Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 6901–6905, August 1991
Biochemistry

The leucine zipper symmetrically positions the adjacent basic regions for specific DNA binding

(protein–DNA interactions/bZIP domain/yeast GCN4/transcription factor)

WILLIAM T. PU AND KEVIN STRUHL

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

Communicated by Thomas A. Steitz, May 13, 1991 (received for review February 22, 1991)

ABSTRACT **The bZIP structural motif present in several eukaryotic transcription factors is defined by the leucine zipper, a coiled-coil dimerization interface, and an adjacent basic region that directly interacts with DNA. To examine the functional importance of the highly conserved spacing between the leucine zipper and the basic region, we have analyzed the DNA-binding ability of yeast GCN4 proteins containing amino acid insertions between these two subdomains. Proteins containing a surprisingly wide variety of seven-amino acid insertions, but none containing two-, four-, or six-amino acid insertions, are functional. However, heterodimers between wild-type GCN4 and functional derivatives containing seven amino acid insertions are unable to bind DNA. These observations provide strong experimental support for several aspects of the scissors grip and induced fork models for DNA-binding by bZIP proteins. Specifically, they demonstrate that continuous α-helices symmetrically diverging from the leucine zipper correctly position the two basic regions for specific binding to abutting DNA half-sites. In addition, the results indicate that GCN4 homodimers are primarily responsible for transcriptional activation in yeast cells.**

The yeast GCN4 protein belongs to the class of eukaryotic transcription factors whose DNA-binding domains contain a bZIP structural motif (1). bZIP proteins bind as dimers to dyad-symmetric target sequences (2–5), indicative of protein–DNA complexes in which two protein monomers interact with adjacent half-sites. The conserved bZIP structural motif that defines this class of proteins consists of a leucine zipper dimerization element (1) that structurally resembles a coiled coil (6, 7) and an adjacent basic region that determines DNA-binding specificity (8). The leucine zipper and basic region are distinct structural subdomains because they can be interchanged between different family members to generate chimeric proteins with predicted dimerization and DNA-binding specificities (8–11). Moreover, synthetic leucine zipper peptides form dimers of appropriate specificity (6, 12), and a synthetic basic region (dimerized via a disulfide bond) can specifically bind DNA, although with reduced affinity (13).

The spacing between the leucine zipper and basic region (defined by the distance between the N-terminal leucine in the zipper and the conserved pair of C-terminal arginines in the basic region) is seven amino acids in all known bZIP proteins. Moreover, insertions of five amino acids between the leucine zipper and the basic region of transcription factor C/EBP abolish DNA binding (8). Largely on the basis of the conserved spacing between the subdomains, it was proposed that the leucine zipper correctly and symmetrically positions the two basic regions for specific DNA binding to adjacent half-sites (14). Further, this "scissors grip" model suggested that diverging α-helices from the leucine zipper continued into the adjacent basic regions. A more recently proposed "induced fork" model makes these same predictions (7) but differs from the scissors grip model with respect to whether the invariant asparagine in the basic region directly contacts DNA or forms an "N-cap" that permits a sharp bend for wrapping around the DNA. The fact that the GCN4 bZIP domain is almost entirely α-helical when bound to its target site (7, 15) is consistent with both models.

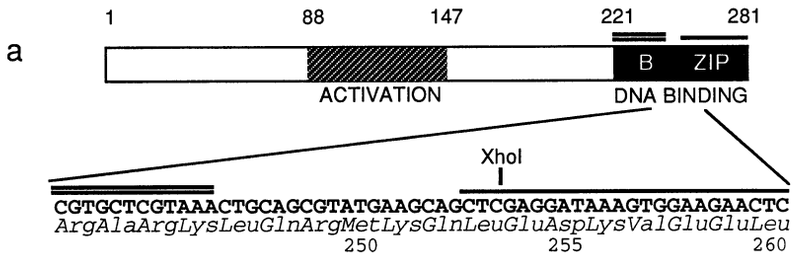
Here, we test critical aspects of the scissors grip and induced fork models by analyzing the DNA-binding properties of GCN4 mutants in which the spacing between the leucine zipper and basic region has been altered. We show that although the spacing between these two subdomains is crucial for DNA binding, it can be altered by the insertion of an integral number of α-helical turns (seven amino acids) without affecting GCN4 function. Further, efficient DNA binding requires that the two basic regions of the GCN4 dimer must be symmetrically related with respect to the leucine zipper. These observations strongly suggest that continuous α-helices symmetrically diverging from the leucine zipper correctly position the two basic regions for specific binding to abutting DNA half-sites.

