Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 2401–2405, April 1987
Genetics

Mutations that inactivate a yeast transcriptional regulatory protein cluster in an evolutionarily conserved DNA binding domain

(GAL4/DNA binding protein/Zn^{2+} binding domain/*Saccharomyces cerevisiae*)

MARK JOHNSTON AND JIM DOVER

Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110

Communicated by Stuart Kornfeld, December 19, 1986

ABSTRACT **The protein encoded by the *GAL4* gene of the yeast *Saccharomyces cerevisiae* binds to DNA upstream of several genes and activates transcription. To try to understand these processes, we have undertaken a genetic analysis of *GAL4*. Here we report that nearly all missense mutations in *GAL4*, selected *in vivo* to lack function of the protein, cluster in the small region of the gene that encodes the DNA binding domain. About half of these mutations alters a cysteine-rich region of the protein highly homologous to several eukaryotic DNA binding proteins; the other half alters some of the 20 amino acids adjacent to the cysteine-rich region. Nearly all of the missense mutations that alter the DNA binding domain abolish the DNA binding activity of GAL4 protein measured *in vitro*. In contrast, nearly all of the mutations that alter the 3′ 95% of the gene that encodes the transcription activation function are nonsense or frameshift mutations. These results support the idea that the conserved cysteine-rich sequence motif is directly involved in binding of several eukaryotic transcriptional regulatory proteins to DNA.**

The *GAL4* gene of *Saccharomyces cerevisiae* encodes a protein required to activate transcription of several genes involved in galactose utilization (1). This protein (hereafter termed GAL4 protein) binds to specific DNA sequences located upstream of each regulated gene (2–7); the events that occur subsequent to DNA binding that lead to activation of transcription are unknown.

Each gene activated by GAL4 protein contains at least one, but usually two or more, GAL4 protein binding sites located 95–410 base pairs (bp) upstream of the transcription initiation site (6, 7). Each binding site is related to a 23-bp consensus sequence that displays partial 2-fold rotational (dyad) symmetry. Because these GAL4 protein binding sites cause transcription activation in either orientation (8) and at variable distances from a gene, they have been compared to enhancer sequences of higher eukaryotes (9). A central issue in the study of eukaryotic gene expression is the question of how proteins that activate transcription, which presumably act by binding to enhancer elements, can work over relatively great distances. The GAL4 protein provides an excellent opportunity to gain insight into this problem.

The GAL4 protein is unusually large, consisting of 881 amino acids (predicted M_r 99,350) (10, 11). Among the functions of GAL4 protein are (*i*) DNA binding, (*ii*) transcription activation, (*iii*) ability to enter the nucleus, (*iv*) interaction with the product of the *GAL80* gene (a GAL gene regulatory protein that inhibits the ability of GAL4 protein to activate transcription), (*v*) possible direct involvement in catabolite repression, and presumably (*vi*) multimer formation. Some of the functional domains of this protein have been roughly localized. The DNA binding function resides in the first 74 amino acids (12), as do determinants specifying nuclear localization (13). The transcription activation function is encoded in the 3′-terminal 95% of the gene (14). The region of the protein responsible for interaction with *GAL80*-encoded protein also lies in the carboxyl-terminal 95% of the protein (15).

To locate more precisely functional regions of the GAL4 protein, we have undertaken a genetic analysis of GAL4. Here we report that most of the missense mutations that inactivate GAL4 protein lie in or adjacent to the region of the gene that encodes a cysteine-rich sequence that has been proposed to be a DNA binding domain in several proteins. We show that nearly all of these mutations abolish the DNA binding activity of GAL4 protein. These results strongly suggest that this region of the protein is directly involved in binding to DNA.

MATERIALS AND METHODS

Yeast and Bacterial Strains and Plasmids. Mutants of *GAL4* were isolated in YM725 (*ura3-52 his3-200 ade2-101 lys2-801 trp1-901 tyr1-501 met⁻ can^R gal4-542 gal80-538*) and YM1764 (isogenic to YM725, except *MET⁺ CAN^S*). The *his3*, *trp1*, *gal4*, and *gal80* mutations in these strains are deletions of all coding sequences that were constructed *in vitro* and transplaced into the yeast genome (16). The mutants we isolated carry mutations in *GAL4* present on a yeast plasmid in these strains. This plasmid (pBM292) is derived from pTC3 (17) and contains *CEN3*, *TRP1*, *ARS1*, and *GAL4*. The *GAL4* gene is present as a *Bam*HI–*Hin*dIII fragment encompassing nucleotides 1–3332 (10). These strains also carry a yeast plasmid (pBM252) with the *HIS3* gene fused to the *GAL10* promoter (18). The GAL4 protein was produced in *Escherichia coli* strain JM105 (19). The *gal4* mutants were selected on minimal plates containing 5% glycerol as carbon source and 0.1% 2-deoxygalactose, as described (20).

