and the percentage of each base in the coding strand of the randomized codon (determined by sequencing DNA from 6 to 12 unselected *E. coli* transformants).

highly conserved residues (Fig. 1). The functional role of these residues is largely unknown, although models of the bZIP protein-DNA complex predict that several of these conserved residues make specific interactions that are essential for high-affinity DNA binding. To analyze the role of these residues, we individually randomized the codons for three of the highly conserved residues of GCN4 (Asn-235, Ala-238, and Ala-239). Asn-235 was also randomized in Jun-GCN4, a chimera in which the Jun basic region replaced the GCN4 basic region. The quality of the resultant mutant libraries, determined by sequencing unselected DNAs, is summarized in Table 1. Although the library of GCN4 Asn-235 derivatives certainly contains all possible substitutions, the other libraries probably do not, as a result of skewed nucleotide distributions in the degenerate oligonucleotides. However, given the sizes of these biased libraries, it is very likely that most amino acid changes are represented.

Analysis of Asn-235 substitutions in GCN4. When the GCN4 position 235 library was introduced into a yeast *gcn4* deletion strain (KY1365), roughly 10% of the Ura⁺ transfor-

mants displayed GCN4 function. Plasmids were recovered from strains both with and without GCN4 activity, and the region encoding the GCN4 DNA-binding domain was sequenced. The plasmids were reintroduced into KY1365, and GCN4 activity was determined by the ability of strains harboring the plasmid to grow on AT (Fig. 3). Although the majority of amino acid substitutions of Asn-235 eliminate GCN4 function, we were surprised to find that derivatives containing tryptophan, glutamine, or alanine display GCN4 activity. The protein containing the nonconservative tryptophan substitution displays nearly wild-type function (only slightly reduced growth in 40 mM AT), whereas the more conservative glutamine and alanine substitutions show very weak activity (barely detectable growth in 5 mM AT).

To directly analyze DNA-binding activities, we synthesized several mutant proteins in vitro and measured their affinity for a DNA fragment containing the optimal GCN4 binding site (Fig. 4A). The Trp-235 protein binds efficiently but somewhat less avidly than wild-type GCN4 (Asn-235), the Ala-235 and Gln-235 proteins bind less efficiently, and the Gly-235 protein binds extremely weakly. These DNA-binding activities correspond to the in vivo activities of these proteins. The lack of GCN4 function in strains expressing Gly-235 protein, despite the ability of this protein to bind DNA, agrees with our experience that the threshold for DNA-binding activity needed for detectable function in vivo is higher than that needed for detectable DNA binding in vitro. The Lys-235 and Glu-235 proteins do not bind the fragment as a homodimer. However, the Glu-235 protein binds detectably as a heterodimer with GCN4-C131, a truncated wild-type protein that retains the DNA-binding domain (11), whereas the Lys-235 protein does not. This result suggests that the glutamic acid substitution confers a less

