It has recently been shown that a gene required for the differentiation of the touch receptor neurons in *C. elegans* also contains a homoeobox[37]. Moreover, several homoeobox-containing genes cloned in vertebrates seem to be primarily expressed in the nervous system[38-42]. These observations indicate that the relevance of homoeobox-containing genes in specifying cell fate in the nervous system may not be confined to *Drosophila*.

We thank Drs T. Kornberg and T. Hsieh for cDNA libraries, Dr S. Benzer for Mab21A6, Drs C. O'Kane and W. Gehring for transformant line 37a, Drs A. Ghysen and C. O'Kane for results prior to publication, Dr T. Schwarz, S. Shepherd, Dr C. Turck and M. Doherty for technical assistance, and Drs W. Bender, A. Ghysen and P. 'OFarrell for their comments. This study was supported by the Howard Hughes Medical Institute and the NIH. K.B. had a fellowship from the Fonds Nationals Suisse, L.Y.J. and Y.N.J. are Howard Hughes investigators.

Received 5 April; accepted 22 April 1988.

1. Bate, C. M. in *Handbook of Sensory Physiology* Vo. IX (ed. Jacobson, M.) 1-53 (Springer, Berlin, Heidelberg, New York, 1978).
2. Campos-Ortega, J. A. & Hartenstein, V. in *The Embryonic Development of Drosophila melanogaster* (Springer, Berlin, Heidelberg, New York, 1985).
3. Dambly-Chaudiere, C. & Ghysen, A. *Wilhelm Roux Arch. dev. Biol.* **195,** 222-228 (1986).
4. Ghysen, A., Dambly-Chaudiere, C., Aceves, E., Jan, L. Y. & Jan, Y. N. *Wilhelm Roux Arch. dev. Biol.* **195,** 281-289 (1986).
5. Bodmer, R. & Jan, Y. N. *Wilhelm Roux Arch. dev. Biol.* **196,** 69-77 (1987).
6. Hertweck, H. *Z. wiss. Zool.* **139,** 559-663 (1931).
7. Zacharuk, R. Y. in *Comprehensive Insect Physiology, Biochemistry and Pharmacology* Vol. 6 (eds Gilbert, L. I. & Kerkut, G. A.) 1-69 (Pergamon, New York, 1985).
8. McIver, S. B. in *Comprehensive Insect Physiology, Biochemistry and Pharmacology* Vol. 6 (eds Gilbert, L. I. & Kerkut, G. A.) 71-132 (Pergamon, New York, 1985).
9. Hartenstein, V. *Development* (in the press).
10. Bodmer, R. *et al. Cell* **51,** 293-307 (1987).
11. Gehring, W. J. *Science* **236,** 1245-1252 (1987).
12. Scott, M. P. & Carroll, S. B. *Cell* **51,** 689-698 (1987).
13. Desplan, C., Theis, J. & O'Farrell, P. H. *Nature* **318,** 630-635 (1985).
14. Hoey, T. & Levine, M. *Nature* (in the press).
15. Johnson, T. K. & Judd, B. H. *Genetics* **92,** 485-502 (1979).
16. Jack, J. W. *Cell* **42,** 869-876 (1985).
17. Cavener, D. R. *Nucleic Acids Res.* **15,** 1353-1361 (1987).
18. Patel, R. S., Odermatt, E., Schwarzbauer, J. E. & Hynes, R. O. *EMBO J.* **6,** 2565-2572 (1987).
19. Kornberg, T., Siden, I., O'Farrell, P. H. & Simon, M. *Cell* **40,** 45-53 (1985).
20. Regulski, M., McGinnis, N., Chadwick, R. & McGinnis, W. *EMBO J.* **6,** 767-777 (1987).
21. Kassis, J. A., Poole, S. J., Wright, D. K. & O'Farrell, P. H. *EMBO J.* **5,** 3583-3589 (1986).
22. Guarente, L. *Cell* **52,** 303-305 (1988).
23. Zipursky, S. L., Ventakesh, T. R., Teplow, D. B. & Benzer, S. *Cell* **36,** 15-26 (1984).
24. Pollack, J. A. *Calif. Inst. Tech. Biol. A. Rep.* **203,** 203 (1985).
25. Jan, Y. N., Ghysen, A., Christoph, I., Barbel, S. & Jan, L. Y. *J. Neurosci.* **5,** 2453-2464 (1985).
26. Harris, W. A., Stark, W. S. & Walker, J. A. *J. Physiol., Lond.* **256,** 415-439 (1976).
27. Campos-Ortega, J. A., Jurgens, G. & Hofbauer, A. *Wilhelm Roux Arch. EntwMech Org.* **186,** 27-50 (1979).
28. Tomlinson, A. & Ready, D. F. *Devl Biol.* **123,** 264-275 (1987).
29. Hafen, E., Basler, K., Edstroem, J. E. & Rubin, G. M. *Science* **236,** 55-63 (1987).
30. Banerjee, U., Renfranz, P. J., Pollock, J. A. & Benzer, S. *Cell* **49,** 281-291 (1987).
31. Tomlinson, A., Bowtell, D. D. L., Hafen, E. & Rubin, G. M. *Cell* **51,** 143-150 (1987).
32. Banerjee, U., Renfranz, P. J., Hinton, D. R., Rabin, B. A. & Benzer, S. *Cell* **51,** 151-158 (1987).
33. DiNardo, S., Kuner, J. M., Theis, J. & O'Farrell, P. H. *Cell* **43,** 59-69 (1985).
34. Garcio-Bellido, A. *Am. Zool.* **17,** 613-629 (1977).
35. Scott, M. P. & O'Farrell, P. H. *A. Rev. Cell Biol.* **2,** 49-80 (1986).
36. Doe, C. Q., Hiromi, Y., Gehring, W. J. & Goodman, C. S. *Science* **239,** 170-175 (1988).
37. Way, J. & Chalfie, M. *Cell* (in the press).
38. Awgulewitsch, A., Utset, M. F., Hart, C. P., McGinnis, W. & Ruddle, F. H. *Nature* **320,** 328-335 (1986).
39. Simeone, A. *et al. Nature* **320,** 763-765 (1986).
40. Carrasco, A. E. & Malacinski, G. M. *Devl Biol.* **121,** 69-81 (1987).
41. Joyner, A. L. & Martin, G. R. *Genes Dev.* **1,** 29-38 (1987).
42. Odenwald, W. F. *Genes Dev.* **1,** 482-496 (1987).
43. Poole, S. J., Kauvar, L. M., Drees, B. & Kornberg, T. *Cell* **40,** 37-43 (1985).
44. Nolan, J. M., Maxwell, P. L., Wyckoff, E. & Hsieh, T. *Proc. natn. Acad. Sci. U.S.A.* **83,** 3664-3668 (1986).
45. Papazian, D. M., Schwarz, T. L., Tempel, B. L., Jan, Y. N. & Jan, L. Y. *Science* **237,** 749-757 (1987).
46. Schwarz, T. L., Tempel, B. L., Papazian, D. M., Jan, Y. N. & Jan, L. Y. *Nature* **331,** 137-142 (1988).
47. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74,** 5463-5467 (1977).
48. Schneuwly, S., Kuroiwa, A., Baumgartner, P. & Gehring, W. J. *EMBO J.* **5,** 733-739 (1986).
49. Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. & Noll, M. *Cell* **47,** 1033-1040 (1986).
50. Newman, G. R., Jasani, B. & Williams, E. D. *Microscopy* **132,** RP1-RP2 (1983).
51. O'Kane, C. J. & Gehring W. J. *Proc. natn. Acad. Sci. U.S.A.* **84,** 9123-9127 (1987).

Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein

Ian A. Hope*, Subramony Mahadevan & Kevin Struhl

Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115, USA

Derivatives of the yeast GCN4 transcription factor containing acidic regions of 35 to 40 amino acids fused directly to the DNA-binding domain are fully functional in vivo. *High resolution deletion analysis and proteolytic mapping suggest that the activation region is a repeated structure composed of small units acting additively. Acidic character is a feature of the structural motif, possibly a dimer of α-helices from two GCN4 monomers, that may be important for interactions with the basic transcriptional machinery.*

INITIATION and regulation of eukaryotic messenger RNA synthesis by RNA polymerase II requires transcriptional activator proteins that interact with specific promoter DNA sequences[1-3]. The DNA-binding functions of these activator proteins are localized to autonomous domains that contain less than 100 amino-acid residues[4-8]. Analysis of yeast GCN4 (ref 5) and GAL4 (refs 4, 9) proteins indicates that DNA-binding and transcriptional activation are independent functions carried out by distinct regions of the protein.

Surprisingly, the GCN4 and GAL4 transcriptional activation functions are performed by relatively short regions of acidic character that do not have rigid sequence requirements[2,5,10]. Different portions of the GCN4 acidic region are equally capable of activating transcription, and there is no sequence homology between the GCN4 and GAL4 activation regions. Furthermore, acidity is the only obvious common characteristic of transcriptional activation regions selected from random segments of *Escherichia coli* DNA[11], and mutations of a GAL4 derivative that increase or decrease activation usually increase or decrease negative charge, respectively[12]. In addition, the jun oncoprotein, a vertebrate transcription factor[13,14], activates transcription in yeast through an acidic region[15].

Although a growing body of evidence indicates the importance of acidic residues, little is known about the important structural characteristics of transcriptional-activation regions. This lack of understanding is due partly to the fact that no single activation region has been analysed in detail. Here, we investigate the

* Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

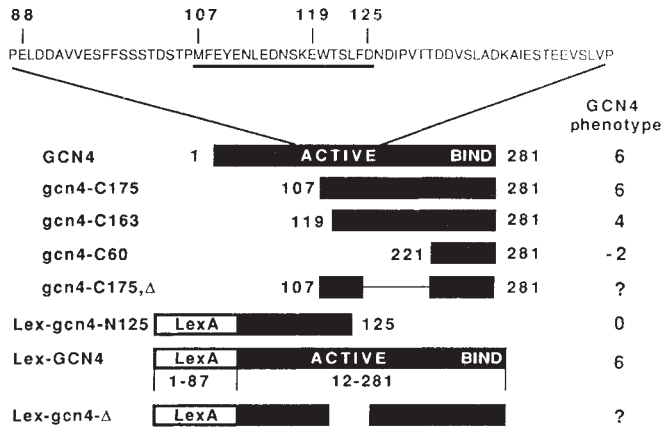

Fig. 1 Structural and functional analysis of GCN4 and LexA-GCN4 derivatives. The structures of various GCN4 and LexA-GCN4 proteins are indicated by black bars with the N- and C-terminal residues defined as in the wild-type proteins. The wild-type GCN4 protein contains 281 amino acids and consists of a fully functional and dimeric DNA-binding domain (BIND) located within the 60 C-terminal amino acids[5,20], and a centrally located transcriptional activation region (ACTIVE)[4] whose primary sequence (residues 87–152) is shown above. The 19 amino-acid segment implicated as being functionally important from the phenotypes of gcn4-C175 and LexA-N125 is underlined. The GCN4 phenotypes are shown in an arbitrary scale, defined in the methods and used in the Tables, where 6 indicates wild-type activity, 4 indicates partial activity as observed for gcn4-C163, 0 indicates no GCN4, and −2 indicates the repression phenotype as observed for gcn4-C60 (ref. 5).

