Determinants of Binding-Site Specificity Among Yeast C_6 Zinc Cluster Proteins

Richard J. Reece and Mark Ptashne*

Related DNA binding proteins often recognize similar DNA sites but can distinguish among them with the use of different protein-DNA contacts. Here, it is shown that members of the C_6 zinc cluster family of yeast transcriptional activators distinguish related DNA sites by a different mechanism. The DNA binding site for each of these proteins contains identical nucleotide triplets (CGG . . . CCG) but differs in the spacings between the triplets. It is shown that zinc clusters of these proteins work interchangeably to recognize the conserved triplets and that the region 19 amino acids to the carboxyl-terminal side of the zinc cluster, comprising the linker and the beginning of a dimerization element as inferred from the GAL4 crystal structure, directs the protein to its preferred site.

The yeast transcriptional activator GAL4 contacts DNA with a six-cysteine DNA binding motif called a C_6 zinc cluster (*1*). Data from x-ray crystallography and other experiments indicate that the GAL4 dimer contacts the outer 3 base pairs (bp) of a 17-bp site ($CGGN_{11}CCG$) (*2*, *3*). PUT3 and PPR1 are two other yeast transcriptional activators that contain C_6 zinc clusters similar to that of GAL4 (*4*). Like GAL4, each recognizes a DNA site containing two rotationally symmetric CGG triplets, but the triplets of these sites are separated by 10 and 6 bp, respectively (*5*). We examined whether PUT3 and PPR1, like GAL4, use their C_6 zinc clusters to recognize the CGG triplets in their binding sites and, if so, what factors determine their different DNA binding specificities.

The NH_2-terminal sequences of GAL4, PUT3, and PPR1 including the C_6 zinc cluster are shown in Fig. 1. A fragment of GAL4 containing the zinc cluster, linker, and dimerization regions (GAL4 amino acids 1 to 100) binds as a dimer specifically to a GAL4 site (Table 1, line 1) (*6*). The corresponding fragments of PUT3 (residues 31 to 126) and of PPR1 (residues 29 to 123) bind specifically to their respective sites (Table 1, lines 3 and 4) as determined by gel mobility assays (*7*). Each fragment, either PUT3(31–126) or PPR1(29–123), binds at least 200-fold more tightly to its own DNA binding site than to either heterologous site. The binding affinity of GAL4(1–100) for a GAL4 site is 10-fold greater than its affinity for a PUT3 site and over 650-fold greater than its affinity for a PPR1 site (Table 1). Comparison of PUT3(1–126) with PUT3(31–126) (Table 1, lines 2 and 3) indicates that the NH_2-terminal extension of the C_6 zinc cluster has no effect on DNA binding affinity or specificity.

A series of fragments and chimeras of GAL4, PUT3, and PPR1 were tested for DNA binding in gel mobility assays (Fig. 2). Under these conditions, PPR1(29–123) binds to a PPR1 site but fails to bind detectably to either a PUT3 or a GAL4 site; PUT3(31–126) binds only to the PUT3 site; and GAL4(1–100) binds to GAL4 and PUT3 sites but not to a PPR1 site (Fig. 2, lanes 2, 7, and 10). The binding of GAL4 to both GAL4 and PUT3 sites is consistent with the 10-fold difference in the affinity of this protein for each of these sites (Table 1, line 1).

Hybrids containing the PUT3 or PPR1 zinc cluster in place of that of GAL4 bind DNA with the specificity of GAL4, and a hybrid in which the zinc cluster of PUT3 was replaced by that of GAL4 binds with the specificity of PUT3 (Fig. 2, lanes 3, 5, and 8). For example, GAL4(1–38)+PUT3(61–126), a PUT3 molecule containing the zinc cluster of GAL4, binds to a PUT3 site (Fig. 2B, lane 3) but not to a GAL4 site (Fig. 2A, lane 3) nor to a PPR1 site (Fig. 2C, lane 3). Each of the chimeras described has a DNA binding affinity characteristic of the protein whose sequences are COOH-terminal to the zinc cluster (Table 1, lines 5, 6, and 8). These results indicate that the C_6 zinc clusters of the three proteins are interchangeable and make equivalent contacts with the CGG triplets found in each site. The GAL4 residues implicated by crystallography (*2*) to make specific base pair contacts, Lys^{17} and Lys^{18}, are Arg