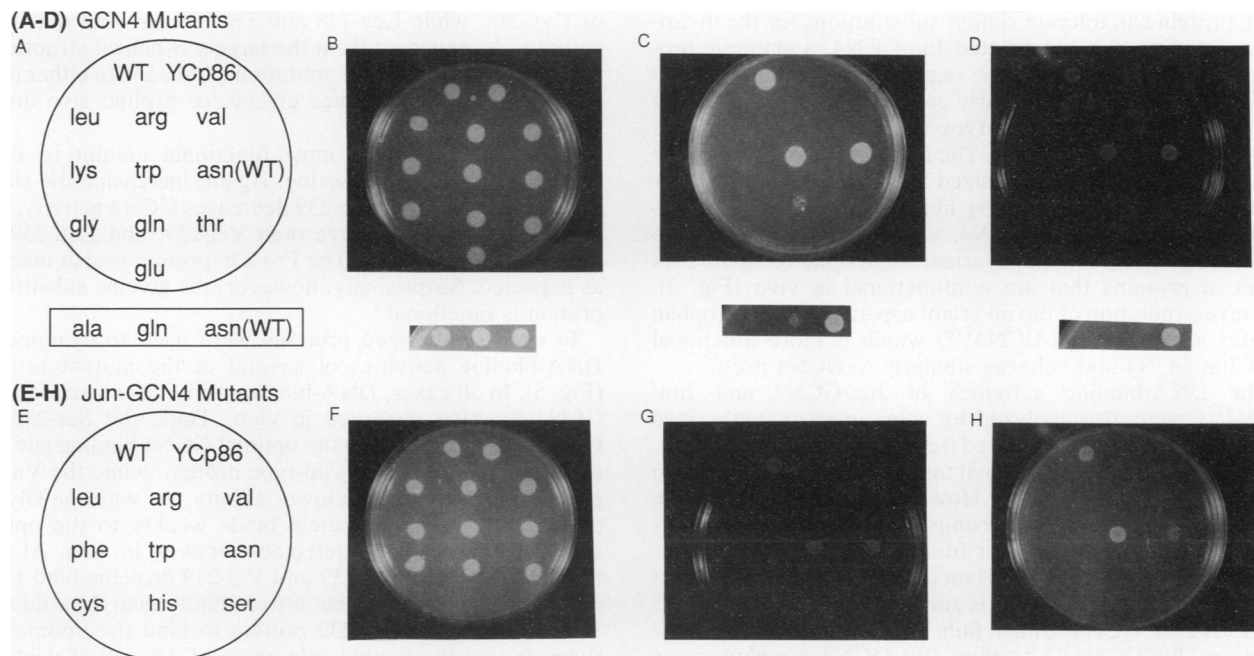


FIG. 3. In vivo GCN4 activity of derivatives mutated at position 235. Approximately 1,000 cells containing the indicated GCN4 or Jun-GCN4 derivative (A and E) were spotted on plates containing Casamino Acids and lacking uracil (B and F) or plates containing minimal medium with appropriate supplements and either 5 mM (C and G) or 40 mM (D and H) AT. YCp86 is a control plasmid that lacks GCN4 coding sequences. WT, wild type.

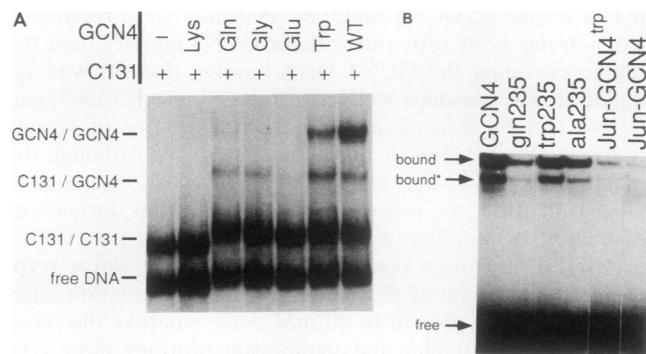


FIG. 4. DNA-binding activities of derivatives mutated at position 235. (A) In vitro synthesized GCN4 (wild type [WT]) or derivatives with the indicated amino acid residue at position 235 were combined with in vitro-synthesized GCN4-C131, a truncated GCN4 derivative which contains the entire DNA-binding domain (11), and incubated with a 32 P-labelled DNA fragment containing the optimal GCN4 binding site. Protein-DNA complexes involving GCN4 homodimers (GCN4/GCN4), GCN4-C131 heterodimers (C131-GCN4), and C131 homodimers (C131-C131) are indicated. The intensities of the bands representing the heterodimer complex were lower than expected because the two proteins were mixed after synthesis (12). (B) Similar analysis that includes proteins with the indicated residue at position 235 of Jun-GCN4. The complex indicated as bound* represents a heterodimer between the full-length GCN4 derivative and an artifactual translation product containing the 175 C-terminal residues that is generated by aberrant initiation at an internal AUG codon. The presence and amount of the bound* complex depends on the batch of wheat germ extract and the age of the synthesized proteins.

severe defect in DNA binding than does the lysine substitution.

