MOLECULAR AND CELLULAR BIOLOGY, July 1996, p. 3773–3780 Vol. 16, No. 7
0270-7306/96/$04.00+0

DNA Sequence Preferences of GAL4 and PPR1: How a Subset of Zn_2Cys_6 Binuclear Cluster Proteins Recognizes DNA

STANLEY D. LIANG,[1]† RONEN MARMORSTEIN,[1]‡ STEPHEN C. HARRISON,[1,2] AND MARK PTASHNE[1]*

Department of Molecular and Cellular Biology, Harvard University,[1] and Howard Hughes Medical Institute,[2] Cambridge, Massachusetts 02138

Received 9 February 1996/Returned for modification 19 March 1996/Accepted 9 April 1996

Biophysical and genetic experiments have defined how the *Saccharomyces cerevisiae* protein GAL4 and a subset of related proteins recognize specific DNA sequences. We assessed DNA sequence preferences of GAL4 and a related protein, PPR1, in an in vitro DNA binding assay. For GAL4, the palindromic CGG triplets at the ends of the 17-bp recognition site are essential for tight binding, whereas the identities of the internal 11 bp are much less important, results consistent with the GAL4-DNA crystal structure. Small reductions in affinity due to mutations at the center-most 5 bp are consistent with the idea that an observed constriction in the minor groove in the crystalline GAL4-DNA complex is sequence dependent. The crystal structure suggests that this sequence dependence is due to phosphate contacts mediated by arginine 51, as part of a network of hydrogen bonds. Here we show that the mutant protein GAL4(1-100)R51A fails to discriminate sites with alterations in the center of the site from the wild-type site. PPR1, a relative of GAL4, also recognizes palindromic CGG triplets at the ends of its 12-bp recognition sequence. The identities of the internal 6 bp do not influence the binding of PPR1. We also show that the PPR1 site consists of a 12-bp duplex rather than 16 bp as reported previously: the two T residues immediately 5′ to the CGG sequence in each half site, although highly conserved, are not important for binding by PPR1. Thus, GAL4 and PPR1 share common CGG half sites, but they prefer DNA sequences with the palindromic CGG separated by the appropriate number of base pairs, 11 for GAL4 and 6 for PPR1.

GAL4 and PPR1 (pyrimidine pathway regulator 1), two *Saccharomyces cerevisiae* transcription regulatory proteins, are members of a family of at least 12 fungal transcription factors containing a region termed the Zn_2Cys_6 binuclear cluster (reviewed in reference 11). This region has six absolutely conserved cysteines and a number of other highly conserved residues. It has been shown for various members of the family that the Zn_2Cys_6 region is essential for DNA binding (reviewed in references 15 and 24). The DNA sites of several members of the family, including those recognized by GAL4 and PPR1, contain palindromic CGG triplets within their DNA sites (Fig. 1a). Thus, GAL4 and PPR1 are members of a subset of Zn_2Cys_6-containing proteins whose members recognize twofold symmetric CGG half sites separated by distinct numbers of base pairs (24, 29).

GAL4 is an activator of transcription of various galactose-inducible genes. These genes possess GAL4 binding sites positioned in the upstream regions of their promoters (3, 4, 9). Examination of 16 known natural sites reveals a 17-bp pseudopalindromic consensus sequence (9). This consensus site is a high-affinity binding site for GAL4 derivatives in vitro (this study) and confers GAL4-dependent transcriptional activation in vivo when placed upstream of a test reporter gene, both in yeast cells (18, 21a) and in mammalian tissue culture cells (36). The site has highly conserved palindromic CGG sequences at its ends separated by 11 bp.

* Corresponding author. Mailing address: Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Ave., Cambridge, MA 02138. Phone: (617) 495-9559. Fax: (617) 495-0758.

† Present address: Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032.

‡ Present address: The Wistar Institute and Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104-4268.