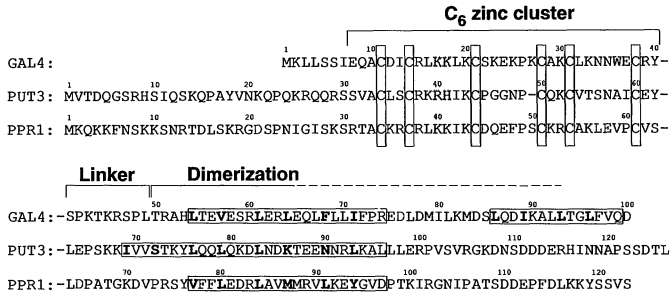

Fig. 1. Partial primary amino acid sequences of GAL4, PUT3, and PPR1. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. The sequences are shown with the six cysteine residues of each zinc cluster aligned. As defined by the GAL4 crystal structure (*2*), the positions of the C_6 zinc cluster, linker, and dimerization element of GAL4 are indicated, and the GAL4 residues implicated to make specific base pair contacts, together with the corresponding residues in PUT3 and PPR1, are shaded. The positions of potential coiled-coil structures in each sequence are boxed, with the generally hydrophobic first and fourth positions of each heptad repeat shown in bold (*16*).

Table 1. The DNA binding affinities of GAL4, PUT3, and PPR1 fragments and chimeras. DNA binding reactions were performed as described (Fig. 2), and the dissociation constants (K_d) were determined from the concentration of protein required to shift 50% of the DNA (*17*). Protein concentrations were estimated by the method of Bradford (*18*) with bovine serum albumin as a standard and are expressed as molarities of dimer. ND indicates not determined.

	Protein	K_d (nM)		
		GAL4 site CGGaggactgtcctCCG	PUT3 site CGGgaagcgcttcCCG	PPR1 site CGGcaattgCCG
1	GAL4(1–100)	3	30	>2000
2	PUT3(1–126)	>2000	20	>2000
3	PUT3(31–126)	>2000	20	>2000
4	PPR1(29–123)	>2000	>2000	10
5	GAL4(1–38)+PUT3(61–126)	>2000	20	>2000
6	PUT3(31–60)+GAL4(39–100)	3	30	>2000
7	PPR1(29–61)+GAL4(39–100)	750	ND	>2000
8	PPR1(29–63)+GAL4(41–100)	10	100	>2000
9	GAL4(1–61)+PUT3(84–126)	4	50	>2000
10	PUT3(31–79)+GAL4(58–100)	>2000	20	>2000
11	PPR1(29–80)+GAL4(58–100)	>2000	>2000	10

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138.

*To whom correspondence should be addressed.

and His, respectively, in the corresponding positions of PUT3 (Fig. 1); these residues should be able to make similar structural contacts.

The PPR1-GAL4 hybrid (Fig. 2, lanes 8) was constructed by including two amino acids of the PPR1 sequence COOH-terminal to the sixth cysteine of the zinc cluster (8). The crystal structure of GAL4 suggests that the tyrosine at position 40 (two residues beyond the sixth cysteine) is involved in maintaining the integrity of the zinc cluster domain. Consistent with this, a hybrid bearing NH_2-terminal PPR1 sequences ending at the sixth cysteine [PPR1(29–61)+GAL4(39–100)] was deficient in DNA binding (Table 1, line 7). Unlike PPR1, PUT3 contains a tyrosine residue at the position equivalent to Tyr^{40} of GAL4, and the hybrid bearing NH_2-terminal PUT3 sequences that terminate exactly at the sixth cysteine is functional (Fig. 2, lanes 5).