Construction of a Plasmid for Expression of GAL4 Protein in *E. coli*. The *GAL4* gene was inserted downstream of the *tac* promoter in pKK223-3 (ref. 21; obtained from Pharmacia) by ligating *Eco*RI–*Hin*dIII-cut plasmid to the *Sph* I–*Hin*dIII fragment of *GAL4* [encompassing nucleotides 473–3332 (10), which encode the last 872 amino acids of GAL4 protein] and a synthetic oligonucleotide (with *Eco*RI and *Sph* I "sticky" ends) that encodes the first 9 amino acids of GAL4 protein.

Preparation of *E. coli* Crude Extracts. Crude extracts of *E. coli* were prepared essentially as described (12). Cells (50 ml) grown to an OD_{600} of ≈0.7 were induced for *GAL4* expression by the addition of isopropyl β-D-thiogalactoside (IPTG) to a final concentration of 1 mM. Two hours after IPTG addition, cells were harvested, washed once with 1 ml of buffer A(50), resuspended in 0.5 ml of buffer A(200) in a 1.5-ml Eppendorf tube, and lysed by sonication on ice (10 2-sec bursts, each separated by 10 sec of cooling). (Buffer A is described in ref. 22. The number in parentheses is the KCl concentration of

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.

Abbreviations: bp, base pair(s); IPTG, isopropyl β-D-thiogalactoside.

the buffer in mM.) Aliquots of the cleared lysate obtained after centrifugation in a Sorvall SS34 rotor at 15,000 rpm for 3 hr were frozen in liquid N_2 and stored at −80°C.

DNA Binding Assays. The binding of GAL4 protein to DNA was measured by the nitrocellulose filter binding assay (23), as described (12, 22). DNA with GAL4 protein binding sites used for nitrocellulose filter binding assays was pUC18 (19) carrying the 685-bp *Eco*RI–*Bam*HI fragment of pBM150 (18), which spans the region between *GAL1* and *GAL10*. Present on this piece of DNA are four binding sites for GAL4 protein (6, 7). Each 20-μl reaction mixture contained buffer A(50) (without the protease inhibitors), 10 μg of sonicated salmon sperm DNA (Sigma), 1 mM IPTG, 10 fmol of [^{32}P]DNA (50,000–200,000 cpm), and 1–10 μl of crude extract [diluted in buffer A(50), or undiluted]. The reaction mixtures were incubated at 25°C for 10 min, placed on ice, and filtered through BA85 nitrocellulose filter discs (24 mm in diameter, Schleicher & Schuell) immediately after dilution with 1 ml of ice-cold buffer A(50) (without protease inhibitors) at a rate of 2–3 ml/min. The filters were washed with 1 ml of ice-cold buffer A(50), dried, and assayed for radioactivity. Background (with no added crude extract) was always <5% of total cpm.

RESULTS

Missense Mutations That Inactivate GAL4 Protein Cluster in the DNA Binding Domain. To locate functional regions of the GAL4 protein, we identified the sequence changes in a large number of spontaneous and UV-induced *gal4* mutants isolated using a method that selects for mutants unable to express the *GAL1* gene. The selection method relies on the fact that 2-deoxygalactose is toxic to yeast if it is phosphorylated, a reaction catalyzed by the *GAL1*-encoded enzyme galactokinase (20). Among the mutants deficient in galactokinase, and therefore resistant to 2-deoxygalactose, are *gal4* mutants, which we identified by complementation tests with a known *gal4* mutant. These *gal4* mutants were also recognized by their inability to express a *HIS3* gene fused to the *GAL1* promoter on a yeast plasmid. All of the *gal4* mutations we identified appear to prevent *GAL* gene expression completely: they cause a Gal$^-$, His$^-$ phenotype in *his3*$^-$ strains that carry the *HIS3*$^+$ gene fused to the *GAL1* promoter. Each mutation was mapped by UV-induced mitotic recombination to regions of the *GAL4* gene of ≈200 bp. The sequence changes in these mutants were then determined by sequencing the appropriate region from each mutant. Details of isolation and characterization of these mutants will be described elsewhere, along with a complete list of their sequence changes. Our results are summarized in Fig. 1.

A disproportionate number of mutations map in the 5′-terminal 5% of the gene, which encodes the DNA binding function (Fig. 1*A*). Forty-seven of the 216 mutations we mapped (21.8%) lie in the 153 nucleotides that encode the first 51 amino acids (5.8%) of GAL4 protein. Most of the mutations (27/31) in the 5′-terminal portion of GAL4 are missense mutations. The sequence changes caused by these mutations are presented in Fig. 1*B*. They are located in two regions of the DNA binding domain: one group alters amino acids in the cysteine-rich region that is homologous to several eukaryotic DNA binding proteins (ref. 24; see *Discussion*); the other group alters some of the 20 amino acids immediately adjacent to the cysteine-rich region.