GAL4 is 881 amino acids in length. A polypeptide consisting of the amino-terminal 100 residues forms a dimer in solution and binds DNA with high affinity. A polypeptide consisting of the N-terminal 65 residues of GAL4 is a monomer in solution that binds DNA specifically as a dimer, although with reduced affinity (5, 24). Both the nuclear magnetic resonance solution structure of GAL4(1-65) (2, 20) and a cocrystal structure of the consensus 17-bp DNA site complexed to GAL4(1-65) have been determined (24). The crystalline complex demonstrates that GAL4 binds DNA as a homodimer. Each subunit of GAL4(1-65) comprises three segments (Fig. 1b). An N-terminal domain ligates two closely spaced Zn ions with six cysteine residues, forming a Zn_2Cys_6 binuclear cluster. The C terminus is helical and participates in a coiled-coil dimerization element. A nine-residue extended strand, termed the linker, connects the amino- and carboxyl-terminal domains. The cocrystal structure displays a picture of specific binding of GAL4 to its 17-bp site: the two Zn domains of a dimer make base-pair-specific contacts to the highly conserved CGG triplets at the ends of the site, and the linker and dimerization elements contact the phosphate backbone within the inner 11 bp. The DNA in the complex deviates little from B-form geometry, except for a 3-Å (0.3-nm) constriction of the minor groove at the center of the site. This constriction appears to be stabilized by phosphate contacts from arginine residues within the dimerization element.

PPR1 is the activator of transcription of genes in the pyrimidine metabolic pathway (18, 30). These genes possess PPR1 binding sites at positions upstream of their promoters. The consensus PPR1 site is a twofold symmetric 16-bp sequence (17, 30). It contains palindromic CGG sequences separated by 6 bp, with two T residues just 5′ to each CGG triplet. The region of PPR1 that corresponds to the segment of GAL4 contacting the CGG triplet is highly conserved (Fig. 1a), in-

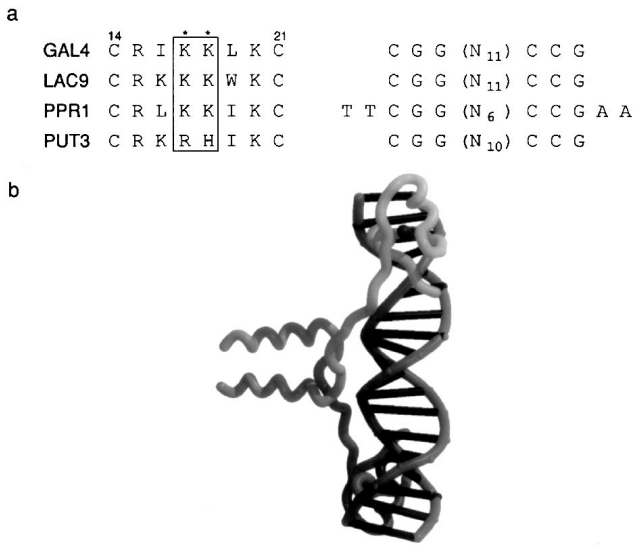

FIG. 1. (a) The left panel shows a comparison of amino acid sequences of four homologous proteins in a region that, in GAL4, contacts DNA base pair residues. In GAL4, this region is part of the first helix strand of the Zn_2Cys_6 binuclear cluster. Lysines 17 and 18 in GAL4 and the equivalents in the other proteins are boxed. These two residues in GAL4 contact the functional groups of the CGG sequences in the GAL4 site. The right panel shows the DNA sites of these four proteins. Only the sense strand is shown for simplicity. (b) GAL4(1-65)-DNA structure as observed in the crystalline complex (24).

cluding the two critical lysine residues. Domain swap experiments (29) have shown that the Zn_2Cys_6 domain of PPR1 can substitute for that of GAL4, the resulting hybrid protein retaining the specificity of GAL4. The Zn_2Cys_6 domain of GAL4 can also be replaced by that of PUT3, the yeast regulator of expression of genes in the proline utilization pathway, without loss of GAL4 specificity.