The differences in DNA binding specificity among GAL4, PUT3, and PPR1 are determined by the 19 amino acids directly to the COOH-terminal side of each zinc cluster. A molecule composed of amino acids 31 to 79 of PUT3 (the six-cysteine region and the 19 COOH-terminal residues) fused to amino acids 58 to 100 of GAL4 bound DNA with the specificity of PUT3 (Fig. 2, lanes 6). Also, a molecule composed of the six-cysteine region and the 19 COOH-terminal amino acids of PPR1 fused to the dimerization elements of GAL4 [PPR1(29–80)+GAL4(58–100)] bound DNA with the specificity of PPR1 (Fig. 2, lanes 9). Finally, a protein consisting of the GAL4 zinc cluster and COOH-terminal residues fused to the dimerization elements of PUT3 [GAL4(1–61)+PUT3(84–126)] bound DNA with the specificity of GAL4 (Fig. 2, lanes 4). Each of these hybrid proteins retains the approximate DNA binding affinity of the protein whose specificity it exhibits (Table 1, lines 9, 10, and 11).

These results are represented diagrammatically in Fig. 3 and show that the specificity determinant of GAL4 functions in three hybrid contexts: when fused to the zinc clusters of either PUT3 or PPR1 at one end and to GAL4 sequences at the other (Fig. 3, lines e and f) and when fused to the GAL4 zinc cluster at one end and to PUT3 sequences at the other (Fig. 3, line g). The corresponding determinant of PUT3 functions in two hybrid contexts (Fig. 3, lines d and h). Crystallography reveals that this region in GAL4 encompasses the linker and the NH_2-terminal end of the dimerization element (Fig. 1).

These data, taken with previous results, suggest that the linker and NH_2-terminal end of the dimerization element act as a unit to correctly position the zinc clusters for specific DNA binding (9), an idea also suggested in previous work (*10*). Consistent with this view, amino acids 48 to 56 of GAL4 (the end of the linker and the start of the dimerization element) are identical to those in the corresponding region of LAC9, a *Kluyveromyces lactis* protein that binds DNA with a specificity identical to that of GAL4; these two proteins share little overall sequence homology (*11*). The sequence similarities of GAL4 and PPR1 (Fig. 1) suggest that the linkers of these two proteins are of similar lengths and so the PPR1 linker may form some structure, other than the extended chain found in GAL4, that determines the separation of the zinc clusters. LAC9, LEU3, and HAP1, other members of the yeast C_6 zinc cluster family, also contain potential coiled-coil dimerization elements located on the COOH-terminal side of their zinc clusters (*11*, *12*), and it is likely that DNA binding site selection for these proteins is made by a mechanism similar to that described. However, other members of the family—for example, ARGR2 and MAL63 (*13*)—do not obviously contain coiled-coil elements on the COOH-terminal side of the zinc cluster and it is not clear whether the model for specificity outlined above will be appropriate for these proteins. For example, the DNA binding sites for ARGR2 appear to be asymmetric, containing only a single CGG triplet (*14*). Therefore, some members of the C_6 zinc cluster family may bind as monomers or perhaps in association with other proteins.

We have observed GAL4 binding to CGG triplets separated by 10 and 12 bp with an affinity about one-tenth of that for a native GAL4 site spacing of 11 bp (Table 1, line 1) (*15*), which perhaps indicates that the linker between the zinc cluster and the dimerization element of GAL4 is flexible, as suggested by the crystal structure (*2*). We have also noted PPR1 binding to CGG triplets separated by 7 and 8 bp, again with reduced affinity when compared to the native 6-bp spacing (*15*). In contrast, we have been unable to show PUT3 binding to any site other than one having the native 10-bp spacing, which perhaps indicates that the PUT3 linker is more rigid than the GAL4 linker. We have also observed that in yeast cells, full-length GAL4 is able to activate transcription from PUT3 DNA binding sites, whereas PUT3 is unable to activate from GAL4 DNA binding sites (*15*).