In contrast to these results, all but 3 of 38 mutations that lie in the 3′-terminal 95% of the gene are either nonsense or frameshift mutations. Apparently, most amino acid changes in this part of the protein do not reduce *GAL* gene expression enough to allow growth under the selection conditions. The sequence changes caused by the three missense mutations in this region of the protein are shown in Fig. 1*C*.

Detection of GAL4 Protein DNA Binding Activity *in Vitro*. To assay the DNA binding activity of GAL4 protein, we produced it in *E. coli* by constructing a plasmid that directs the expression of *GAL4* from the *tac* promoter (see *Materials and Methods*). We detected GAL4 protein in crude cell extracts by its ability to retain on nitrocellulose filters [^{32}P]DNA containing GAL4 protein binding sites. Fig. 2 shows that this assay reliably detects the DNA binding activity of GAL4 protein produced in *E. coli*: if extract of a strain that expresses *GAL4* is used, only DNA that contains GAL4 protein binding sites is retained on filters (Fig. 2*A*); extracts of strains that do not express *GAL4* do not retain

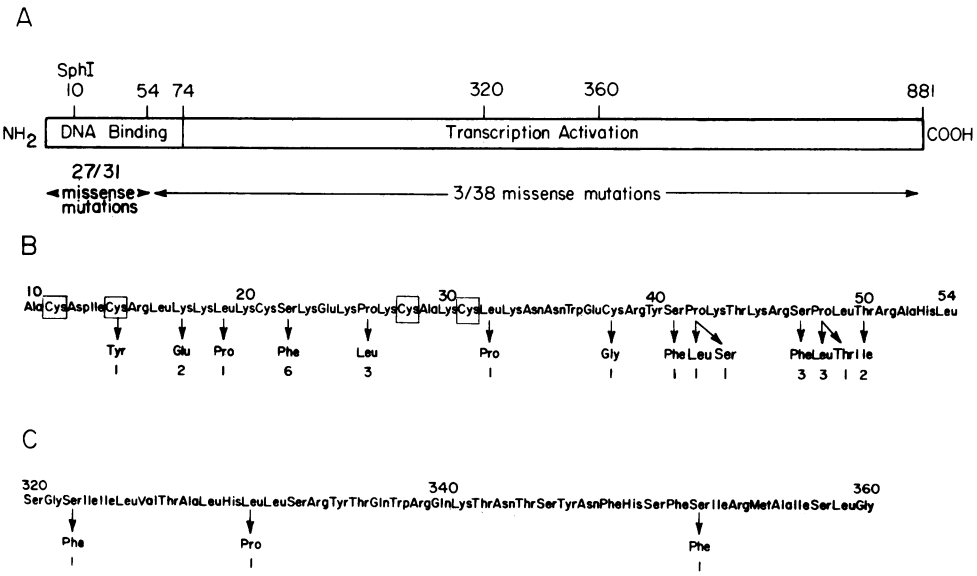

FIG. 1. (*A*) Schematic representation of GAL4 protein. Indicated in the box are the locations of functional regions of GAL4 protein, deduced from the evidence of Keegan *et al.* (12) and Brent and Ptashne (14). Below the box is shown the location of *gal4* mutations selected *in vivo* to lack GAL4 protein function. Shown above the box are locations of landmark amino acids referred to in the text and in the bottom part of the figure. (*B*) Sequence changes caused by mutations that alter the region between amino acids 10 and 54. The amino acid changes caused by each mutation are indicated below the wild-type sequence. The numbers below each amino acid change are the number of times that a particular mutation was isolated. (*C*) Sequence changes caused by mutations that alter the region between amino acids 320 and 360.