Analysis of Asn-235 substitutions in the Jun-GCN4 hybrid protein. To determine whether the basic region of another bZIP protein can tolerate similar substitutions for the invariant asparagine, we constructed Jun-GCN4, a chimeric protein in which the GCN4 basic region is replaced by the Jun basic region. Jun-GCN4 clearly confers GCN4 function, but it is less effective than wild-type GCN4 (grows much more slowly in 40 mM AT; Fig. 3). The invariant asparagine of the chimeric protein was randomized by using degenerate oligonucleotides, and the resulting library was analyzed as described above. As with GCN4, we found that most amino acid substitutions for the invariant asparagine of Jun-GCN4 result in proteins that are nonfunctional in vivo (Fig. 3). However, mutation of the invariant asparagine to tryptophan creates a protein (Jun-GCN4^{Trp}) which is more functional than Jun-GCN4 and behaves similarly to GCN4 itself.

The DNA-binding activities of Jun-GCN4 and Jun-GCN4^{Trp} were also analyzed by using in vitro-synthesized proteins (Fig. 4B). As expected from their properties in vivo, Jun-GCN4^{Trp} binds the optimal target site more efficiently in vitro than does Jun-GCN4. However, in comparison with GCN4 derivatives that have comparable functional activities in vivo, both Jun-GCN4 and Jun-GCN4^{Trp} bind DNA surprisingly poorly. Jun-GCN4 is moderately functional in yeast cells, but its binding affinity is similar to that of the Gly-235 derivative of GCN4 which fails to complement the *gcn4* deletion. Jun-GCN4^{Trp} confers full GCN4-complementing activity, while the Gln-235 derivative displays barely detectable function, even though the two proteins have comparable DNA-binding activities in vitro. Since both Jun-GCN4 chimeric proteins are much more effective in vivo than

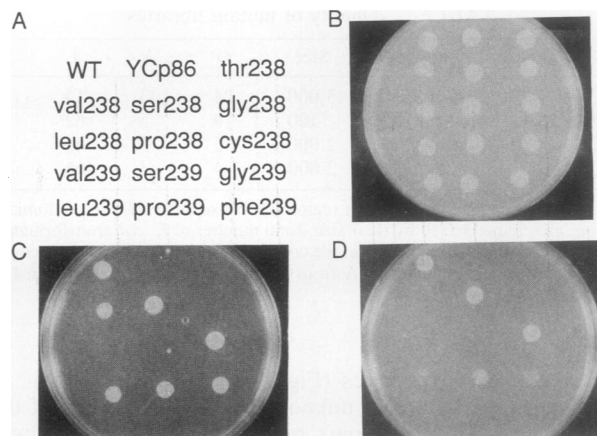


FIG. 5. In vivo GCN4 activity of derivatives mutated at positions 238 or 239. Approximately 1,000 cells containing the indicated GCN4 mutant proteins (A) were spotted on plates containing Casamino Acids and lacking uracil (B) or plates containing minimal medium with appropriate supplements and either 5 mM (C) or 40 mM (D) AT. WT, wild type.

expected from their DNA-binding activities, the phenomenon is probably due to some property of the Jun basic region (see Discussion for some potential explanations).

Analysis of Ala-238 and Ala-239 substitutions in GCN4. When libraries of GCN4 proteins randomized at either position 238 or 239 of GCN4 were introduced into KY1365 cells, approximately 50% of the plasmids conferred some degree of GCN4 function. At position 238, the most functional mutants, Ser-238 and Cys-238, have residues with small, nonpolar side chains at position 238 (Fig. 5). Increasing the size of this side chain decreases GCN4 activity: Val-238 is significantly less functional than Ala-238, Ser-238, or Cys-238, while Leu-238 and Thr-238 have no detectable activity. As expected from the largely α -helical structure of the basic domain (20, 34), mutation of Ala-238 to either of the helix-destabilizing residues glycine or proline also disrupt function.

At position 239, the most functional mutant is also a change from alanine to serine. Again, increasing the size of the side chain at position 239 decreases GCN4 activity, with Ser-239 being more active than Val-239, and Leu-239 and Phe-239 being inactive. The Pro-239 protein is also inactive, as expected. Surprisingly, however, the glycine substitution protein is functional.