MATERIALS AND METHODS

Construction of Insertion Mutants. The DNA molecules used in this paper were derived from YCp88-GCN4, a *URA3* centromeric vector that permits expression of GCN4 in yeast cells from the *DED1* promoter and *in vitro* from the bacteriophage SP6 promoter (16). To facilitate manipulations, the *GCN4* gene was modified without changing the encoded protein by using oligonucleotides to introduce restriction sites throughout the bZIP domain, including the *Xho* I site where all insertion mutations were made (Fig. 1). To generate libraries containing either LEX_4 or LEX_5 insertions between the leucine zipper and basic region, degenerate double-stranded oligonucleotides (boxed positions were programmed to be 79% wild-type and 7% each other nucleotide; Fig. 1) were inserted in either orientation at the *Xho* I site. These were obtained by converting CGACTCGAG-GATAAAGTGGAACTCGAGT or AGGCTCGAGTTCT-TCCACTTTATCCTCGAG (underlined residues indicate positions of degeneracy) to double-stranded DNA by mutually primed synthesis (17), followed by cleavage with *Xho* I. Derivatives containing two- or four-amino acid inserts were constructed by inserting *Xho* I linkers. The insertion mutants are named by the position and length of the additional residues; e.g., 252In7-1 means the first isolate of a seven-residue insertion at position 252.

Phenotypic Analysis. Libraries encoding the LEX_4 or LEX_5 insertion proteins were introduced into the *Saccharomyces cerevisiae gcn4* deletion strain KY803 by selecting for uracil-independent transformants, and the resulting strains were assayed for GCN4 function by growth in 20–80 mM aminotriazole as described previously (16). Plasmid DNAs were

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.



c

DPAALKRARNTEAARRSRARKLQRMKQ LEDKVEE LLSKNYH LENEVAR LKKLVGER

Name		Phenotype
253In2-1	MKQ L**EV** EDKVEE L	−
253In2-2	MKQ L**DL** EDKVEE L	−
253In4-1	MKQ L**DLEV** EDKVEE L	−
252In6-1	MKQ **LEANVA** LEDKVEE L	−
252In6-2	... **LEDKVE** ...	−
252In6-3	... **LEFHCI** ...	−
252In6-4	... **LEFQFM** ...	−
252In6-5	... **LEAQVV** ...	−
252In6-6	... **LECHFI** ...	−
252In7-1	MKQ **LEFFRFI** LEDKVEE L	++
252In7-2	... **LELLHCI** ...	+++
252In7-3	... **LEFNHFI** ...	++
252In7-4	... **LEIFYNI** ...	++
252In7-6	... **LEDKGDE** ...	++
252In7-7	... **LEDVEEE** ...	+++
252In7-8	... **LELFQFI** ...	+
252In7-9	... **LEDEVVE** ...	++
252In7-10	... **LEFLHFI** ...	++
252In7-11	... **LEFFQFE** ...	+
252In7-12	... **LEDIVEE** ...	++
252In7-14	... **LESYHLI** ...	++
252In7-15	... **LESYHFI** ...	++
252In7-16	... **LEFFHFI** ...	++
252In7-17	... **LEDNVQG** ...	++
252In7-18	... **LEYVNFI** ...	++
252In7-20	... **LEHFNCI** ...	−
252In7-21	... **LEFFICI** ...	−
252In7-22	... **LEFFPFI** ...	−
252In7-27	... **LEFCPFI** ...	−