Methods. The plasmids used to produce GCN4 and LexA-GCN4 proteins in yeast all derive from YCp88-GCN4 and YCp88-LexA-GCN4 (ref. 5), and contain the appropriate protein-coding region downstream of the constitutive *DED1* promoter in yeast/*E. coli* shuttle vector maintained at 1–2 copies per yeast cell. The GCN4 coding regions in YCp88-gcn4-C175 and YCp88-LexA-GCN4 were modified to generate the internally deleted GCN4 derivatives whose structures and phenotypes are shown in Tables 1–3 and Fig. 2. For the experiment described in Table 1, YCp88-gcn4-C175 DNA was cleaved with *Kpn*I (corresponding to amino-acid residue 151), treated to various extents with *Bal*31 nuclease and circularized. Upon introduction into yeast, derivatives conferring GCN4 function were selected by growth in the presence of 10 mM aminotriazole (see below) and subjected to DNA-sequence analysis. For the remainder of the experiments, appropriate DNAs were treated with *Bal*31 nuclease, ligated to an 8 base pair *Sal*I linker, and cloned into M13 derivatives. The deletion end-points were determined by DNA sequence analysis, and appropriate pairs of derivatives were combined to generate deletion mutants. The YCp88 DNAs were introduced into yeast strain KY803 (ref. 5), which is completely deleted for the GCN4 coding region, or KY330, an isogenic strain containing a plasmid which carries the *lacZ* gene downstream of a LexA binding site to permit measurement of transcriptional activation through the LexA DNA-binding domain. Levels of transcriptional activation through the GCN4 DNA-binding domain (total GCN4 activity) were determined by comparing growth rates on minimal plates supplemented with 20, 10 or 5 mM aminotriazole (AT) as described previously[5]. Growth in the presence of AT, which acts on the *HIS3* gene product to inhibit histidine biosynthesis, depends on active GCN4 to induce *HIS3* transcription, and the relative degree of AT resistance is directly related to the level of *HIS3* transcription[25]. To ensure reproducibility, all strains were tested on plates that were made at the same time, each derivative was tested on at least two indpendent occasions, several independent transformants for each derivative were tested, and cell growth was monitored by at least two individuals. The phenotypic classes are defined as follows: 6, growth at wild-type rates in 20 mM AT; 5, reduced growth in 20 mM AT, but wild-type growth in 10 mM AT; 4, slow growth in 10 mM, but wild-type growth in 5 mM AT; 3, slow growth in 5 mM AT, but normal growth in the absence of AT; 2 very slow growth in 5 mM AT, but normal growth in the absence of AT; 1 no growth in 5 mM AT, but nearly normal wild-type growth in the absence of AT. A phenotype of 0 is defined as identical to strains containing a *GCN4* deletion mutation; such strains grow slowly in the absence of AT. GCN4 derivatives with a functional DNA-binding domain but lacking the transcriptional-activation regions repress transcription below the normal basal level; the growth of such cells is strongly inhibited on minimal plates without extra amino-acid supplements[5]. Phenotypes −1 and −2 represent mild or severe repression respectively. Transcriptional activation through the LexA DNA-binding domain was measured by standard plate and liquid culture assays of β-galactosidase activities.

structure and function of the GCN4 activation region by systematic and high resolution deletion analysis and by proteolysis of wild-type and deleted versions of the GCN4 protein.

Minimal activation region

The initial experiment was designed to determine the minimal region sufficient for wild-type levels of transcriptional activation. Previously, gcn4-C175, the C-terminal 175 amino acids of GCN4, was shown to be the derivative with most deletions that retained full activity[5]. In the present experiment varying deletions within the gcn4-C175 coding region were made by restriction cleavage between the sequences coding for the activation region and the DNA-binding domain, followed by treatment with *Bal*31 nuclease. Derivatives still capable of activating transcription were identified by subjecting the collection of internal deletion mutants to a genetic selection (Fig. 1, Table 1).

The smallest fully functional derivative, Δ11, consists of 36 amino acids from the acidic region directly fused to a DNA-binding region containing 89 amino acids (Table 1). This indicates that a large, flexible spacer between the transcription activation region and the DNA-binding domain is not required. The minimal transcriptional activation region necessary for full function appears to be just 32 amino acids because the phenotype of a derivative containing only the N-terminal amino acids, Δ12, is indistinguishable from the wild-type protein, whereas all deletions that extend further towards the N-terminus of gcn4-C175 show reduced transcriptional activation. Interestingly, the level of activation correlates very well with the length of the activation region but poorly with the overall size of the deletion or with the size of DNA-binding domain. The sole exception, Δ16, may have less GCN4 activity because the deletion encroaches furthest into the DNA-binding domain. The shortest N-terminal segment with any detectable activity was the 25 amino-acid segment of Δ17 and Δ18, suggesting that segments shorter than 25 amino acids are insufficient for transcriptional activation.

Functional redundancy

Previous N- and C-terminal deletions indicated either that the 19 amino-acid segment between residues 107 and 125 is sufficient

Table 1 Transcriptional activation by internal deletions of gcn4-C175 selected for GCN4 function

GCN4 derivative	GCN4 amino acids present		Transcriptional activity
	N-terminus	C-terminus	
gcn4-C175, Δ11	107–142	193–281	6
gcn4-C175, Δ12	107–138	162–281	6
gcn4-C175, Δ13	107–132	164–281	5
gcn4-C175, Δ14	107–134	165–281	5
gcn4-C175, Δ15	107–132	176–281	3
gcn4-C175, Δ16	107–137	201–281	3
gcn4-C175, Δ17	107–131	200–281	2
gcn4-C175, Δ18	107–131	170–281	2

The GCN4 amino acids which contribute to the N- and C-terminal segments are given for each derivative. The GCN4 activities are indicated on an arbitrary scale defined as in Fig. 1, where 6 indicates wild-type levels of activation, 0 indicates no activation, and −2 indicates repression.