In vitro-synthesized proteins were used to examine the DNA-binding activities of several of the mutant proteins (Fig. 6). In all cases, DNA-binding activities correlated with GCN4 function observed in vivo. Thus, the Ser-238 and Cys-238 proteins bind to the optimal GCN4 binding site with the same affinity as the wild-type protein, while the Val-238 protein binds with much lower affinity. As with the Gly-235 protein, the Leu-238 protein binds weakly to the optimal site, despite having no detectable activity in vivo. At position 239, both the Gly-239 and Val-239 proteins bind to the optimal site with somewhat lower affinity than the wild type. The ability of the Gly-239 protein to bind the optimal site suggests that the methyl side chain of Ala-239 of wild-type GCN4 does not significantly contribute to binding affinity.

Uracil interference pattern of Ala-238 and Ala-239 mutants. It has been suggested that Ala-238 and Ala-239 are involved in sequence-specific DNA binding because they are highly

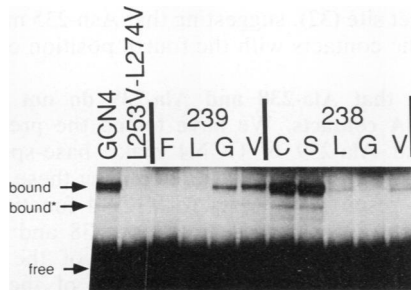


FIG. 6. DNA-binding activities of GCN4 derivatives mutated at position 238 or 239. In vitro-synthesized proteins with the indicated amino acid substitution were incubated with the optimal GCN4 binding site, and protein-DNA complexes were separated from unbound DNA. The complex indicated as bound* represents a heterodimer between the full-length GCN4 derivative and an artifactual translation product containing the 175 C-terminal residues that is generated by aberrant initiation at an internal AUG codon. The L253V-L274V derivative of GCN4 has valines in place of two conserved leucines of the leucine zipper and does not bind DNA.

conserved and may be located on the DNA contact face of the basic region α helix (20). Based on other examples of base-specific, protein-DNA interactions involving alanine residues (13, 35), the 5'-methyl group of thymine is the most likely partner to interact with the alanine methyl group. Thus, replacement of the critical thymine with uracil, which differs from thymine by its lack of the 5-methyl group, would be expected to decrease DNA-binding affinity (5, 25, 29). Furthermore, mutation of the critical alanine should abolish the effect that the particular uracil substitution has on binding affinity (3).

The role of the 5-methyl groups of thymine in high-affinity DNA binding by GCN4 was examined by a novel uracil interference assay that we developed recently. By using PCR, a DNA fragment containing the GCN4 recognition sequence was randomly substituted with deoxyuracil. DNA molecules that could be tightly bound by GCN4 were selected by purifying the DNA-GCN4 complex. The distribution of uracil residues before and after selection was then analyzed by cleaving the DNA at uracil residues, using uracil-*N*-glycosylase followed by piperidine (Fig. 7).

For GCN4, substitution of uracil for the two central thymines of the half-site, but not the outermost thymine, significantly interferes with DNA binding, indicating that the methyl groups on the four core thymines of the optimal site are important recognition elements. However, the critical thymine residues on the noncoding strand (ATGAGTCAT; Fig. 7A) appear more important than their counterparts on the coding strand (ATGACTCAT; Fig. 7B) because they are more sensitive to uracil substitution. In addition, uracil substitutions on the coding strand reveal that the thymine at position -3 is relatively important, whereas the thymine at position +1 has only a modest effect on binding (positions defined with respect to the central C · G base pair) (18). Though unexpected, these asymmetric interactions with the thymine residues are consistent with the inherent asymmetry of the optimal GCN4 binding site and the relative importance of the left half-site over the right half-site (18, 25, 27).