FIG. 1. Construction and characterization of insertion mutants. (*a*) Structure of GCN4, showing the locations of the activation and DNA-binding (bZIP) domains (numbers refer to amino acid residues), the DNA sequence of the junction between the basic region (double bar) and leucine zipper (single bar), and the amino acid sequence (italics). The *GCN4* allele encodes the wild-type protein but has been modified to introduce restriction sites, including the indicated *Xho* I site, where all insertion mutations were made. (*b*) Degenerate double-stranded oligonucleotides [boxed positions were programmed to be 79% wild-type (WT) and 7% each other nucleotide] were inserted in either orientation at the *Xho* I site, thus generating either LEX_4 (left) or LEX_5 (right) insertions between the leucine zipper and basic region. (*c*) Insertion mutants used in this study. Listed below the sequence (standard one-letter symbols) of wild-type GCN4 are the names (e.g., 252In7-1 means the first isolate of a seven-residue insertion at position 252), inserted sequences (bold), and phenotypes of the insertion mutants. The phenotype refers to the ability of these mutants to complement a *gcn4* deletion strain (see *Materials and Methods*).

recovered from individual yeast strains by transformation into *Escherichia coli*, and the sequences of the inserted regions were determined. These DNAs were also reintroduced into KY803 to confirm their growth phenotypes, which are designated as follows: +++, indistinguishable from wild type; ++, slightly worse than wild type; +, weak but detectable; −, no growth.

DNA-Binding Experiments. For analysis of individual GCN4 derivatives, equal amounts of *in vitro* synthesized proteins (determined by SDS/PAGE) were incubated as described previously (18) with a ^{32}P-labeled 435-base-pair (bp) DNA fragment containing the optimal GCN4 binding site at 8 nM. The protein–DNA complex was separated from unbound [^{32}P]-DNA by PAGE. To determine DNA binding by heterodimers, equal amounts of the unlabeled GCN4 derivatives were combined with a constant amount of ^{35}S-labeled GCN4-C131 (contains the 131 C-terminal residues; referred to as ^{35}S-C131) (16), incubated with the 435-bp fragment containing the optimal GCN4 site at 8 nM. The protein–DNA complexes involving the GCN4/^{35}S-C131 heterodimer and the ^{35}S-C131/^{35}S-C131 homodimer were electrophoretically separated. L253V-L274V is a non-DNA-binding derivative in which two leucines of the leucine zipper are replaced by valine.

Glutaraldehyde Crosslinking. Equal amounts of *in vitro* synthesized, ^{35}S-labeled proteins were incubated for 10 min at room temperature in 20 mM potassium phosphate at pH 7.0/100 mM KCl/3 mM $MgCl_2$/1 mM EDTA, in either the presence or the absence of 6 μM GCN4p, the 58-residue GCN4 DNA-binding domain produced in *E. coli* (15). The proteins were crosslinked for 2 hr in 0.005% glutaraldehyde, and the reaction was stopped by adding Tris·HCl at pH 7.5 to 25 mM. The crosslinked homodimer of the full-length protein (^{35}S-GCN4/^{35}S-GCN4), the crosslinked heterodimer between the full-length derivative and the C58 peptide (^{35}S-GCN4/GCN4p), and the uncrosslinked full-length monomer (^{35}S-GCN4) were separated by SDS/PAGE. The conditions for specific crosslinking were optimized for the concentration of glutaraldehyde, incubation time, pH, and concentration of GCN4p. One possibility for the apparent inefficient crosslinking is that glutaraldehyde might preferentially react with the monomeric leucine zipper, thus inhibiting dimerization.