	116 123 130 137 144	Amino acids	Net charge	GCN4 activity
	- - - -| +- | - - | - - | -+ -| - -			
Δ35	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAIESTEEVSL	wt	-11	
Δ36	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAIESTEEgstk	41	-10	
Δ37	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAIESTEgstk	40	-9	4
Δ38	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAIESTEvdq	40	-11	
Δ39	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAIESgrpvk	38	-7	
Δ40	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAvdq	35	-9	
Δ41	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAgrpvk	35	-6	
Δ42	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKArstk	35	-6	3
Δ43	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLAgstk	32	-7	
Δ44	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLgstk	31	-7	
Δ45	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSgrpvk	30	-6	2
Δ46	MFEYENLEDNSKEWTSLFDNDIPgstk	23	-5	
Δ47	MFEYENLEDNSKEWTSLFDNgrpvk	20	-3	
Δ48	MFEYENLEDNSKEWTSLFDrstk	19	-3	1
Δ49	MFEYENLEDNSKEWTSLFDgrpvk	19	-3	
Δ50	MFEYENLEDNSKEWTSLFgstk	18	-3	0
Δ51	MFEYENLEDNSKEWTSLwstk	17	-3	
Δ52	MFEYENLEDNSKEWTgrpvk	15	-2	
Δ53	MFEYENLEDNSKEWgrpvk	14	-2	
Δ54	MFEYENLEDNSKgstk	12	-2	-1
Δ55	MFEYENLEDNSvdq	11	-5	
Δ56	MFEYENLEDgrpvk	9	-2	
Δ57	MFEYENLEDrstk	9	-2	
Δ58	MFEYENLEgstk	8	-2	
Δ59	MFEYENLvdq	7	-3	-2
Δ60	MFEgrpvk	3	+1	

Fig. 2 High-resolution deletion analysis. For each derivative, the indicated amino-acid-sequence of the activation region was fused to a 100-residue GCN4 DNA-binding domain (at position 181). Residues shown in capital letters derive from the native GCN4 protein, whereas residues in small letters derive from the *Sal*I linker and joint sequences. The number of GCN4 amino acids (excluding linker sequences), the net negative charge, and the level of GCN4 activity (determined as described in Fig. 1) are shown. Shown above the amino-acid sequences are the location of negatively and positively charged residues, and the probable boundaries that distinguish the phenotypic classes.

for the transcriptional activation or that GCN4 contains redundant activities[5]. As the data in Table 1 imply that the 19 amino-acid segment is insufficient for activation, it appears that residues towards either the N-terminus (as in LexA-gcn4-N125) or towards the C-terminus (as in gcn4-C175) contribute to a functional transcriptional-activation region. This redundancy was investigated by constructing internal deletions around and including the 19 amino-acid segment in LexA-GCN4 (Fig. 1), a bifunctional activator protein that can stimulate transcription through both the LexA and GCN4 DNA-binding domains[5]. Unlike the previous experiment in which functional derivatives were selected from a population, the derivatives analysed in Table 2 were generated by systematically fusing a series of N-terminal segments of LexA-GCN4 to a series of C-terminal segments. The resulting matrix of deletions were phenotypically analysed without being previously selected for GCN4 function.

The primary observation is that numerous deletions throughout the acidic region (amino-acid residues 88–147) still permit transcriptional activation by both DNA-binding domains. Many derivatives remove part or all of the 19 amino-acid segment, yet still show full or partial transcriptional activation. Thus, even though the previous analysis attached critical importance to this region, analysis of internal deletions clearly reveals that this 19 residue segment is neither sufficient nor essential for transcriptional activation activity. Although these results could be complicated slight variations in the stability of derivatives *in vivo* (for which there are no data as GCN4 could not be detected biochemically), it is remarkable that so many derivatives are functional.

The level of transcriptional activation correlates fairly well with the amount of the acidic region that remains, but poorly to the particular regions that are retained. The exceptions are fusions to residue 133 (Δ25–Δ28), which have anomalously low activity. In addition, the region between residues 12 and 92 can impart some transcriptional activity in the absence of the primary acidic activation region (compare Δ23, Δ30, Δ34 with Δ22 and Δ29). This may be due to the short region between residues 50 and 70 that has the same proportion of acidic residues as the primary activation region.

Table 2 Transcriptional activation by both the LexA and GCN4 components of a series of internal deletions of LexA-GCN4

GCN4 derivative	N-terminus LexA	N-terminus GCN4	C-terminus GCN4	GCN4 Activity	β-galactosidase Plate	β-galactosidase Liquid
LexA-gcn4-Δ19	1–87	12–14	85–281	6	++	350
LexA-gcn4-Δ20	1–87	12–14	103–281	6	++	440
LexA-gcn4-Δ21	1–87	12–92	103–281	6	++	450
LexA-gcn4-Δ22	1–87	12–14	124–281	−2	−	6
LexA-gcn4-Δ23	1–87	12–92	124–281	2	+	85
LexA-gcn4-Δ24	1–87	12–117	124–281	5	++	240
LexA-gcn4-Δ25	1–87	12–14	133–281	0	−	2
LexA-gcn4-Δ26	1–87	12–92	133–281	0	−	3
LexA-gcn4-Δ27	1–87	12–117	133–281	0	+	40
LexA-gcn4-Δ28	1–87	12–129	133–281	4	++	170
LexA-gcn4-Δ29	1–87	12–14	141–281	−2	−	8
LexA-gcn4-Δ30	1–87	12–92	141–281	2	+	30
LexA-gcn4-Δ31	1–87	12–117	141–281	4	++	130
LexA-gcn4-Δ32	1–87	12–129	141–281	5	++	400
LexA-gcn4-Δ33	1–87	12–138	141–281	6	++	320
LexA-gcn4-Δ34	1–87	12–95	144–281	2	+	40
LexA-GCN4	1–87	12–281		6	++	300

The amino-acid composition of each derivative is indicated in order from the N- and C-termini. The N- and C-terminal segments are joined by a *Sal*I linker that encodes an additional arginine and proline residue. GCN4 activities are indicated in the arbitrary scale defined in Fig. 1 where 6 indicates wild-type levels of activation, 0 indicates no activation and −2 indicates repression. LexA activities were determined by measuring β-galactoidase activity produced from a *lacZ* gene downstream of a LexA binding site (see Fig. 1). For the β-galactosidase plate assay, hydrolysis of X-Gal gave either a strong blue (++), weak blue (+), or no blue (−) colour to yeast colonies. β-galactosidase activities for the liquid assay are defined as in previous work[5].