In this study, we assessed biochemically the importance for tight binding by GAL4 of each base pair in the GAL4 twofold symmetric 17-bp DNA site. To this end, we determined the relative affinities of GAL4(1-100) for its wild-type (WT) consensus 17-bp DNA site and mutant sites consisting of all possible doubly symmetric mutations of the site. We use an equilibrium competition gel shift assay to show that the palindromic CGG triplets at the ends of the GAL4 17-bp site are crucial for binding by GAL4. The identities of the inner 11 bp are relatively unimportant for binding, although the inner most 5 bp do contribute slightly to protein specificity.

We also present experiments showing that the palindromic CGG triplets of the PPR1 DNA site are crucial for PPR1 binding. But unlike the case for GAL4, the identity of none of the other base pairs, neither the 6 bp between the CGG triplets nor the 4 bp to either side of the CGG triplets, matters.

Although GAL4 and PPR1 both recognize DNA targets that contain two CGG half sites, we show that the spacing between the twofold symmetric CGG half sites is crucial for binding. Thus, GAL4 cannot bind to a PPR1 site ($CGGN_6CCG$), and PPR1 has very low affinity for a GAL4 site ($CGGN_{11}CCG$). The optimal spacing between two CGG half sites for GAL4 is 11 bp, although spacing of 10 or 12 bp is only somewhat less favorable. However, in vivo experiments show that a 16-mer is a poor site for GAL4 in the cell, whereas the 18-mer is a strong site. PPR1 cannot bind to sites that have CGG triplets separated by a spacing of 5 or 4 bp.

MATERIALS AND METHODS

Oligonucleotides. Oligonucleotides were made on an Applied Biosystems Inc. DNA synthesizer or on a Milligen DNA synthesizer. Oligonucleotides were purified by reverse-phase high-pressure liquid chromatography on a Dynamax 300 column (Rainen). The concentration of single-stranded oligonucleotides was calculated on the basis of the approximation of 1 mg/ml for an A_{260} absorbance of 30.3.

The oligonucleotides were labeled with ^{32}P at the 5′ ends as previously described (26). The oligonucleotides were annealed by heating to 90°C and slowly cooling to 4°C in 10 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES; pH 7.5)–150 mM NaCl. The extent of duplex formation was confirmed to be approximately 100% by electrophoresis on ethidium bromide-stained polyacrylamide gel.

Plasmid constructions. Mutant GAL4(1-100)R51A was generated by site-directed mutagenesis as previously described (17). Plasmid pBluescript KS+ (Stratagene) was linearized with *Eco*RV, and an *Sph*I 8-mer linker (New England Biolabs) was inserted. An *Sph*I-to-*Xho*I fragment, containing DNA encoding amino acids 11 to 74 of GAL4, was subcloned from pMH75 (32) into the KS+ Sph1 plasmid, creating pSL76. Single-stranded DNA of pSL76 was produced in and purified from *Escherichia coli* CJ236 after infection with helper M13 phage (Promega). An oligonucleotide with the mutant sequence embedded was synthesized (Applied Biosystems Inc. DNA synthesizer), annealed to the single-stranded DNA, extended with Klenow enzyme (New England Biolabs) in the presence of all four deoxynucleoside triphosphates (Pharmacia), ligated (Bethesda Research Laboratories ligase), and transformed into *E. coli* MM294. The correct clones were identified by DNA sequencing (26, 33). The *Sph*I-to-*Xho*I GAL4 fragment was transferred to pTAC-GAL4(1-100) (24), substituting for the WT fragment. The correct clones were again identified by DNA sequencing. GAL4(1-100)R51A is pTAC-SL209.