RESULTS

Insertions Between the Leucine Zipper and the Basic Region. The spacing between the leucine zipper and the basic region is highly conserved, and DNA-binding of transcription factor

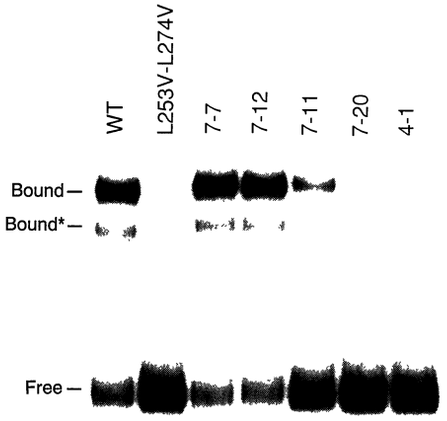


FIG. 2. DNA-binding activities of GCN4 derivatives. Equal amounts of the indicated proteins synthesized *in vitro* were incubated with a ^{32}P-labeled DNA fragment containing the optimal GCN4 binding site, and the protein–DNA complexes (bound) were electrophoretically separated from unbound [^{32}P]DNA. The complex indicated as bound* most likely 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. Homodimers of the C-terminal fragment are not observed because the artifactual translation products represent only 5–10% of the protein synthesized *in vitro*. The presence and amount of the bound* complex depends on the batch of translation extract.

C/EBP is abolished upon insertion of five amino acids between these subdomains (8). Consistent with these observations, insertion of two or four amino acids between the GCN4 leucine zipper and the basic region abolishes GCN4 function *in vivo* (Fig. 1). However, the failure of these insertion mutations to bind DNA might simply reflect a general disruption of the protein; hence they do not prove the hypothesis that the zipper orients the basic region on target DNA.

If continuous α-helix extending N-terminally from the leucine zipper is required for positioning the basic region, one might predict that the correct spatial relationship would be retained upon the insertion of an integral number of α-helical turns (e.g., seven residues) between the zipper and basic region. Accordingly, we introduced a library of GCN4 proteins containing LEX_5 heptapeptide insertions into a yeast *gcn4* deletion strain and examined the resulting transformants for GCN4 function (Fig. 1). Strikingly, the vast majority (35 out of 40 tested) of the LEX_5 heptapeptide insertion proteins display GCN4 function (ranging from weak to wild-type levels of activity), and the inserted sequences vary considerably. For example, in the region containing the five variable amino acids, 7-7 contains four acidic residues while 7-8 and 7-16 contain four hydrophobic residues. As expected from the α-helix hypothesis, none of the 19 derivatives conferring GCN4 function has a proline residue in the variable region; however, two of them (7-6 and 7-17) contain a single glycine. Of 10 derivatives examined that lack GCN4 function (Fig. 1), 6 are trivially accounted for by ochre, amber, or frameshift mutations, 2 (7-22 and 7-27) contain a proline residue, and 2 are not easily explained (especially 7-21, which differs at only two positions from 7-16 and 7-1).

As a control, a library of GCN4 proteins containing LEX_4 hexapeptide insertions was analyzed in a similar manner. In sharp contrast to the results with heptapeptide insertions, none of 144 transformants containing an LEX_4 insertion conferred GCN4 function. Plasmids recovered from six transformants lacking GCN4 function revealed a variety of six-amino acid insertions (Fig. 1). As the LEX_4 and LEX_5 insertions differ in length but not amino acid composition, it is clear that the spacing between the zipper and basic region is critical for function. Nevertheless, a correct spacing relationship can be restored by the insertion of seven residues, which corresponds to two α-helical turns.

Dimerization and DNA Binding by Insertion Proteins. Several of these mutant proteins were synthesized *in vitro* and assayed for their DNA-binding activity (Fig. 2). In all cases, DNA-binding activity of the proteins synthesized *in vitro* correlates with transcriptional activation *in vivo*. The 7-7 and 7-12 proteins bind with wild-type affinity, 7-11 binds with reduced affinity, and the nonfunctional protein 7-20 fails to bind detectably. Thus, while insertion of many different heptapeptides is compatible with GCN4 activity *in vivo*, the sequence of the insertion can affect DNA-binding activity.