High resolution deletions

To investigate the nature of the transcriptional-activation region in more detail, we generated a large and sequential series of very fine scale internal deletions that could be easily related to each other. Constant C-terminal segments, containing the DNA-binding domain, were fused to a series of N-terminal segments ranging in size from 3 to 41 residues (Fig. 2). Although all the derivatives contain the 100 C-terminal amino acids, note that there are three classes of fusions representing the three possible reading frames; this affects the amino-acid sequence at the junction between the activation peptide and the DNA-binding domain. Thus, derivatives that contain the same number of GCN4 residues from the activation region can have a different number of acidic residues.

The phenotypes of the resulting derivatives fit a simple and striking pattern. Unlike structural elements such as active sites or domains in a protein, progressive deletion did not reveal a position where there was a sudden complete loss in activity. Rather, a series of small, step-wise reductions in activity was observed. Seven discrete phenotypes could be readily distinguished, ranging from high GCN4 activity for the largest derivatives to a transcriptional-repression phenotype when only the GCN4 DNA-binding domain is present. The range of activities observed and the lack of any derivatives which failed to fit into the simple pattern argue against any complications, such as effects on protein stability. Therefore, the different levels of activity observed almost certainly reflect differences in the amount of transcriptional-activation function retained by each derivative.

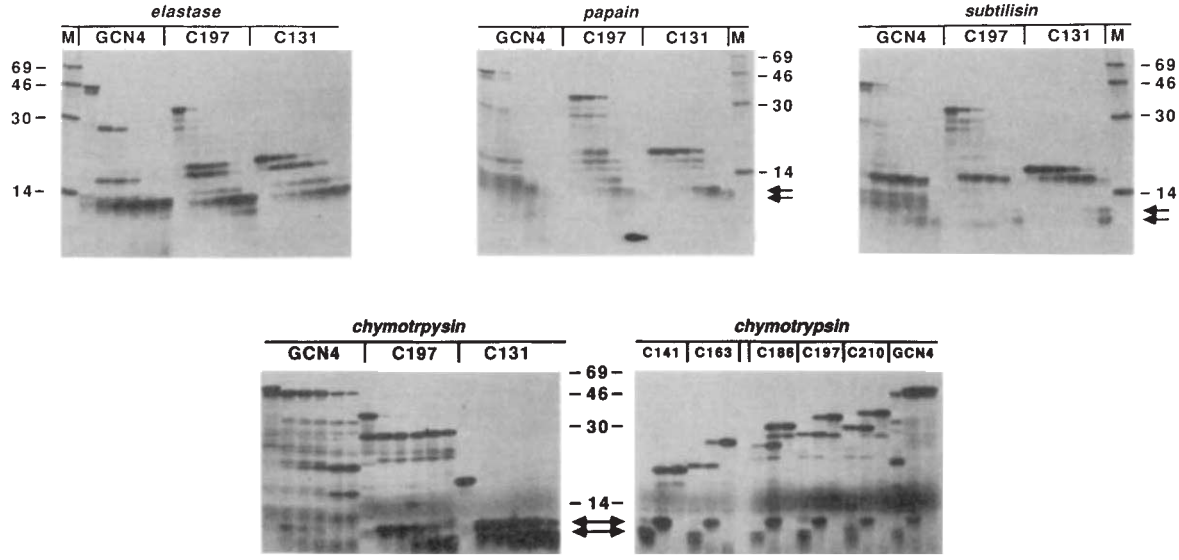

Fig. 3 Protease digestion of GCN4 and deleted derivatives. For all the panels except for the bottom right (chymotryspin), the indicated derivative was treated with 0, 1×, 2×, 4×, 8× or 16× units of protease, where × equals 6.25×10^{-4} for elastase, 6.25×10^{-5} for papain, 10^{-5} for subtilisin, and 10^{-4} for chymotrypsin. For the chymotrypsin experiment shown on the far right, each derivative was treated with 0, 1× or 10× units of chymotrypsin. The sizes of the cleavage products described below are approximate and do not always add up to the size of the initial substrates probably because of the very anomolous electrophoretic migration of GCN4 and its derivatives (GCN4 migrates at a position corresponding to a relative molecular mass (M_r) of 45,000(45K) even though its true M_r is only 31K)[16]. Elastase initially cleaved GCN4 to a 16 K C-terminal segment present for all three derivatives, and an N-terminal segment of 26 K. The former was cleaved to a more stable 13.5 K segment. The C-terminal segment appeared less resistant to treatment with papain than to other proteases. C-terminal segments of greater stability than those derived from the N-termini could be discerned at 16 K, however, and then at 12.5 K. Note that *N*-terminal fragments would be expected to appear more intense than C-terminal fragments on the autoradiogram because of the distribution of the ^{35}S-methionine residues (three are located at the extreme N-terminus at residues 1, 14 and 38, one is centrally located within the acidic region at 107, and one is located in the DNA-binding domain at residue 250). With subtilisin, the N-terminal segment did not appear as a single discrete fragment, whereas a stable C-terminal segment of 16 K was observed for all derivatives. For chymotrypsin, GCN4 was cleaved to 42 K N-terminal segment and a 12K C-terminal segment which was cleaved further to 11 K. The arrows in figures correspond to bands representing the stable proteolytic fragments containing the GCN4 DNA-binding domain.