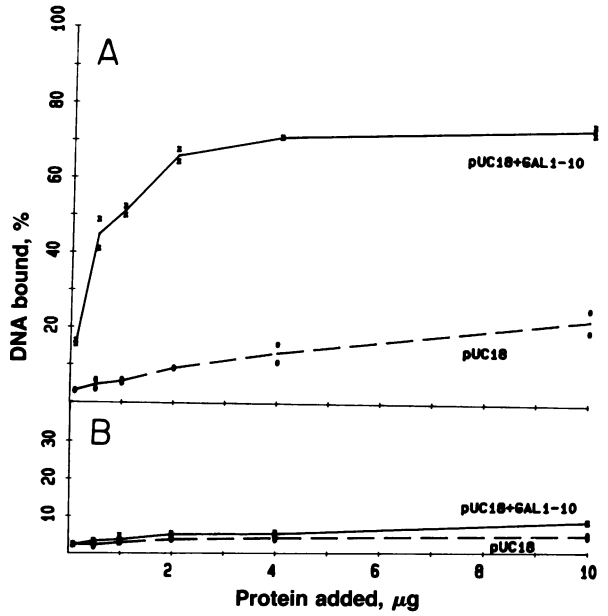

FIG. 2. Nitrocellulose filter binding assays of GAL4$^+$ and gal4$^-$ cell extracts. (*A*) An extract from a strain (BM1123) producing GAL4 protein was used in the assay. (*B*) An extract from a strain (BM1131) carrying the expression plasmid without the *GAL4* gene inserted was used. DNA containing GAL4 protein binding sites (pUC18+ GAL1–10) and DNA without these sites (pUC18) were used.

DNA on nitrocellulose filters (Fig. 2*B*). Similar results have been reported by Keegan *et al.* (12).

Most of the Mutations in the DNA Binding Domain Abolish DNA Binding Activity of GAL4 Protein. To assess the effects of mutations on the DNA binding activity of GAL4 protein, each mutated *GAL4* gene was expressed in *E. coli* and crude extracts were assayed for DNA binding activity by the nitrocellulose filter binding assay. Construction of plasmids containing the mutations was facilitated by an *Sph* I site located upstream of all of the mutations, allowing replacement of the wild-type *Sph* I–*Hin*dIII *GAL4* fragment in the expression vector with the same fragment from each mutant (see Fig. 1). Most of the mutations that alter the DNA binding domain abolish the DNA binding activity of GAL4 protein assayed *in vitro* (Fig. 3). Two mutations that alter Pro-42 and two that affect Pro-48 reduce, but do not abolish, DNA binding of GAL4 protein. These results strongly suggest that amino acids in and adjacent to the cysteine-rich region of the protein constitute the DNA binding domain and may identify amino acids directly involved in this function.

Mutations That Alter Amino Acids Outside the DNA Binding Domain Do Not Affect DNA Binding Activity of GAL4 Protein. The three missense mutations that lie outside the region of *GAL4* that encodes the DNA binding domain have no detectable effect on the DNA binding activity of GAL4 protein *in vitro* (Fig. 4). These mutations may affect some other function essential for GAL4 protein activity (e.g., nuclear transport or transcription activation).

The results of the DNA binding assays are summarized in Table 1. Also reported are qualitative levels of the mutant GAL4 proteins (relative to wild type) estimated by immunoblotting, using antibody against the first 74 amino acids of GAL4 protein (see Table 1 legend). Though some of the mutant proteins (e.g., Pro-26 > Leu, Ser-22 > Phe) appear to have reduced levels in *E. coli*, most of them (13 of 17) are present at levels that are at least 45% of wild type. Therefore, the reduced DNA binding activity of the altered GAL4 proteins cannot be solely attributed to their instability in *E. coli*.

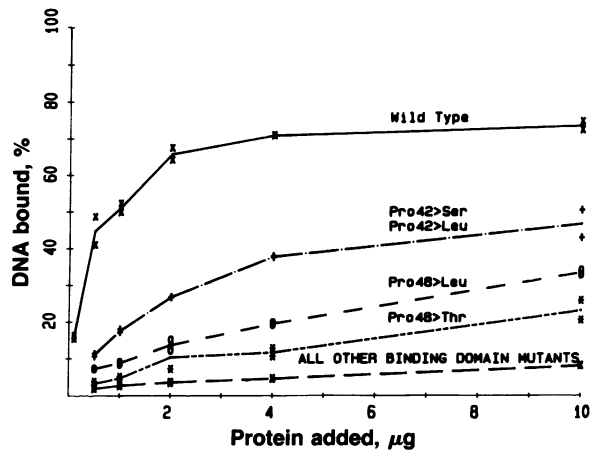

FIG. 3. Nitrocellulose filter binding assays of GAL4 proteins altered in the DNA binding domain. Extracts of cells producing GAL4 proteins with the amino acid changes that alter the DNA binding domain shown in Fig. 1*B* were used. Data representing "all other DNA binding domain mutants" are for the Cys-38 > Gly mutation; data obtained for the other mutants are similar. A complete tabulation of the mutations tested is listed in Table 1. None of these extracts retains on the filters above background levels DNA that does not contain GAL4 protein binding sites (data not shown).

DISCUSSION

We report here that nearly all of the missense mutations that inactivate *GAL4* alter the small portion of the protein that is the DNA binding domain. The nonrandom distribution of *gal4* mutations is highly significant: 90% (27/30) of the missense mutations alter the amino-terminal 5.8% of the protein that is sufficient for DNA binding. Because nearly all of the mutations that alter the DNA binding domain abolish the DNA binding activity of GAL4 protein *in vitro*, we think it is likely that the amino acids altered in these mutants are directly involved in binding of GAL4 protein to DNA.