To investigate the possibility that the nonfunctional insertions could eliminate DNA-binding activity by interfering with dimerization, we performed glutaraldehyde crosslinking experiments with radiolabeled proteins (Fig. 3). Although

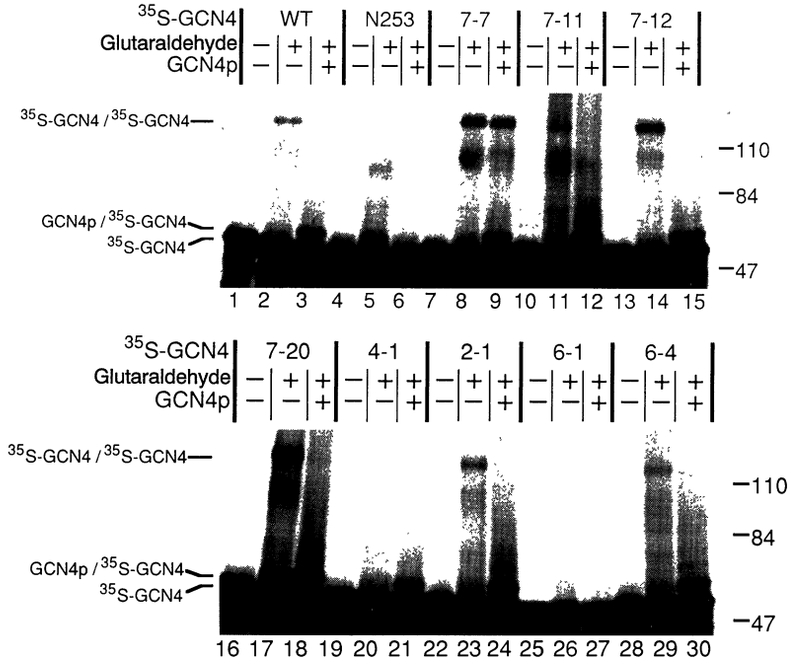


FIG. 3. Homodimer and heterodimer formation. Equal amounts of *in vitro* synthesized, ^{35}S-labeled GCN4 (WT), N253 (contains the N-terminal 253 residues and thus lacks almost the entire leucine zipper), and the indicated insertion derivatives were crosslinked with glutaraldehyde in the presence or absence of GCN4p (15). Bands representing the crosslinked homodimer of the full-length protein (^{35}S-GCN4/^{35}S-GCN4), the crosslinked heterodimer between the full-length derivative and the C58 peptide (GCN4p/^{35}S-GCN4), and the uncrosslinked full-length monomer (^{35}S-GCN4) are indicated (left) as are molecular weight markers ($\times 10^{-3}$) (right).

only a low percentage of wild-type GCN4 molecules are crosslinked under our experimental conditions, the crosslinked products represent specific dimer formation dependent on the leucine zipper (compare lanes 2 and 5). Moreover, an excess of unlabeled GCN4p (15), a 58-residue DNA-binding domain produced in *E. coli*, changes the major radiolabeled crosslinked product from a full-length homodimer to a heterodimer between the full-length protein and GCN4p.

As expected, the functional heptapeptide insertion proteins (7-7, 7-11, and 7-12) efficiently form homodimers. Of the proteins unable to bind DNA, proteins 7-20, 2-1, and 6-4 form homodimers, while proteins 4-1 and 6-1 do not (4-1 does form heterodimers with GCN4p). For those derivatives that dimerize, it is likely that their insertions interfere with the ability of the basic regions to interact properly with DNA. The inability of the 7-20 protein to bind DNA is clearly an effect specific to the sequence LEHFNCI, since the majority of seven-amino acid insertions are tolerated. In contrast, the fact that two- and six-amino acid insertion proteins can dimerize but not bind DNA suggests that the defect in positioning of the basic region in these mutants is due to the length rather than the sequence of the insertion.