Analysis of the Val-238, Cys-238, and Gly-239 derivatives of GCN4 indicates that the four core thymine residues remain as essential recognition elements. On the noncoding strand, all of these proteins behave equivalently, with either of the uracil substitutions essentially eliminating binding. On

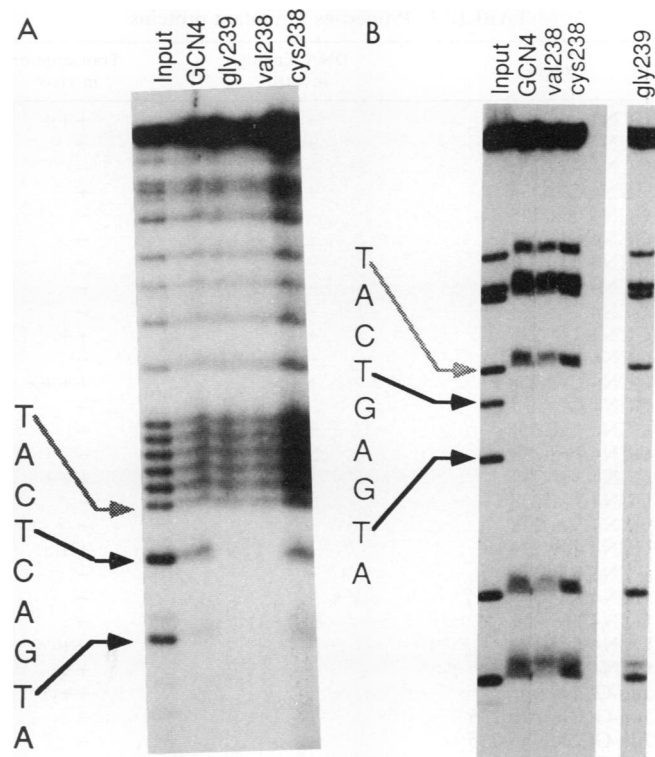


FIG. 7. Uracil interference assay. Complexes between the indicated proteins and a uracil-substituted DNA fragment containing the optimal GCN4 binding site were isolated, and the recovered DNA was cleaved specifically at uracil residues by treatment with uracil-*N*-glycosylase followed by piperidine. (A) 5' end label on the coding strand (defined with respect to the *his3* structural gene); (B) 5' end label on the noncoding strand.

the coding strand, the Cys-238 protein behaves indistinguishably from GCN4, but the Val-238 and Gly-239 proteins display a surprising increase in sensitivity to uracil substitutions. This apparent increased importance of the thymines on the noncoding strand may be related to the somewhat reduced DNA-binding activity of both the Val-238 and Gly-239 proteins. Whatever the molecular basis for this effect, these results conflict with those predicted by the model that the conserved alanines interact directly with the critical thymines in the target site.

DISCUSSION

Evidence that N-cap formation is not an essential role of Asn-235. The scissors grip (33) and induced fork (20) models of the bZIP protein-DNA complex differ in the predicted roles of the invariant asparagine residue in the basic region. The scissors grip model proposes that the asparagine forms an N-cap structure that breaks the basic region α helix, whereas induced fork model proposes that the asparagine directly contacts DNA. To address these models, we examined the functional consequences of amino acid substitutions of Asn-235 in GCN4 and Jun-GCN4 (summarized in Table 2). For GCN4, substitution of asparagine 235 with tryptophan, alanine, glutamine, and glycine generates proteins that bind DNA in vitro. Tryptophan can also functionally substitute for the invariant asparagine in Jun-GCN4, with a resultant increase in DNA-binding affinity.

TABLE 2. Properties of mutant proteins

Protein	DNA binding in vitro ^a	Transcription in vivo ^b
GCN4	+++	+++
GCN4 Ala-235	+	+
GCN4 Arg-235	-	-
GCN4 Gln-235	+	+
GCN4 Glu-235	-	-
GCN4 Gly-235	±	-
GCN4 Leu-235	-	-
GCN4 Lys-235	-	-
GCN4 Thr-235	-	-
GCN4 Trp-235	++	++
GCN4 Val-235	-	-
GCN4 Cys-238	+++	+++
GCN4 Gly-238	-	-
GCN4 Leu-238	±	-
GCN4 Pro-238	NT	-
GCN4 Ser-238	+++	+++
GCN4 Thr-238	NT	-
GCN4 Val-238	+	+
GCN4 Gly-239	++	++
GCN4 Leu-239	-	-
GCN4 Phe-239	-	-
GCN4 Pro-239	NT	-
GCN4 Ser-239	NT	++
GCN4 Val-239	++	++
Jun-GCN4	±	++
Jun-GCN4 Arg-235	-	-
Jun-GCN4 Cys-235	-	-
Jun-GCN4 His-235	-	-
Jun-GCN4 Leu-235	-	-
Jun-GCN4 Phe-235	-	-
Jun-GCN4 Ser-235	-	-
Jun-GCN4 Trp-235	+	+++
Jun-GCN4 Val-235	-	-

