3

GCN4, the Yeast Version of the *jun-fos* Oncogene Family

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

In the yeast *Saccharomyces cerevisiae*, the 50 to 100 genes involved in amino acid biosynthesis are regulated coordinately. Under normal growth conditions, these genes are transcribed at a basal level of 1 to 2 mRNA molecules per cell. However, when yeast cells are starved for amino acids, they induce the transcription of all the coregulated genes by a factor of 3 to 5. The starvation response can be elicited by metabolic poisons or mutations that inhibit the synthesis of any single amino acid or the level of tRNA charging. This general control mechanism differs from the situation in *E. coli* where regulation occurs in individual biosynthetic pathways.

General control is best viewed as a mechanism to regulate protein synthesis by controlling the amount of amino acid precursors. This regulatory mechanism is important for controlling cell growth because yeast cells initiate new cell division cycles only if they have sufficient amino acids to complete the cycle; amino acid auxotrophs arrest at the start of the cycle upon being switched to medium lacking the required amino acid. Moreover, mutant strains defective in regulating transcription of the amino acid biosynthetic genes also arrest at the start of the cell cycle when grown at high temperatures. Thus, transcriptional regulation of amino acid biosynthetic genes represents part of a more global mechanism that regulates cell growth.

DNA-Binding and Transcriptional Activation Regions of the GCN4 Protein

Transcriptional induction of the amino acid biosynthetic genes is mediated by GCN4, a protein that binds specifically to the promoter regions of these genes.[1] GCN4 binds as a dimer[2] to target sites whose consensus is the 9-bp palindrome ATGA(C/G)TCAT. This consensus sequence also represents the optimal GCN4 binding site as determined by saturation mutagenesis[3] and by selection of binding sites from random-sequence DNA.[4] The 60 C-terminal amino acids are sufficient both for specific DNA-binding[5] and for dimerization.[6]

Although the GCN4 DNA-binding domain is necessary for recognizing the appropriate promoters, a short acidic region in the center of the protein is required for transcriptional activation.[5] Derivatives containing only the DNA-binding domain do not activate transcription *in vivo* and indeed can actually repress transcription in certain promoters. The transcriptional activation region of GCN4 stimulates transcription when fused to a heterologous DNA-binding domain, the *E. coli* LexA repressor. The resulting LexA hybrid proteins activate transcription of promoters that contain LexA binding sites as upstream elements.

Derivatives retaining only 35 to 40 amino acids from this acidic region are sufficient for wild-type levels of transcriptional activation when fused directly to the GCN4 DNA-binding domain.[5,6] Moreover, the distance and orientation of the activation region with respect to the DNA-binding domain is functionally unimportant. This indicates that the activation region is an autonomous function and that there is no requirement for a spacer between the activation region and the DNA-binding domain. These observations, along with those from the laboratory of Mark Ptashne, Harvard University, strongly suggest that transcriptional activation regions are not defined by a specific primary sequence but rather by a more general structural feature presumably involving net negative charge.

Progressive deletion of the GCN4 transcriptional activation region did not reveal a position where there was a sudden complete loss in activity but rather a series of small, stepwise reductions in activity.[6] GCN4 activity appears directly related to the size of the

transcription activation region remaining. In no case did a shorter region activate transcription more efficiently than a longer region. In contrast, there is no such precise relationship of transcriptional activity to the number of acidic residues. In particular, there are several examples in which derivatives with fewer acidic residues activate transcription better than derivatives with more acidic residues. The strong correlation between the length of the GCN4 activation region and the level of transcriptional activity is strongly suggestive of a repeating structure consisting of units that act additively.

Jun Oncoprotein Binds the Same Sequences as GCN4 and Activates Transcription in Yeast

The *jun* oncogene was originally derived from a defective avian sarcoma virus that causes fibrosarcomas in chickens and transforms chick embryo fibroblasts in cell culture. The amino acid sequence of the C-terminus of Jun protein is 45 percent identical to the GCN4 DNA-binding domain, including a 30 amino acid region in which there are 17 identical residues and four conservative differences. To investigate whether the Jun oncoprotein and GCN4 bind to the same DNA sequences, molecules were constructed in which the region encoding the GCN4 DNA-binding domain was replaced by the homologous Jun region.[7]

A protein containing the C-terminal 112 amino acids of Jun in place of the GCN4 DNA-binding domain functionally substitutes for GCN4 in its ability to induce the expression of *HIS3* and other amino acid biosynthetic genes. A related protein containing only 99 C-terminal amino acids of Jun also substitutes for GCN4 but with slightly less efficiency. The Jun DNA-binding domain activates *HIS3* transcription only if the promoter region contains a functional GCN4 binding site. Moreover, maximal *HIS3* induction mediated by the Jun hybrid protein occurs in combination with the optimal GCN4 binding site. Thus, the GCN4 and Jun DNA-binding regions behave homologously on a set of target sites, thereby indicating that they recognize very similar DNA sequences by a common structural motif.

The Jun proteins just described contain the intact GCN4 activa-

tion region and thus do not address the question of whether Jun itself can activate transcription in yeast. However, a LexA-Jun hybrid protein in which the entire *jun* coding region is fused directly to the LexA DNA-binding domain can functionally replace GCN4.[8] In addition, LexA-Jun activates transcription through the heterologous LexA DNA-binding domain at a level comparable to that achieved by LexA-GCN4-Jun or LexA-GCN4. These observations indicate that the *jun* oncogene contains a sequence(s) that functions as an efficient transcriptional activation region in yeast.