Methods. GCN4 and derivatives were generated by transcription and translation *in vitro* as previously described[16]. All protease digestions were of 0.5 μl ~10^4 acid precipitable c.p.m. of ^{35}S-methionine-labelled protein) of *in vitro* translation products in a final volume of 10 μl. Reaction conditions were as follows: 10 mM Tris *p*H 8.8, 37 °C for elastase; 10 mM Tris *p*H 6.8, 25 °C; for papain; 10 mM Tris *p*H 7.5, 37 °C for subtilisin; 10 mM Tris *p*H 7.9, 37 °C for chymotrypsin. After 45 min incubation, reactions were stopped by adding Laemmli sample buffer and heating to 100 °C for 5 min before applying to a 15 % SDS polyacrylamide gel for electrophoresis. The ^{35}S-methionine-labelled protein fragments were revealed by autoradiography.

GCN4 activity appears directly related to the size of the transcriptional-activation region remaining. There is not a single case where a shorter region activates transcription more efficiently than a longer region. In contrast, there is no such precise relationship of transcriptional activity to the number of acidic residues. First, the boundaries that distinguish the phenotypic classes do not always correspond with the presence or absence of particular acidic residues. Second, there are several examples in which derivatives with fewer acidic residues activate transcription better than derivatives with more acidic residues. Third, there are several examples in which derivatives with an equal number of acidic residues fall into different phenotypic classes.

Spacing requirements

Results presented here and elsewhere[5,10] suggest that the distance and orientation of the activation region with respect to the DNA-binding domain is functionally unimportant. To examine this more explicitly, several of the short N-terminal fragments described in the previous section were fused to a series of C-terminal segments of different sizes. The result was that for each N-terminal segment, fusion to C-terminal segments ranging from 61 to 130 amino acids produced similar phenotypes. This provides direct evidence that the activation region encodes an autonomous function that is equally active on a series of DNA-binding domains. In addition, it indicates that the distance between the activation region and DNA-binding domain is unimportant even at short distances, so that a spacer between these two functions is not necessary (although the requirement for a very small spacer cannot be excluded).

One apparent anomaly is that all derivatives containing DNA-binding domains less than 110 amino acids show reduced levels of transcriptional activation when compared to the corresponding derivatives wth DNA-binding domains of 130 residues. Of several possible explanations, perhaps the most probable is that the shorter DNA-binding domains are slightly defective for specific DNA-binding *in vivo*. In this regard, such derivatives have a slightly increased non-specific DNA-binding activity *in vitro*[5]. The wild-type phenotype of Δ11, which has an 89 residue DNA-binding domain, may represent an exception, especially because this derivative was pre-selected for wild-type function.

Structural analysis

The fact that GCN4 can be extensively deleted without reducing its activity *in vivo* strongly suggests that tertiary structure is unimportant for the transcriptional-activation function. To investigate GCN4 structure in biochemical terms, ^{35}S-labelled proteins[16] were treated with a series of proteases which preferentially cleave unstructured protein regions (Fig. 3). A protease-resistant C-terminal domain was readily released by treatment with elastase, subtilisin, or papain. Under the same conditions, these proteases completely cleaved the N-terminal portion of the protein, although with variation in pattern and rate. These results indicate that GCN4 contains an independently structured DNA-binding (C-terminal) domain with the remainder of the protein being relatively unstructured.

In contrast, cleavage by chymotrypsin generates two equally stable intermediates, suggesting the apparently contradictory view that the large N-terminal region is indeed structured. Similar results were obtained, however, with N-terminally deleted derivatives (gcn4-C210, C197, C186, C163); that is, large N-terminal segments of GCN4 were resistant to protease digestion as if they formed part of a structured domain, yet they could be removed without destroying the integrity of that domain. Interestingly, a stable N-terminal fragment was not generated for gcn4-C141 or -C131, as if deletion beyond residue 118 (the centre of the acidic activation region) leads to a sudden loss in the resistance of the N-terminal segment to chymotrypsin. Thus, the unusual pattern seen with chymotrypsin correlates completely with the presence of a functional transcriptional-activation region. One interpretation of this result is that the activation region has a local structure that inhibits chymotrypsin cleavage of the otherwise unstructured N-terminal region of GCN4.

Molecular implications

Perhaps the most striking observation is that progressive deletion of the activation region causes a series of small, step-wise reductions of activity rather than defining a position where there is a precipitous loss of activity. This indicates that transcriptional-activation regions do not have a defined tertiary structure such as is found in active sites or domains in a protein, a view that is supported by our proteolysis experiments, by the short size and non-stringent amino-acid sequence requirements, and by the large number of protein deletions that remain functional. More importantly, 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 which act additively.

Although acidic character is clearly involved in this activity, it is not the only crucial feature because the level of transcriptional activation is only moderately correlated with net negative charge (Fig. 2). One clue to the structure of the GCN4 activation region is that the boundaries defining the step-wise levels of transcriptional activation appear to occur every seven amino-acid residues (residues 116, 123, 130, 137 and 144). The only exception to this rule is that derivatives in phenotypic classes 0 and 1 have endpoints within a single heptapeptide unit. Although this interpretation is clearly speculative, a repeat unit of seven amino acids is provocative because it could represent two turns of an α-helix. The acidic region of GCN4 could form three α-helices that are amphipathic, having acidic and hydrophobic residues along separate surfaces[17,18].