^a Determined as shown in Fig. 4 and 6. NT, not tested.

^b Assayed by AT resistance in a *gcn4* complementation assay as shown in Fig. 3 and 5.

Amino acid preferences at the N-cap position have been calculated from a collection of 215 α helices derived from solved protein structures (24). The most strongly preferred amino acids at the N-cap position are asparagine, serine, aspartate, and threonine. Of these, only asparagine at position 235 is compatible with DNA binding; the serine substitution of Jun-GCN4 and the serine, aspartate, and threonine substitutions of GCN4 do not function in vivo. In contrast, the Trp-235 and Gln-235 derivatives of GCN4 bind DNA despite tryptophan and glutamine being among the least preferred residues at the N-cap position. Thus, the ability of a protein containing an amino acid substituted for Asn-235 to bind DNA does not correlate with the likelihood that the residue forms an N cap. This strongly argues against the prediction of the scissors grip model that the crucial role of Asn-235 is to form an N cap. However, our results do not rule out the more general feature of this model that the basic region bends to maximize the protein-DNA interface.

Although the invariant asparagine does not appear to form an N cap, it is clearly important for activity. Besides being present in all known bZIP proteins, most of the Asn-235 substitutions of GCN4 and Jun-GCN4 examined here abolish DNA-binding activity in vitro. Moreover, of three residues shown to functionally substitute for Asn-235 in GCN4 and Jun-GCN4, all clearly reduce DNA-binding affinity. Finally, we have shown that the Ala-235, Gln-235, and Trp-235 proteins all have altered specificities at the ± 4 position of the

GCN4 target site (32), suggesting that Asn-235 makes direct, base-specific contacts with the fourth position of the GCN4 target site.

Evidence that Ala-238 and Ala-239 do not make base-specific DNA contacts. We have tested the prediction that Ala-238 and Ala-239 of GCN4 make base-specific DNA contacts (20) by individually randomizing these codons and testing the resulting mutants for GCN4 function (summarized in Table 2). At both positions 238 and 239, DNA-binding affinity decreases as the size of the side chain increases, consistent with the location of these residues along the putative DNA contact face of the basic region α helix. However, substitution of glycine at position 239 only slightly decreases DNA-binding affinity from wild-type levels, suggesting that the side chain at 239 does not contribute significantly to DNA binding. Moreover, the Gly-239 protein interacts equally or more strongly with thymine methyl groups in the recognition sequence than GCN4, strongly suggesting that Ala-239 is not essential for these interactions. Since the methyl side chain of alanine is relatively unimportant and the only known base-specific interaction by alanine is with the 5-methyl group of thymine (13, 35), it seems unlikely that Ala-239 is involved in direct base pair contacts. However, we cannot exclude the possibility that Ala-239 directly interacts with some other functional group(s) of the DNA.

The ability of the Cys-238, Ser-238, Val-238, and Leu-238 proteins to bind DNA suggests that Ala-238 also does not make a hydrophobic interaction with thymine. In order to make such a contact, Ala-238 would have to approach the thymine very closely, so that it would be unlikely that a bulkier group such as valine or leucine could be sterically accommodated. Moreover, the core thymine residues are equally or more important for binding by the Cys-238 and Val-238 proteins, arguing against a direct contact by Ala-238. Instead, by analogy with randomization experiments on the dimerization interface of λ repressor (23), the relationship between DNA-binding activity and the size of the hydrophobic side chain at positions 238 and 239 suggests the importance of hydrophobic packing interactions between protein and DNA.