DNA sites were synthesized as blunt or overhanging oligonucleotides (Applied Biosystems Inc. DNA synthesizer). The overhanging oligonucleotides were rendered blunt by treating with mung bean nuclease (New England Biolabs). The oligonucleotides were inserted into the *Pvu*II site of pSP72 (Promega). The constructs were sequenced by standard procedures (33). Reporter constructs with one copy of a given site at −100 from TATA were made by inserting the *Xho*I-to-*Sph*I small fragments from the pSP72-based plasmids into a yeast integrating vector, a gift of J. Pearlberg (2, 27), at *Sal*I and *Sph*I, both unique sites. This places the site at −100 from TATA. The reporter plasmids with two copies of a given site were made by three-way ligation. The pSP72-based plasmids were digested with *Xho*I and *Hin*dIII, and the small fragments containing the sites were purified. These were inserted into the −100-TATA yeast integrating reporter plasmid, which was digested to completion at the unique *Sal*I site. The resulting clones had two copies of each of the *Xho*I-to-*Hin*dIII fragments inserted in the *Sal*I site. The reporters with two copies of a given site at −94 from TATA were constructed also by three-way ligation. In this case, the *Xho*I-to-*Hin*dIII fragments from the pSP72-based clones were inserted into the *Xho*I site of pLR1del1 (37). The reporters with two 16-mers, 17-mers, or 18-mers at −150 or −191 from TATA were constructed by cloning the *Xho*I fragments containing the sites from the pLR1del1 derivatives to the −150 or −191 site from TATA reporter clones, again gifts from J. Pearlberg (2, 27).

Effector plasmids used in this work all use the yeast *ADH1* (alcohol dehydrogenase) promoter to express the various proteins. PMH75 (32), a gift from M. Hollis, expresses full-length GAL4. It has the *TRP1* gene and an autonomous replicating sequence-centromere (ARS-CEN). PKW14, pKW23, and pKW41 (gifts from K. Wood) (39) express GAL4 amino acids 1 to 147 and activation domains I and II, GAL4 amino acids 1 to 147 and B17, and GAL4 amino acids 1 to 147 [referred to in this report as GAL4(1-147)+I+II, GAL4(1-147)+B17, and GAL4(1-147)]. They derive from pMH75.

Protein production. GAL4(1-100), GAL4(1-100)R51A, and PPR1(29-123) proteins were purified to homogeneity as described previously (24, 25).

DNA binding conditions. The binding reaction mixtures, in 30-μl volume, typically contained 10 mM HEPES (pH 7.5), 0.5 mM $MgCl_2$, 50 nM ZnCl, 150 mM NaCl, 0.1 mg of bovine serum albumin per ml, 5% glycerol, protein, and DNA. The DNA consisted of a ^{32}P-labeled GAL4 site (20 pM, unless otherwise indicated in figure legends) or PPR1 site (100 pM), in addition to variable concentration of unlabeled DNA for competition experiments. The reaction mixtures were incubated for 30 min at 25°C. Free DNA molecules were separated from protein-DNA complexes by electrophoresis at 100 V on 10% nondenaturing polyacrylamide gels containing 1% glycerol, 45 mM Tris-HCl (pH 8), 45 mM borate, and 1 mM EDTA.

Equilibrium competition gel shift assay. In the equilibrium competition gel shift assay, the protein concentration is chosen so that 70% or more of a fixed concentration of labeled WT DNA site is bound. Titrations are done with unlabeled oligonucleotides which compete for protein binding with labeled WT site. All DNA fragments, labeled or not, are added first, and the protein is added last. The protein and DNA are allowed to equilibrate for 30 min at 25°C. The mixtures are then subjected to electrophoresis. The gel is dried on a gel dryer (Bio-Rad), and quantitation of bound and free ^{32}P-labeled DNA site is performed on a Fuji BAS2000 image plate reader. The percent of ^{32}P-labeled bound DNA site is then plotted against the concentration of unlabeled competitor DNA on a semilog plot (concentrations of unlabeled sites on log phase). A comparison