Many of the *gal4* mutations we identified alter a region of the protein that is highly homologous to several other eukaryotic DNA binding and transcriptional regulatory proteins (Fig. 5). The homologies among some of these proteins have been recognized by others (12, 24, 27, 37, 39); we list them here to show their relationship to GAL4 protein. The GAL4 protein is highly homologous to another yeast protein that activates transcription, encoded by the *PPR1* gene (26): in a region of 24 amino acids (from Ala-10 through Lys-33),

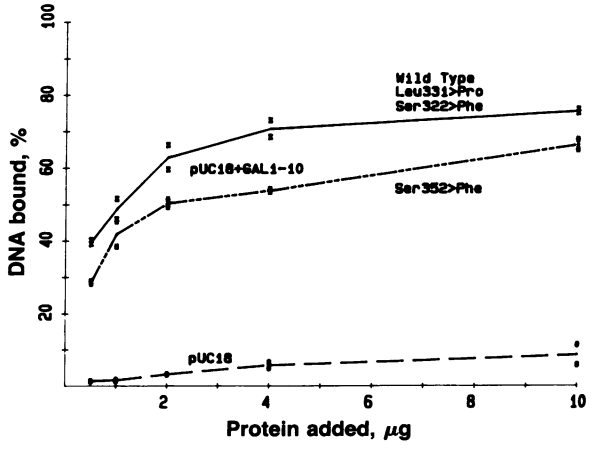

FIG. 4. Nitrocellulose filter binding assays of GAL4 proteins altered in the transcription activation domain. Extracts of cells producing GAL4 proteins with the amino acid changes shown in Fig. 1*C* were used.

Table 1. Effect of mutations on DNA binding activity of GAL4 protein

Allele	Amino acid change	DNA binding activity, % of wild type	GAL4 protein levels
wt	None	100.0	++
1260	Leu-331 > Pro	92.0	++
1259	Ser-322 > Phe	90.3	+
1203	Ser-352 > Phe	70.1	++
1834	Pro-42 > Ser	49.4	++
1836	Pro-42 > Leu	46.2	++
990	Pro-48 > Leu	25.6	+
843	Pro-48 > Thr	16.1	+−
1817	Cys-14 > Tyr	9.4	++
1840	Leu-19 > Pro	6.9	+
1819	Cys-38 > Gly	6.1	++
1258	Ser-22 > Phe	4.2	+−
1192	Lys-17 > Glu	4.0	+
1264	Thr-50 > Ile	3.8	+
840	Ser-47 > Phe	3.3	+−
842	Ser-41 > Phe	3.2	+
1188	Leu-32 > Pro	1.4	++
841	Pro-26 > Leu	1.3	+−
—	Vector without GAL	5.4	

The % DNA bound by 4 μg of extract containing wild-type (wt) protein was taken as 100% activity. The amount of DNA bound by 4 μg of extract containing each mutated protein is reported relative to wild type. The levels of GAL4 proteins in bacterial crude extracts were estimated from densitometer scanning of autoradiograms of immunoblots (done as described in ref. 25) using antibody against the first 74 amino acids of GAL4 protein (gift of L. Keegan, G. Gill, and M. Ptashne; see ref. 12). The levels reported are the average of three experiments (except for Pro-48 > Thr, which is an average of two experiments, and Ser-322 > Phe, which was determined only once) and are as follows: ++, 75–135% of wild-type levels; +, 45–70% of wild-type levels; +−, 10–25% of wild-type levels.

14 are *identical*; taking into account conservative amino acid differences, these two proteins are homologous at 17 of 24 positions. Especially striking is the conservation of the *number* and *location* of cysteine residues between these two proteins. This *same* organization of cysteine residues occurs in a number of other DNA binding and transcriptional regulatory proteins (shown in the middle of Fig. 5); a very *similar* organization of cysteine or cysteine and histidine residues occurs in other DNA binding and transcriptional regulatory proteins (bottom part of Fig. 5).

This cysteine-rich region, which in the proteins most homologous to GAL4 protein has the structure Cys-Xaa_2-Cys-Xaa_{13}-Cys-Xaa_2-Cys, is thought to bind a zinc atom between the four cysteines. The 13 amino acids between the two sets of cysteines are thought to form a "finger" of protein that contacts DNA. This idea is suggested by the observation that RNA polymerase III transcription factor A (TFIIIA) has a similar structure (with two histidines substituted for the second set of cysteines) (38), and this protein (which binds to DNA) has been shown to contain approximately one zinc atom for each of the nine repeats of the cysteine-histidine region that it contains (40). It remains to be shown that the other proteins listed in Fig. 5, including GAL4 protein, contain zinc, but the results presented here support the idea that this sequence motif is a common DNA binding domain.