Regulation of GCN4 activity in vivo. The in vitro DNA-binding activity of a mutant protein generally correlates very well with its functional activity in vivo. However, Jun-GCN4 and Jun-GCN4^{Trp} proteins are much more functional in vivo than GCN4 mutant proteins with comparable DNA-binding activities (Table 2). Measurements of *his3* mRNA levels confirm that Jun-GCN4^{Trp} and Jun-GCN4 activate *his3* expression much more strongly than GCN4, as expected (4).

This unexpected result could be explained in several ways. First, the surprisingly high activity of the Jun-GCN4 derivatives in comparison with the GCN4 derivatives might simply reflect higher protein levels. This would have to arise from differences between the GCN4 and Jun basic regions that affect protein stability, because the proteins are expressed in the identical manner and are nearly identical in sequence. Second, although GCN4 and Jun appear to bind identical DNA sequences (28), it is conceivable that they have slightly different binding specificities. Since the yeast genome contains many binding sites that compete for the limited amount of GCN4 that exists in vivo (7), proteins with nonidentical binding specificities could differ in the spectra of target genes that are transcriptionally activated. Third, the GCN4 and Jun basic regions could be differentially modified in yeast cells, with a concomitant change in binding affinity that would not be detected in our in vitro binding experi-

ments. Fourth, the GCN4 and Jun basic regions could differ in their ability to interact with a coactivator or corepressor protein. In this regard, these basic regions are only about 50% identical, with most of the differences presumed to lie on the solvent-exposed surface that should be accessible for protein-protein interactions. By analogy, the GAL11 coactivator increases the transcriptional activity of GAL4 derivatives, possibly by associating directly with the GAL4 DNA-binding domain (8, 17).

Evolutionary considerations. It is generally believed that evolutionarily conserved amino acid residues are functionally important. However, our results indicate that three highly conserved residues in the basic region of bZIP proteins are not essential for DNA-binding activity. A number of substitutions are compatible with function, and some of them have no detectable effect in vitro or in vivo (Table 2).

Why are these residues so highly conserved? It is possible that our biological and biochemical assays are not sufficiently sensitive to detect subtle functional defects. A small difference, though nearly inconsequential in the life of an individual cell, could be magnified through numerous generations of competitive growth. Another possibility is that these conserved bZIP residues might be important not for binding per se but rather for maximizing the range of distinct DNA sequences that can be efficiently bound. For example, the genomic binding sites in GCN4-regulated promoters, though clearly related, differ in their precise DNA sequences (7). Finally, the residues might be conserved not for their inherent functional importance but rather for their compatibility with a variety of amino acid residues at nearby positions in the basic region. In this sense, the conserved residues would contribute to evolutionary and regulatory flexibility by allowing more possible DNA-binding proteins. Regardless of the validity of any of these specific explanations, the results presented here indicate the complications in attempting to use evolutionary conservation for predicting functional importance.

ACKNOWLEDGMENTS

We thank Chris Brandl for DNA containing the altered GCN4 coding region, Dimitris Tzamarias for constructing strain KY1365, Alice Vincent for compiling the list of bZIP proteins, and Ann Hochschild and Michael Weiss for fruitful discussion.

This work was supported by a Johnson and Johnson fellowship to W.T.P. and by grants to K.S. from the National Institutes of Health (GM 30186) and the Lucille Markey Trust.