Heterodimers Between GCN4 and Heptapeptide Insertion Proteins Do Not Bind DNA. The above results demonstrate that the leucine zipper is required for properly orienting the adjacent basic region for DNA binding. To ask whether the leucine zipper must symmetrically orient the two basic regions along the adjacent DNA half-sites, we determined the DNA-binding activity of heterodimers between wild-type GCN4 and several heptapeptide insertion proteins (7-7, 7-11, 7-12). Although all of these proteins efficiently interact with a GCN4 target site and hence form fully functional DNA-binding homodimers, mixtures of the mutants with wild-type GCN4 do not result in detectable levels of DNA-binding heterodimers (Fig. 4). The inability to bind is not due to a failure to form heterodimers, since crosslinking experiments show that the mutant proteins form heterodimers with a wild-type DNA-binding domain (Fig. 3). Thus, both the wild-type and heptapeptide insertion proteins contain an acceptable spacing between the leucine zipper and the basic region, but the distinct spacings are not mutually compatible.

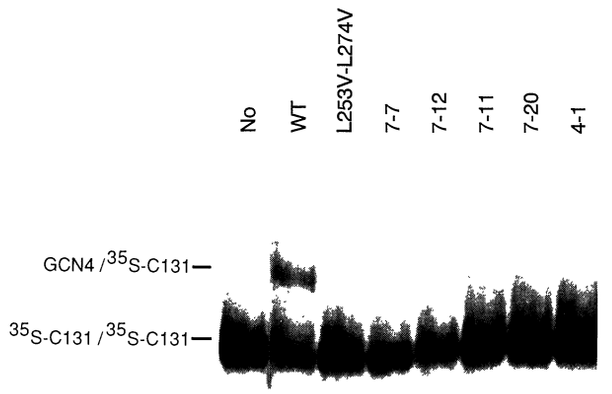


FIG. 4. Heterodimers between GCN4 and insertion mutants do not bind DNA. Equal amounts of the indicated unlabeled GCN4 derivatives were combined with a constant amount of ^{35}S-C131 (16) and incubated with DNA fragment containing the optimal GCN4 site at 8 nM. The positions of complexes involving the GCN4/^{35}S-C131 heterodimer and the ^{35}S-C131/^{35}S-C131 homodimer are indicated.

DISCUSSION

Our results (summarized diagrammatically in Fig. 5) provide strong experimental support for several of the key proposals of the scissors grip (14) and induced fork (7) models. First, the spacing between the zipper and basic region is critical for function, because fewer than 1% of LEX_4 insertion proteins but more than 80% of the LEX_5 insertion proteins are functionally active, despite the similarities of the inserted amino acid sequences. Second, the DNA-binding ability of derivatives containing insertions of an integral number of α-helical turns (seven amino acids) but not a nonintegral number of turns (two, four, or six amino acids) is strongly indicative of a continuous α-helical region between the two subdomains. Third, the indistinguishable DNA-binding activities of GCN4 and several heptapeptide insertion proteins indicate that the α-helix connecting the zipper and the basic region is required for the correct structural relationship between these subdomains. Fourth, the failure of heterodimers between GCN4 and functional heptapeptide insertion proteins to bind DNA demonstrates that the two basic regions of the dimer must be related by a dyad to properly interact with the binding site. Moreover, the component α-helices of the GCN4 leucine zipper must associate in a specific alignment to generate a symmetrical coiled coil; i.e., the heptad repeats cannot shift in register. These observations, however, do not distinguish between the scissors grip and induced fork models regarding the proposed bend in the basic region and the role of the invariant asparagine residue.