It is noteworthy that the region of GAL4 protein in and adjacent to the cysteine-rich region is extraordinarily basic: more than one-third of the amino acids from Cys-11 through Arg-51 are basic (14 of 41 are arginine or lysine). An even higher fraction of the residues between the two pairs of cysteines is basic: 7 of these 13 amino acids are arginine or lysine! This is consistent with the proposal that these regions of the protein are *directly* involved in DNA binding.

About half of the missense mutations that alter DNA binding of GAL4 protein alter some of the 20 amino acids immediately adjacent to the cysteine-rich region. These mutations thus define a second region of the protein important for DNA binding. The sequence of this region shows little homology to other DNA binding proteins. It is noteworthy, however, that this region of GAL4 protein is highly homologous to the analogous protein of *Kluyveromyces lactis*, encoded by the

	10				*			*		*			*											*
GAL4	Ala	Cys	Asp	Ile	Cys	Arg	Leu	Lys	Lys	Leu	Lys	Cys	Ser	Lys	Glu	Lys	Pro	Lys	Cys	Ala	Lys	Cys	Leu
	33																							
PPR1	Ala	Cys	Lys	Arg	Cys	Arg	Leu	Lys	Lys	Ile	Lys	Cys	Asp	Gln	Glu	Phe	Pro	Ser	Cys	Lys	Arg	Cys	Ala
	94																							
LAC9	Ala	Cys	Asp	Ala	Cys	Arg	Lys	Lys	Lys	Trp	Lys	Cys	Ser	Lys	Thr	Val	Pro	Thr	Cys	Thr	Asn	Cys	Leu
	420																							
hGR	Leu	Cys	Leu	Val	Cys	Ser	Asp	Glu	Ala	Ser	Gly	Cys	His	Tyr	Gly	Val	Leu	Thr	Cys	Gly	Ser	Cys	Lys
	184																							
hER	Tyr	Cys	Ala	Val	Cys	Asn	Asp	Tyr	Ala	Ser	Gly	Tyr	His	Tyr	Gly	Val	Trp	Ser	Cys	Glu	Gly	Cys	Lys
	36																							
v-erbA	Gln	Cys	Val	Val	Cys	Gly	Asp	Lys	Ala	Thr	Gly	Thr	His	Tyr	Arg	Cys	Ile	Thr	Cys	Glu	Gly	Cys	Lys
	153																							
Ad2E1A	Gly	Cys	Arg	Ser	Cys	His	Tyr	His	Arg	Arg	Asn	Thr	Gly	Asp	Pro	Asp	Ile	Met	Cys	Ser	Leu	Cys	Tyr
	451																							
MuPV-T	Ser	Cys	Ile	Lys	Cys	Ser	Lys	Glu	Glu	Thr	Arg	Leu	Gly	Ile	His	Trp	Lys	Asn	His	Arg	Lys	His	Ala
	301																							
SV40-T	Met	Cys	Leu	Lys	Cys	Ile	Lys	Lys	Glu	Gln	Pro	Ser	His	Tyr	Lys	Tyr			His	Glu	Lys	His	Tyr
	336																							
ScMet-RS	Glu	Cys	Pro	Lys	Cys	His	Tyr	Asp	Asp	Ala	Arg	Gly	Asp	Gln					Cys	Asp	Lys	Cys	Gly
	66																							
RPO21	Lys	Cys	Gln	Thr	Cys	Gln	Glu	Gly	Met	Asn	Glu	Cys	Pro	Gly					His	Phe	Gly	His	Ile
	66																							
RPO31	Glu	Cys	Ala	Thr	Cys	His	Gly	Asn	Leu	Ala	Ser	Cys	His	Gly					His	Phe	Gly	His	Leu
	105																							
ADR1#1	Val	Cys	Glu	Val	Cys	Thr	Arg	Ala	Phe	Ala	Arg	Gln	Glu	His	Leu	Lys	Arg		His	Tyr	Arg	Ser	His	
	133																							
ADR1#2	Pro	Cys	Gly	Leu	Cys	Asn	Arg	Cys	Phe	Thr	Arg	Arg	Asp	Leu	Leu	Ile	Arg		His	Arg	Gln	Lys	Ile	His
	(7)																							
TFIIIA	Val	Cys	X	X	Asp	Gly	Cys	Asp	Lys	Arg	Phe	Thr	Lys	Lys	X	X	Leu	Lys	Arg	His	X	X	X	His