REFERENCES

- 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**:922-926.
- Arndt, K., and G. Fink. 1986. GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5' TGACTC 3' sequences. *Proc. Natl. Acad. Sci. USA* **83**:8516-8520.
- Brunelle, A., and R. F. Schleif. 1987. Missing contact probing of DNA-protein interactions. *Proc. Natl. Acad. Sci. USA* **84**:6673-6676.
- Collart, M., and K. Struhl. Unpublished data.
- Dubendorff, J. W., P. L. deHaseth, M. S. Rosendahl, and M. H. Caruthers. 1987. DNA functional groups required for formation of open complexes between *E. coli* RNA polymerase and the λ P_R promoter. *J. Biol. Chem.* **262**:892-898.
- Gentz, R., F. I. Rauscher, C. Abate, and T. Curran. 1989. Parallel association of Fos and Jun leucine zippers juxtaposes DNA binding domains. *Science* **243**:1695-1699.
- Hill, D. E., I. A. Hope, J. P. Macke, and K. Struhl. 1986. Saturation mutagenesis of the yeast *HIS3* regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. *Science* **234**:451-457.
- Himmelfarb, H. J., J. Pearlberg, D. H. Last, and M. Ptashne. 1990. GAL11P: a yeast mutation that potentiates the effect of weak GAL4-derived activators. *Cell* **63**:1299-1309.
- 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.
- Hope, I. A., and K. Struhl. 1985. GCN4 protein, synthesized in vitro, binds to *HIS3* regulatory sequences: implications for the general control of amino acid biosynthetic genes in yeast. *Cell* **43**:177-188.
- Hope, I. A., and K. Struhl. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46**:885-894.
- Hope, I. A., and K. Struhl. 1987. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *EMBO J.* **6**:2781-2784.
- Jordan, S. R., and C. O. Pabo. 1988. Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions. *Science* **242**:893-899.
- Kouzarides, T., and E. Ziff. 1988. The role of the leucine zipper in the fos-jun interaction. *Nature (London)* **336**:646-651.
- 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.
- 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.
- Nishizawa, M., Y. Suzuki, Y. Nogi, K. Matsumoto, and T. Fukasawa. 1990. Yeast GAL11 protein mediates the transcriptional activation signal of two different transacting factors, GAL4 and general regulatory factor 1/repressor/activator site binding protein 1/translation upstream factor. *Proc. Natl. Acad. Sci. USA* **87**:5373-5377.
- Oliphant, A. R., C. J. Brandl, and K. Struhl. 1989. Defining sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of the yeast GCN4 protein. *Mol. Cell. Biol.* **9**:2944-2949.
- Oliphant, A. R., A. L. Nussbaum, and K. Struhl. 1986. Cloning of random-sequence oligodeoxynucleotides. *Gene* **44**:177-183.
- 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.
- O'Shea, E. K., R. Rutkowski, and P. S. Kim. 1989. Evidence that the leucine zipper is a coiled coil. *Science* **243**:538-542.
- Pu, W. T., and K. Struhl. 1991. The leucine zipper symmetrically positions the adjacent basic regions for specific binding to DNA. *Proc. Natl. Acad. Sci. USA* **88**:6901-6905.
- Reidhaar-Olson, J. F., and R. T. Sauer. 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* **241**:53-57.
- Richardson, J. S., and D. C. Richardson. 1988. Amino acid preferences for specific locations at ends of α helices. *Science* **240**:1648-1652.
- Risse, G., K. Jooss, M. Neuberg, H.-J. Bruller, and R. Muller. 1989. Asymmetrical recognition of the palindromic AP1 binding site (TRE) by Fos protein complexes. *EMBO J.* **8**:3825-3832.
- Sellers, J. W., and K. Struhl. 1989. Changing fos oncoprotein to a DNA-binding protein with GCN4 dimerization specificity by swapping "leucine zippers." *Nature (London)* **341**:74-76.
- 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.
- Struhl, K. 1987. The DNA-binding domains of the jun oncoprotein and the yeast GCN4 transcriptional activator are functionally homologous. *Cell* **50**:841-846.
- Takeda, Y., A. Sarai, and V. M. Rivera. 1989. Analysis of sequence-specific interactions between Cro repressor and operator DNA by systematic base substitution experiments. *Proc. Natl. Acad. Sci. USA* **86**:439-443.

30. **Talanian, R. V., C. J. McKnight, and P. S. Kim.** 1990. Sequence-specific DNA binding by a short peptide dimer. *Science* **249**:769–771.
31. **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.
32. **Tzamarias, D., W. T. Pu, and K. Struhl.** Submitted for publication.
33. **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.
34. **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**:575–578.
35. **Wharton, R. P., and M. Ptashne.** 1987. A new-specificity mutant of 434 repressor that defines an amino acid-base pair contact. *Nature (London)* **326**:888–891.