Our observations also suggest that GCN4 homodimers are the primary, and possibly the sole, mediators of GCN4 function in yeast cells. Because heterodimers between GCN4 and the heptapeptide insertion proteins do not bind DNA, it seems extremely unlikely that these proteins could associate into DNA-binding heterodimers with any given yeast bZIP protein. Thus, the indistinguishable phenotypes of GCN4 and the heptapeptide insertion proteins probably reflect the similar DNA-binding and transcriptional activities of the homodimers. More generally, equivalent derivatives with altered spacing between the leucine zipper and the basic region could be useful for determining whether a specific function(s) of a given bZIP protein is mediated by homodimers or by heterodimers.

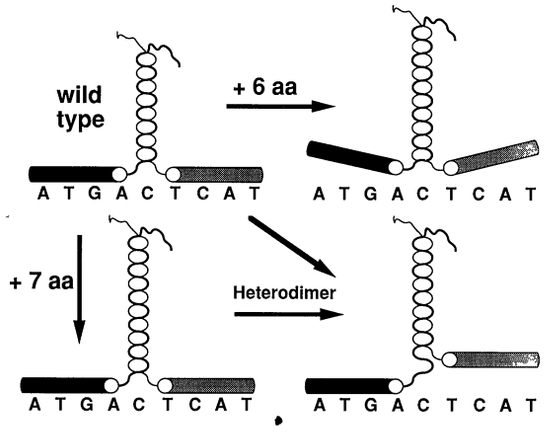


FIG. 5. The leucine zipper symmetrically positions the adjacent basic regions for DNA binding. The spacing between the leucine zipper (intertwined wavy lines perpendicular to the DNA) and both adjacent basic regions (cylinders) must be correct for high-affinity binding (shown by a close parallel arrangement of both basic regions to the DNA sequence). The region between the zipper and basic region is shown as α-helical because this spacing can be altered by the insertion of seven amino acids, an integral number of α-helical turns. When the α-helical periodicity is disrupted by inserting six (also two, four, and five) residues between the two subdomains, the basic regions are improperly oriented along the DNA. A heterodimer between wild-type GCN4 and a derivative containing a seven-amino acid insertion does not bind DNA because its two basic regions cannot be simultaneously positioned along the adjacent half-sites.

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 (GM30186) and the Lucille P. Markey Charitable Trust.

1. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240,** 1759–1764.
2. Hope, I. A. & Struhl, K. (1987) *EMBO J.* **6,** 2781–2784.
3. Halazonetis, T. D., Georgopoulos, K., Greenberg, M. E. & Leder, P. (1988) *Cell* **55,** 917–924.
4. Nakabeppu, Y., Ryder, K. & Nathans, D. (1988) *Cell* **55,** 907–915.
5. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1989) *Science* **243,** 1681–1688.
6. O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1989) *Science* **243,** 538–542.
7. O'Neil, K. T., Hoess, R. H. & DeGrado, W. F. (1990) *Science* **249,** 774–778.
8. Agre, P., Johnson, P. F. & McKnight, S. L. (1989) *Science* **246,** 922–926.
9. Kouzarides, T. & Ziff, E. (1989) *Nature (London)* **340,** 568–571.
10. Sellers, J. W. & Struhl, K. (1989) *Nature (London)* **341,** 74–76.
11. Neuberg, M., Schuermann, M., Hunter, J. B. & Muller, R. (1989) *Nature (London)* **338,** 589–590.
12. O'Shea, E. K., Rutkowski, R., Stafford, W. F. I. & Kim, P. S. (1989) *Science* **245,** 646–648.
13. Talanian, R. V., McKnight, C. J. & Kim, P. S. (1990) *Science* **249,** 769–771.
14. Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) *Science* **246,** 911–916.
15. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C. & Struhl, K. (1990) *Nature (London)* **347,** 575–578.
16. Hope, I. A. & Struhl, K. (1986) *Cell* **46,** 885–894.
17. Oliphant, A. R., Nussbaum, A. L. & Struhl, K. (1986) *Gene* **44,** 177–183.
18. Hope, I. A. & Struhl, K. (1985) *Cell* **43,** 177–188.