Recently it was shown that a synthetic 15 amino-acid region whose sequence is consistent with forming two turns of an amphipathic helix could confer some transcriptional activity in yeast when fused to a 147 amino-acid GAL4 DNA-binding domain[19]. This result seems inconsistent with our observations that acidic regions of 15 residues do not possess transcriptional-activation function. The 147 residue GAL4 domain, however, contains 70 residues beyond those necessary for DNA binding, including an acidic region (net negative charge of −7 between residues 66 and 118). In our opinion, the 15 residue peptide probably does not represent an autonomous activation region, but possibly one or two of the basic units that can act in combination with the otherwise cryptic GAL4 acidic region. It should be mentioned that the major GAL4 activation region and several other yeast activation regions are unlikely to form amphipathic helices. Thus, although an amphipathic helical structure may be compatible with a functional transcriptional-activation region, a simple relationship between this structure and function appears unlikely.

From several considerations, we propose that the activation region is a dimer of intertwined α-helices, one helix from each GCN4 monomer. First, the unusual chymotrypsin pattern reveals a structure in GCN4 that depends on the acidic activation region. Second, the stability of the dimeric DNA-binding domain, even in the absence of target DNA[20], should facilitate the formation and/or stability of the proposed interactions between helices of different monomer subunits. Third, as the LexA DNA-binding domain binds very poorly to its operator because of weak dimerization[21,22], it is probable that transcriptional stimulation through the LexA domain also requires that the activation region should facilitate dimerization[5,11,23]. The dimerization model also explains why amphipathic helices should form functional transcriptional-activation regions, as it would easily permit a structure involving interacting hydrophobic residues that are protected from solvent, and exposed acidic residues. Such amphipathic helices, however, would represent only one mechanism for forming dimers with exposed acidic residues; other mechanisms could be used for proteins such as GAL4.

As we and others have proposed, the activation region does not encode a catalytic activity but rather stimulates transcription by interacting with another protein. From various lines of evidence, we have argued against interactions with the basic histones and in favour of contacts with a TATA-binding protein or RNA polymerase II itself[2,24]. Presumably, such contact would cause an allosteric change in the general transcription machinery and hence would stimulate transcription. Although protein-protein interactions are usually thought to be highly specific, involving contact of complementary surfaces, the wide variety of sequences of transcriptional activation regions argues against an interaction of this type. Presumably, there is an ionic association between the acidic activation region and a basic pocket of the contacted protein. Perhaps the lack of specificity in this interaction is compensated for by the juxtaposition of the interacting elements along the DNA.

We thank Jennifer P. Macke for excellent technical assistance. This work was supported by the NIH.

Table 3 Requirements for spacing between the transcriptional activation and DNA-binding domains for GCN4 activity

GCN4 derivative	GCN4 amino acids present N-terminus	C-terminus	Transitional activity
gcn4-C175, Δ61	107-144	148-281	6
gcn4-C175, Δ62	107-144	172-281	4
gcn4-C175, Δ39	107-144	179-281	4
gcn4-C175, Δ63	107-144	185-281	4
gcn4-C175, Δ64	107-144	201-281	4
gcn4-C175, Δ65	107-144	218-281	4
gcn4-C175, Δ66	107-141	148-281	6
gcn4-C175, Δ41	107-141	179-281	3
gcn4-C175, Δ67	107-141	201-281	3
gcn4-C175, Δ68	107-136	148-281	5
gcn4-C175, Δ45	107-136	179-281	2
gcn4-C175, Δ69	107-136	201-281	2
gcn4-C175, Δ70	107-126	148-281	3
gcn4-C175, Δ47	107-126	179-281	1
gcn4-C175, Δ71	107-125	148-281	3
gcn4-C175, Δ49	107-125	179-281	1
gcn4-C175, Δ72	107-125	201-281	1
gcn4-C175, Δ73	107-121	148-281	0
gcn4-C175, Δ52	107-121	179-281	−1
gcn4-C175, Δ74	107-121	201-281	−1
gcn4-C175, Δ75	107-120	148-281	0
gcn4-C175, Δ53	107-120	179-281	−1
gcn4-C175, Δ76	107-115	148-281	−1
gcn4-C175, Δ56	107-115	179-281	−2
gcn4-C175, Δ77	107-109	148-281	−2
gcn4-C175, Δ60	107-109	179-281	−2

The structure and GCN4 activity of each derivative is described as in Fig. 1, where 6 indicates wild-type levels of transcriptional-activation, 0 indicates no activation, and −2 indicates repression. The N- and C-terminal segments are joined by a *Sal*I linker that encodes an additional arginine and proline residue.

Received 14 March; accepted 25 April 1988.

1. McKnight, S. L. & Tjian. R. *Cell* **46,** 795-805 (1986).
2. Struhl, K. *Cell* **49,** 295-297 (1987).
3. Guarente, L. *A. Rev. Genet.* **21,** 425-452 (1987).
4. Keegan, L., Gill, G. & Ptashne, M. *Science* **231,** 699-704 (1986).
5. Hope, I. A. & Struhl, K. *Cell* **46,** 885-894 (1986).
6. Hollenberg, S. M., Giguere, V., Segui, P. & Evans, R. M. *Cell* **49,** 39-46 (1987).
7. Miesfeld, R., Godowski, P., Maler, B. & Yamamoto, K. *Science* **236,** 423-427 (1987).
8. Kumar, V. *et al. Cell* **51,** 941-951 (1987).
9. Brent, R. & Ptashne, M. *Cell* **43,** 729-736 (1985).
10. Ma, J. & Ptashne, M. *Cell* **48,** 847-853 (1987).
11. Ma, J. & Ptashne, M. *Cell* **51,** 113-119 (1987).
12. Gill, G. & Ptashne, M. *Cell* **51,** 121-126 (1987).
13. Bohmann, D. *et al. Science* **238,** 1386-1392 (1987).
14. Angel, P. *et al. Nature* **332,** 166-171 (1988).
15. Struhl, K. *Nature* **332,** 649-650 (1988).
16. Hope, I. A. & Struhl, K. *Cell* **43,** 177-188 (1985).
17. vonHeijne, G. *EMBO J.* **5,** 1335-1342 (1986).
18. Roise, D., Horvath, S. J., Tomich, J. M., Richards, J. H. & Schatz, G. *EMBO J.* **5,** 1327-1334 (1986).
19. Giniger, E. & Ptashne, M. *Nature* **330,** 670-672 (1987).
20. Hope, I. A. & Struhl, K. *EMBO J.* **6,** 2781-2784 (1987).
21. Little, J. W. & Hill, S. A. *Proc. natn. Acad. Sci. U.S.A.* **82,** 2301-2305 (1985).
22. Hurstel, S., Granger-Schnarr, M. & Schnarr, M. *EMBO J.* **7,** 269-275 (1988).
23. Struhl, K. *Cell* **50,** 841-846 (1987).
24. Chen, W., Tabor, S. & Struhl, K. *Cell* **50,** 1047-1055 (1987).
25. Hill, D. E., Hope, I. A., Macke, J. P. & Struhl, K. *Science* **234,** 451-457 (1986).