FIG. 5. Homologies among eukaryotic DNA binding and transcriptional regulatory proteins. This is only a partial list of proteins that are homologous in the cysteine-rich region; for a more complete list, see ref. 24. The conserved cysteine and histidine residues are boxed. The proteins listed are (in their order of appearance here) GAL4 protein (10), *PPR1*-encoded protein (26), *LAC9*-encoded protein (27), human glucocorticoid receptor (hGR) (28), human estrogen receptor (hER) (29, 30), the protein encoded by the oncogene v-*erbA* (31), adenovirus 2 E1A protein (Ad2E1A) (32), large T antigen of murine polyomavirus (MuPV-T) (33), large T antigen of simian virus 40 (SV40-T) (34), methionyl tRNA synthetase from yeast (ScMet-RS) (35), the largest subunit of RNA polymerase II from yeast (RPO21) (36), the largest subunit of RNA polymerase III from yeast (RPO31) (36), yeast *ADR1*-encoded protein, which contains two copies of the conserved sequence (ADR1-1 and ADR1-2) (37), and the consensus sequence of the nine copies in RNA polymerase III transcription factor A (TFIIIA) (38). (Several proteins that contain cysteine-rich sequences homologous to those above have been left out here because they have not yet been shown to be DNA binding or transcriptional regulatory proteins.) The number above the first amino acid of each sequence is the location of that residue in the protein. The asterisks above GAL4 protein indicate the amino acids affected in the mutants reported here.

LAC9 gene. The *LAC9*-encoded protein is able to substitute for the function of GAL4 protein in *S. cerevisiae* but is homologous to the GAL4 protein only in the amino-terminal region and in a short region at the carboxyl terminus (27). All of the *gal4* mutations except one (Ser-47 > Phe) that alter the stretch of amino acids adjacent to the cysteine-rich region alter residues that are identical between the GAL4 and LAC9 proteins. These amino acids may provide the DNA sequence specificity of binding. In this view, the highly conserved cysteine-rich region might then provide nonspecific interactions with DNA or might provide the proper structure for the adjacent residues to interact with DNA.

It was possible that the mutations we characterized indirectly affect the DNA binding activity of GAL4 protein by destabilizing it in bacteria. However, most of the altered GAL4 proteins are present in *E. coli* extracts at levels roughly similar to wild-type protein (Table 1). Though it could be argued that the effect of some of the mutations (e.g., Pro-26 > Leu) on the DNA binding activity of GAL4 protein is due to its reduced stability in the bacterial extracts, this is clearly not the case for most of the other mutations (e.g., Leu-32 > Pro). Even for the mutant proteins with reduced levels in the bacterial extracts, the reduction in GAL4 protein levels by a factor of 5–10 seems insufficient to account for the apparently complete lack of DNA binding activity. Thus, we infer that the mutations have a direct effect on the DNA binding activity of GAL4 protein.

The three missense mutations we identified that lie outside the DNA binding domain do not affect the ability of GAL4 protein to bind to DNA *in vitro* (Fig. 4). These may alter amino acids in a region of the protein that interacts with other proteins (e.g., RNA polymerase) to achieve transcription activation. Alternatively, these mutations could alter a function essential for *in vivo* (but not *in vitro*) DNA binding, such as the ability of GAL4 protein to be directed to the nucleus. A third possibility is that they may merely destabilize GAL4 protein in yeast (but not in *E. coli*).

The carboxyl-terminal 95% of the GAL4 protein is sufficient for transcription activation: it is able to activate transcription when fused to the DNA binding domain of another protein (14). The dearth of missense mutations that map to the 3′ 95% of the gene was unexpected. Apparently, amino acid changes in this part of the protein have very little effect on function. Perhaps the selection method we used is too stringent. Nevertheless, it appears that this region of the protein is relatively insensitive to alteration. We imagine three possibilities to account for this result. First, there may be only a few amino acids that are critical for the transcription activation function. A second possibility is this region of the protein may contain several domains, each sufficient for function. In this case, only null mutations that alter more than one domain would be expected to block function. This proposed situation for GAL4 protein would be analogous to the redundancy of cis-acting control elements located upstream of many yeast genes. Finally, a single domain of the protein required for transcription activation might be so large, and make so many contacts with proteins with which it interacts, that single amino acid changes do not substantially disrupt function.

We thank our colleagues Athanasios Theologis, David Schlessinger, John Majors, Peter Burgers, Larry Salkoff, and Maynard Olson for advice and comments on the manuscript. We also thank Liam Keegan, Grace Gill, and Mark Ptashne for their gift of antibody to GAL4 protein. This work was supported in part by a grant from the National Institutes of Health (GM32540). M.J. is an Established Investigator of the American Heart Association.