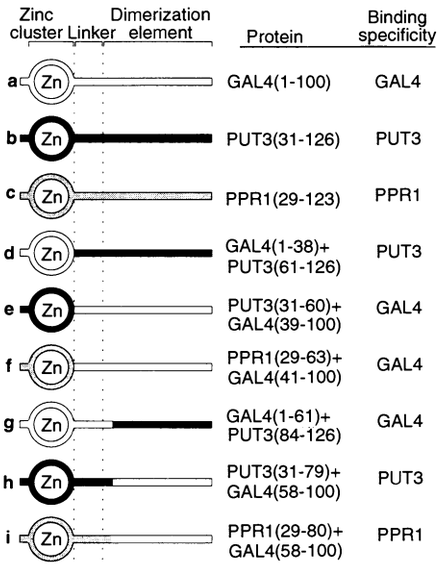

Fig. 3. DNA binding specificities of protein fragments and chimeras used in this study. Each protein is shown diagrammatically with GAL4 sequences in white, PUT3 sequences in black, and PPR1 sequences in gray. The binding specificity of each fragment is indicated as GAL4-like, PUT3-like, or PPR1-like.

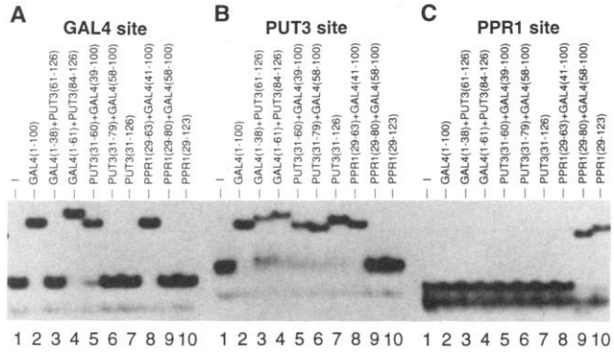

Fig. 2. DNA binding by fragments and chimeras of GAL4, PUT3, and PPR1. Protein (350 nM) was incubated in a 10-μl reaction mixture containing 20 mM Hepes (pH 7.5), 150 mM NaCl, 5 mM $MgCl_2$, 10 μM $ZnSO_4$, 5% glycerol, bovine serum albumin (100 μg/ml), and 10 pM ^{32}P-labeled DNA fragment. After 30 min at 23°C, the mixtures were loaded directly onto pre-electrophoresed 10% (w/v) polyacrylamide gels containing 45 mM tris base, 45 mM boric acid, 1 mM EDTA, and 1% glycerol. Electrophoresis was carried out at 15 V/cm for 60 min, and the gels were subjected to autoradiography. The double-stranded oligonucleotides used represent a GAL4 site (5′-tcCGGaggactgtcctCCGgt-3′) (**A**), a PUT3 site (5′-gatccCGGgaagcgcttcCCGaatt-3′) (**B**), or a PPR1 site (5′-ttCGGcaattgCCGaa-3′) (**C**), with the CGG triplets of each site shown in uppercase letters. Lane 1 of each gel has no protein added to indicate the position at which free DNA migrates.