The Jun protein contains a region between residues 15 and 59 with a net negative charge of -7 and a region between residues 87 and 102 with a net charge of -4. Deletions that remove more than 100 N-terminal residues of *jun* lack both acidic regions and confer extremely low levels of activation. Deletions with end points between residues 54 and 71, which remove one of the acidic regions, show a two to fivefold decrease in the level of expression. Thus, as is the case for the yeast GCN4 and GAL4 activator proteins, the acidic regions of *jun* appear to be important for transcriptional activation in yeast.

The similar DNA-binding properties of GCN4 to both the mammalian transcription factor AP-1 and to the Jun oncoprotein led to demonstration that Jun represents an oncogenic version of a normal cellular transcription factor. The fact that Jun activates transcription in yeast indicates that this oncoprotein can interact functionally with the basic transcription machinery of yeast. The obvious implications are that the basic transcription machineries of eukaryotic organisms from yeast to man are evolutionarily conserved and that mRNA transcriptional initiation in all eukaryotes may occur by a common molecular mechanism.

Converting the Dimerization Specificity of Fos to that of GCN4 by Swapping Leucine Zippers

GCN4, Jun, and the Fos oncoprotein are members of a new class of DNA-binding proteins defined by a structural motif called the "leucine zipper." These proteins all contain a region with four or five

leucine residues spaced exactly seven residues apart whose sequence is consistent with the formation of an amphipathic α-helix. It has been proposed that the leucine zipper consists of two interdigitated α-helices, one from each monomer subunit, that constitute the dimerization function necessary for high affinity binding to DNA; an adjacent region of basic residues is thought to be responsible for specific protein-DNA contacts. However, despite the conserved leucine residues, each protein has a distinct dimerization specificity.

Jun and Fos form a heterodimeric complex that binds to the AP-1 transcriptional regulatory element whose sequence is identical to that bound by GCN4 homodimers. Because the optimal GCN4 binding site is dyad-symmetric such that each GCN4 monomer directly contacts a half-site, by analogy, each subunit of a Fos-Jun heterodimer would be expected to contact a half-site. In this view, the failure of Fos to bind DNA could reflect its inability to dimerize; however, its contribution to specific DNA-binding could also be by an indirect effect mediated through Jun.

To determine if the leucine zipper is sufficient for dimerization specificity and to establish if Fos has the inherent ability to contact DNA, we created a chimeric protein in which the leucine zipper of Fos was precisely replaced by the C-terminal amino acids of GCN4 that contain the leucine zipper.[9] In the resulting protein, the spacing of the leucine zipper and basic domain is identical to that in either Fos or GCN4. This chimeric protein binds with high affinity to the AP-1 site but not to DNA containing a point mutation that significantly reduces binding by GCN4 or by the Jun-Fos complex, thus indicating that the protein has similar DNA sequence recognition properties. More important, the chimeric protein forms DNA-binding heterodimers with GCN4 but not with either Jun or Fos. Thus, the leucine zipper is sufficient to confer dimerization specificity because the chimeric protein contains only the leucine zipper region of GCN4, yet unlike Fos, the chimeric protein efficiently forms heterodimers with GCN4 but not with Jun. Moreover, these observations strongly suggest that the leucine zipper is sufficient for dimerization *per se*.

The leucine zipper motif was initially defined by the presence of four to five leucine residues spaced seven residues apart in a region of

the protein that is permissive for α-helix formation; beyond these properties, the overall amino acid sequence similarity is unimpressive. The fact that swapping the Fos and GCN4 leucine zippers yields a specific DNA-binding protein indicates that these regions are functionally homologous. Nevertheless, it is clear that although the conserved leucines are important for the dimerization, other nonconserved residues in the various zipper regions must be involved in dimerization specificity. Thus, the ability of this class of proteins to form homodimers or heterodimers will depend on the association properties of individual leucine zipper regions.

The Fos oncoprotein clearly has an inherent specific DNA-binding activity because the chimeric protein specifically interacts with the AP-1 site; the GCN4 leucine zipper region alone does not bind DNA. Further, because the GCN4 zipper converts Fos into a Jun-independent DNA-binding protein, the failure of Fos to bind DNA almost certainly reflects its inability to form homodimers. Thus, the Fos and Jun monomer subunits in the heteromeric complex must interact with adjacent half-sites.

Acknowledgments

This work was supported by grant GM30186 from the National Institutes of Health.

References

1. I.A. Hope and K. Struhl, *Cell* **43**, 177 (1985).
2. I.A. Hope and K. Struhl, *EMBO J.* **6**, 2781 (1987).
3. D.E. Hill *et al., Science* **234**, 451 (1986).
4. A.R. Oliphant, C.J. Brandl, and K. Struhl, *Mol. Cell Biol.* **9**, 2944 (1989).
5. I.A. Hope and K. Struhl, *Cell* **46**, 885 (1986).
6. I.A. Hope, S. Mahadevan, and K. Struhl, *Nature (London)* **333**, 635 (1988).
7. K. Struhl, *Cell* **50**, 841 (1987).
8. K. Struhl, *Nature (London)* **332**, 649 (1988).
9. J.W. Sellers and K. Struhl, *Nature (London)* **341**, 74 (1989).