1. Oshima, Y. (1981) in *The Molecular Biology of the Yeast Saccharomyces, Metabolism and Gene Expression*, eds. Strathern, J. N., Jones, E. W. & Broach, J. R. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 159–180.
2. Douglas, H. C. & Hawthorne, D. C. (1964) *Genetics* **49**, 837–844.
3. Tschopp, J. F., Emr, S. D., Field, C. & Sheckman, R. (1986) *J. Bacteriol.* **166**, 313–318.
4. Post-Beittenmiller, M. A., Hamilton, R. W. & Hopper, J. E. (1984) *Mol. Cell. Biol.* **4**, 1238–1245.
5. Shimada, H. & Fukasawa, T. (1985) *Gene* **39**, 1–9.
6. Bram, R. J., Lue, N. F. & Kornberg, R. D. (1986) *EMBO J.* **5**, 603–608.
7. Giniger, E., Varnum, S. & Ptashne, M. (1985) *Cell* **40**, 767–774.
8. Lorch, Y. & Kornberg, R. (1986) *J. Mol. Biol.* **186**, 821–825.
9. Guarente, L. (1984) *Cell* **36**, 799–800.
10. Laughon, A. & Gesteland, R. F. (1984) *Mol. Cell. Biol.* **4**, 260–267.
11. Laughon, A., Driscoll, R., Wills, N. & Gesteland, R. F. (1984) *Mol. Cell. Biol.* **4**, 268–275.
12. Keegan, L., Gill, G. & Ptashne, M. (1986) *Science* **231**, 699–704.
13. Silver, P., Keegan, L. & Ptashne, M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5951–5955.
14. Brent, R. & Ptashne, M. (1985) *Cell* **43**, 729–736.
15. Johnston, S. A., Zavortink, M. J., Debouck, C. & Hopper, J. E. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6553–6557.
16. Scherer, S. & Davis, R. W. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4951–4955.
17. Shaw, K. J. & Olson, M. V. (1984) *Mol. Cell. Biol.* **4**, 657–665.
18. Johnston, M. & Davis, R. W. (1984) *Mol. Cell. Biol.* **4**, 1440–1448.
19. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33**, 103–119.
20. Platt, T. (1984) *Mol. Cell. Biol.* **4**, 994–996.
21. Brosius, J. & Holy, A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6929–6933.
22. Bram, R. & Kornberg, R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 43–47.
23. Riggs, A. D., Suzuki, H. & Bourgeois, S. (1970) *J. Mol. Biol.* **48**, 67–83.
24. Berg, J. (1986) *Science* **232**, 485–487.
25. Francis, G. R. & Waterston, R. H. (1985) *J. Cell. Biol.* **101**, 1532–1549.
26. Kammerer, B., Guyonvarch, A. & Hubert, J. C. (1984) *J. Mol. Biol.* **180**, 239–250.
27. Salmeron, J. M., Jr., & Johnston, S. A. (1986) *Nucleic Acids Res.* **14**, 7767–7781.
28. Hollenberg, S. M., Weinberger, C., Ong, E. S., Cerelli, G., Oro, A., Lebo, R., Thompson, E. B., Rosenfeld, M. G. & Evans, R. M. (1985) *Nature (London)* **318**, 635–641.
29. Greene, G. L., Gilna, P., Waterfield, M., Baker, A., Hort, Y. & Shine, J. (1986) *Science* **231**, 1150–1154.
30. Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J.-M., Argos, P. & Chambon, P. (1986) *Nature (London)* **320**, 134–138.
31. Debuire, B., Henry, C., Benaissa, M., Biserte, G., Clavierie, J. M., Saule, S., Martin, P. & Stehelin, D. (1984) *Science* **224**, 1456–1459.
32. Perricaudet, M., Akusjarvi, G., Virtanen, A. & Pettersson, U. (1979) *Nature (London)* **281**, 694–696.
33. Friedmann, T., Esty, A., LaPorte, P. & Deininger, P. (1979) *Cell* **17**, 715–724.
34. Reddy, V., Thimmappaya, B., Dhar, R., Subramanian, K., Zein, B., Pan, J., Celina, J. & Weisman, S. (1978) *Science* **200**, 494–502.
35. Walter, P., Gangloff, J., Bonnet, J., Boulanger, Y., Ebel, J.-P. & Fasiolo, F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2437–2441.
36. Allison, L. A., Moyle, M., Shales, M. & Ingles, C. J. (1985) *Cell* **42**, 599–610.
37. Hartshorne, T. A., Blumberg, H. & Young, E. T. (1986) *Nature (London)* **320**, 283–287.
38. Ginsberg, A. M., King, B. O. & Roeder, R. G. (1984) *Cell* **39**, 479–489.
39. Krust, A., Green, S., Argos, P., Kumar, V., Walter, P., Bornert, J.-M. & Chambon, P. (1986) *EMBO J.* **5**, 891–897.
40. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4**, 1609–1614.