REFERENCES AND NOTES

1. T. Pan and J. E. Coleman, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2077 (1990); P. J. Kraulis, A. R. C. Raine, P. L. Gadhavi, E. D. Laue, *Nature* **356**, 448 (1992).
2. R. Marmorstein, M. Carey, M. Ptashne, S. C. Harrison, *Nature* **356**, 408 (1992).
3. M. Carey, H. Kakidani, J. Leatherwood, F. Mostashari, M. Ptashne, *J. Mol. Biol.* **209**, 423 (1989).
4. J. E. Marczak and M. C. Brandriss, *Mol. Cell. Biol.* **11**, 2609 (1991); B. Kammerer, A. Guyonvarch, J. C. Hubert, *J. Mol. Biol.* **180**, 239 (1984).
5. A. H. Siddiqui and M. C. Brandriss, *Mol. Cell. Biol.* **8**, 4634 (1988); A. Roy, F. Exinger, R. Losson, *ibid.* **10**, 5257 (1990).
6. J. D. Baleja, R. Marmorstein, S. C. Harrison, G. Wagner, *Nature* **356**, 450 (1992).
7. PUT3 and PPR1 deletion derivatives and all chimeric proteins were constructed with the polymerase chain reaction (oligonucleotide sequences are available on request). Amplified DNA fragments were cloned into pET16b (Novagen, Madison, WI) cut with Nco I–Bam HI. The resulting plasmids contained the fusion gene under the direct control of the T7 promoter. All plasmids were subjected to DNA sequencing analysis to ensure that the correct fusion had been made and that no other mutations had arisen. Plasmids were transformed into *Escherichia coli* strain BL21-(DE3)pLysS (Novagen), and cells were grown at 37°C in Luria broth media containing ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) until an absorbance of 0.3 at 595 nm was reached. We then added $ZnSO_4$ to the cultures to a final concentration of 20 μM, and the cells were induced with isopropyl-β-D-thiogalactopyranoside (BRL) to a final concentration of 2 mM. Growth was continued for a further 2 hours, after which the cells were harvested by centrifugation (5000*g* for 15 min). The cell pellet was resuspended in 2.5% of the original culture volume of buffer A [20 mM Hepes (pH 7.5), 150 mM NaCl, 10% glycerol, 1.4 mM β-mercaptoethanol, and 20 μM $ZnSO_4$], and the cells were sonicated and centrifuged at 15,000*g* for 15 min. The supernatant was added to an equal volume of buffer A and applied to a 3-ml S Sepharose Fast Flow column (Pharmacia) equilibrated in buffer A. The column was washed with 100 column volumes of buffer A and developed with a 30–column volume gradient of buffer A to buffer A + 1 M NaCl. Peak fractions (followed by absorbance at 280 nm) were analyzed by SDS–polyacrylamide gel electrophoresis, pooled, and dialyzed against buffer A. With the use of this single column, each of the fusion proteins was isolated with greater than 95% purity.
8. Other experiments (*15*), either in vivo or with cell extracts, have suggested that a GAL4 or LAC9 molecule, whose six-cysteine region has been replaced by that of PPR1, binds DNA with the specificity of GAL4 but with at most one-tenth its affinity. In further experiments, when the PPR1 zinc cluster and the 14 amino acids to the COOH-terminal side were included in the swap, PPR1 specificity was conferred on the resultant protein but DNA binding activity was weak. Our results indicate that the reduced binding affinity observed in those studies was probably due to the suboptimal placement of the chimeric junction.
9. We have failed in attempts to construct active chimeras that contain just the PUT3 zinc cluster and linker fused to the dimerization region of GAL4. For example, PUT3(31–65)+GAL4(49–100), PUT3(31–67)+GAL4(53–100), PUT3(31–68)+GAL4(54–100), PUT3(31–74)+GAL4(53–100), and PUT3(31–76)+GAL4(62–100) were all inactive in DNA binding in our assays. However, we do not know the structure of PUT3, and it is possible that chimeric junctions were not appropriately positioned. Also, attempts to characterize a GAL4 molecule in which just the linker and start of the dimerization element have been replaced by PUT3 sequences [GAL4(1–38)+PUT3(61–79)+GAL4(58–100)] have been hampered by the insolubility of the protein.
10. J. C. Corton and S. A. Johnston, *Nature* **340**, 724 (1989); M. M. Witte and R. C. Dickson, *Mol. Cell. Biol.* **10**, 5128 (1990).
11. J. M. Salmeron and S. A. Johnston, *Nucleic Acids Res.* **14**, 7767 (1986).
12. P. Friden and P. Schimmel, *Mol. Cell. Biol.* **7**, 2708 (1987); *ibid.* **8**, 2690 (1988); B. Turcotte and L. Guarente, *Genes Dev.* **6**, 2001 (1992); K. Pfeifer, K.-S. Kim, S. Kogan, L. Guarente, *Cell* **56**, 291 (1989).
13. F. Messenguy, E. Dubois, F. Descamps, *Eur. J. Biochem.* **157**, 77 (1986); J. Kim and C. A. Michels, *Curr. Genet.* **14**, 319 (1988).
14. M. DeRijcke, S. Seneca, B. Punyammalee, N. Glansdorf, M. Crabeel, *Mol. Cell. Biol.* **12**, 68 (1992).
15. R. J. Reece and M. Ptashne, unpublished observations.
16. C. Cohen and D. A. D. Parry, *Proteins* **7**, 1 (1990).
17. L. Fairall, S. D. Harrison, A. A. Travers, D. Rhodes, *J. Mol. Biol.* **226**, 349 (1992).
18. M. M. Bradford, *Anal. Biochem.* **72**, 248 (1976).