LETTERS TO NATURE

The radio rings of Hercules A

Andrea Mason*, Philip Morrison† & Alberto C. Sadun‡

*Department of Physics, University of Padova, Italy
† Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
‡ Bradley Observatory, Agnes Scott College, Decatur, Georgia 30030, USA

Hercules A is a powerful, double radio source, about 0.5 Mpc across. Although its two lobes differ, they share a family resemblance in overall form. Along the axis of the somewhat weaker west lobe, however, a unique structure is found: a regular sequence of broken circular rings of enhanced radio intensity. The five rings fit a simple kinematic model in which the rings are projected spherical shells of higher radio brightness. They grew as spherical waves with a uniform speed of expansion, all from one common place of origin, roughly where the narrow input jet widens sharply. The shells drift one after another outward at constant speed with the radiating medium of the lobe. Here we propose a model of inelastic proton–proton collisions within dense filamentary condensations of the input jet for the five initiating particle bursts. Source development contrasts sharply with that inferred for the prototype radio double, Cygnus A, where the radio lobe is nearly stationary and the electrons are newly energized at the outermost end of the lobe. The two sources point to quite distinct mechanisms for generating large radio-emitting plasmas.

The extended double radio source Hercules A(3C348) centres on a galaxy whose nucleus is active both in radio and in optical light. Two long narrow collinear radio-emission features (we call them jets), one on each side of the centre, connect the galaxy to much wider and longer radio-bright regions. These roughly continue along the jet axis, although with some irregularity of form and moderate curvature; we call them radio lobes. The source has been imaged at 5 GHz with the Very Large Array (VLA), its intensity pattern represented both by contour and by grey-scale maps of high spatial resolution over three decades of dynamic range[1]. This superb mapping has disclosed that along the west lobe there is an array of five incomplete, thin, remarkably circular rings of enhanced radio brightness. The farther out the ring, the larger is its diameter. (The smallest arc is less closely circular; it is partially filled like the next larger ring.) We have fitted the nearly round intensity contours roughly to five circular annuli, each having a similar width but a distinct centre and radius. We take these bright rings as real enhancements of radio intensity. The most distant point of the final arc is ~300 kpc from the central galaxy, as projected. The rings do not seem to be artefacts of aperture synthesis; scattering and related illusory images are physically implausible for radio features.

A simple kinematic interpretation is evident. One after the other, each of the shells in turn began to expand, all originating close to a single position, near the plane where the narrow jet first abruptly widens into the lobe. Then each ring grew uniformly in diameter while its centre drifted steadily outwards along the lobe. Figure 1 records the visual fit of ring sizes and positions to the model. No counterpart rings are seen in the brighter east lobe (3/4 of the total radio power), although suggestive disturbances in its form are found at distances from the centre that roughly correspond to the ring distances out along the west lobe[1].

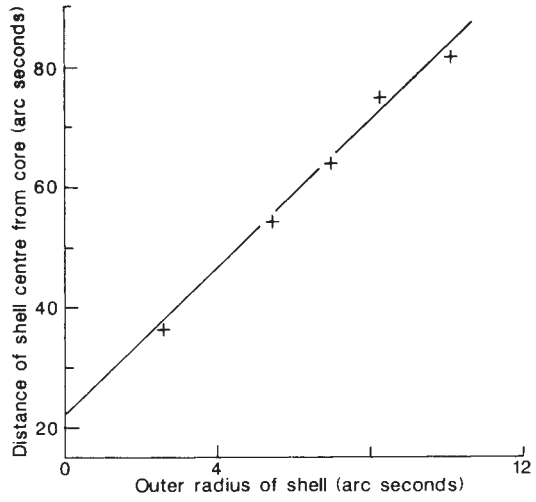

Fig. 1 The shell centre position and shell diameters for all five radio shells of Her A. Their position of common origin is about 20-25 arcsec from the core; the radiating medium of the lobe drifts uniformly outward at 6 times the common shell expansion speed. Comparison with other sources suggests that the drift speed is of the order 0.1 or $0.2c\,h^{-1}$. We have used the values $H_0 = 50$ h km per s-Mpc, with $q_0 = 0.5$. The observed redshift for Her A is $z = 0.15$, so that on the scale assumed 1 arcsec = 3.3 per h kpc projected on the sky.

The implied history of the overall form of Hercules A is as follows. The tiny central engine sends out its bipolar jets of mildly relativistic protons with magnetic field. The west jet includes some fine-grained threads with current density far above average. At those long-stationary places where the lobes begin, the beam collimation was from time to time lost, perhaps by the cumulative effects of entrainment of ambient mass during the jet passage from the centre. The resemblance between east and west lobe morphology speaks against an entirely external dense target encountered by chance in the west jet; both beams vary internally with time in some centrally determined way, perhaps by transient increases in jet current density. Both narrow beams grow unstable as they propagate outward, lose their