19. We thank R. Marmorstein for the gift of purified PPR1(29–123) and many useful discussions. We also thank R. Marmorstein, H. Himmelfarb, N. Lehming, J. Brickman, and J. Robinson for critically reading the manuscript. R.J.R. was supported by fellowships initially from the European Molecular Biology Organization (ALTF 475-1990) and latterly from the Science and Engineering Research Council (B/92RF/1595). This work was supported by grant GM32308 from NIH.

10 February 1993; accepted 14 June 1993

Development of Mature CD8$^+$ Thymocytes: Selection Rather Than Instruction?

Joost P. M. van Meerwijk and Ronald N. Germain*

The role of major histocompatibility complex (MHC) molecules in T cell differentiation was investigated by comparison of thymocyte subpopulations in wild-type mice and β_2-microglobulin (β_2M) mutant mice deficient in MHC class I expression and mature CD8$^+$ cells. On the basis of surface markers, glucocorticoid resistance, in vitro differentiation capacity, and absence in $\beta_2M^{-/-}$ mice, CD4^{intermediate}CD8^{hi} cells with high expression of αβ T cell receptor (TCRαβ) were identified as having been positively selected by MHC class I for development into mature CD8$^+$ T cells. Activated CD4^{int}CD8^{hi} cells bearing intermediate rather than high amounts of TCR were present in both wild-type and $\beta_2M^{-/-}$ animals. These data suggest that recognition of MHC class I molecules is required for full maturation to CD8$^+$ T cells, but not for receptor-initiated commitment to the CD8$^+$ lineage, consistent with a stochastic (selection) model of thymocyte development.

An extensive series of studies on thymocyte differentiation have led to a proposed scheme of development in which precursor cells that lack surface expression of the TCRαβ-CD3 complex and the CD4 or CD8 coreceptors (triple negative cells) become either CD4$^-$CD8$^+$ or CD4$^+$CD8$^-$ intermediates and then CD4^{hi}CD8^{hi} (double positive, DP), TCRαβ^{lo} blasts. These cells are the substrate for two opposing selection events involving the coordinate activity of TCRs, the CD4 or CD8 coreceptors, and MHC class I or class II molecules: positive selection for differentiation to mature CD4^{hi}CD8$^-$ or CD4$^-$CD8^{hi}, TCRαβ^{hi} (CD4 or CD8 single positive, SP) cells, or negative selection that deletes self-reactive cells (*1*). The mechanism that leads to commitment of DP precursor cells to the CD4 or CD8 lineage, however, remains unclear. Some investigators favor "instructive" models, which postulate that differential signaling through CD4–class II or CD8–class I interactions determines the fate of bipotential precursor cells (*2*, *3*). Others propose "selective" models in which stochastic events determine the lineage commitment, but full maturation requires an appropriate match among the MHC restriction (class I or class II) of the TCR and the coreceptor that was retained (*4*, *5*).

Analysis of developmental intermediates could contribute to resolving this issue. The transition from DP-TCRαβ^{lo} to either CD4 or CD8 SP (TCRαβ^{hi}) mature cells must involve progression through intermediate states with increasing TCRαβ expression and with decreasing amounts of the coreceptor whose expression is to be extinguished. Guidos *et al.* described small thymocyte populations of CD4^{hi}CD8^{int}TCRαβ^{int} or CD4^{int}CD8^{hi}TCRαβ^{int} phenotype that show biased V_β usage associated with positive selection by certain MHC molecules (*6*). On the basis of these results, it was concluded that these cells were the product of positive selection events that led to development of CD4 or CD8 SP mature T lymphocytes. In contrast, other investigators have described a CD4$^+$CD8$^+$ DP subpopulation expressing high amounts of TCRαβ. The biased expression of V_β regions by DP-TCRαβ^{hi} cells in normal mice (*7*) and the presence of such cells only in TCR transgenic mice of the positively selecting haplotype (*3*) suggested that they

Lymphocyte Biology Section, Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892.

*To whom correspondence should be addressed.