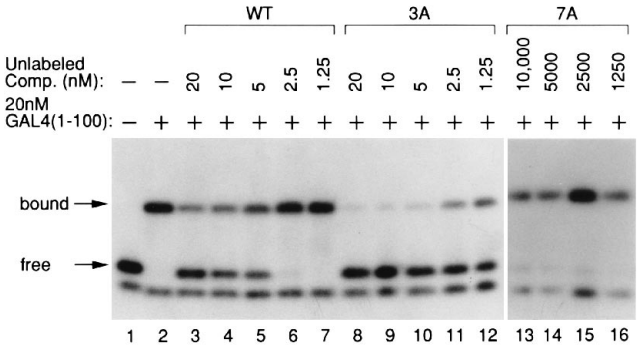

FIG. 2. Sample experiment of the equilibrium competition gel shift DNA binding assay. Each reaction mixture contained GAL4(1-100) and labeled WT GAL4 site at a concentration of 20 pM plus unlabeled WT or mutant site as a competitor (Comp.) DNA at the concentration noted. Mutant 3A (for nomenclature, see the legend to Fig. 3) is a strong site: it competes as well as or better than the WT unlabeled site. Mutant 7A is a weak site: even at a concentration of 10 μM as an unlabeled competitor, it does not compete significantly for binding.

of the concentration, taken from the plot, of unlabeled mutant and WT DNA required to displace half of the protein–^{32}P-labeled DNA complexes directly yields the relative affinity of the mutant DNA site to the WT DNA site for the protein. An unlabeled WT competitor titration is included in every experiment, and comparison of relative affinity between a mutant and a WT site is always done with titrations performed at the same time.

The apparent equilibrium dissociation constant (K_D) of GAL4(1-100) for the WT GAL4 site is calculated as follows. In the absence of competitor, a protein concentration of 2 nM occupies 80% of labeled WT GAL4 DNA sites at a concentration of 20 pM. Addition of unlabeled WT GAL4 DNA sites as competitor DNA at a concentration of 4.2 nM causes half of the labeled DNA sites to dissociate from protein molecules. From the definition $K_D = [P]\,[S]/[PS]$, where $[P]$ is the protein dimer concentration, $[S]$ is the concentration of DNA sites, and $[PS]$ is the concentration of protein dimer-DNA complexes, and the usual carsewatear equations for total protein, $[P]_T = [P] + [PS]$, and total DNA sites, $[S]_T = [PS] + [S]$, we used the data just given to calculate $K_D = 0.5$ nM (5, 7). The apparent K_D was confirmed independently by determining the protein concentration required to occupy 50% of labeled WT DNA sites at a concentration of 20 pM (5, 14).

Predicted affinity for site all 3 out. Site all 3 out (shown in Fig. 5a) is essentially 6G1/2, 7T1/2, and 8A1/2. Since 6G is >2,000 times, 7T is >2,000 times, and 8A is 377 times weaker than the WT site for GAL4, and assuming that the effect of each mutational event is independent from the effects of the others, this site should be the square root of >2,000 × the square root of >2,000 × the square root of 377 (numbers from Fig. 3a). This number is about 38,000.

Yeast strains and manipulations. All yeast strains used (YM709, JPY5, and JPY9) were previously described (2, 27). JPY5 and JPY9 were provided by J. Pearlberg. Yeast transformation was performed by the lithium acetate method as previously described (13).

Number of integrants was controlled for reporters containing the WT site, 16-mer, and 18-mer by the following method. For each reporter, four or more independent colonies representing separate integration events were monitored by the plate assay for β-galactosidase expression. An isolate that had the least intense blue color and a color conforming to the majority of the independent colonies was chosen.

Assays for β-galactosidase. Both the plate and liquid enzyme assays were done by standard methods (12).

RESULTS

The apparent K_D of GAL4(1-100) for a consensus WT 17-bp site is 5×10^{-10} M (under the conditions described in Materials and Methods). The binding affinities, relative to that of the WT DNA site, of mutant sites with all possible doubly symmetric mutations were determined in vitro with purified protein and DNA by an equilibrium competition DNA binding assay (Fig. 2 and Materials and Methods).

Figure 3a shows that binding of GAL4(1-100) is most sensitive to mutations within the palindromic CGG triplets at the ends of the site. This is particularly true for positions 6 and 7. Doubly symmetric substitution of any of the other three possible base pairs at either position 6 or position 7 resulted in undetectable binding by GAL4 (more than 2,000-fold weaker than WT binding). At position 8, doubly symmetric substitutions led to significantly reduced affinity (an average of 342-fold weaker than WT binding). These results suggest that bp 6 to 8, seen to be contacted in the crystal structure, do indeed determine specificity of binding.

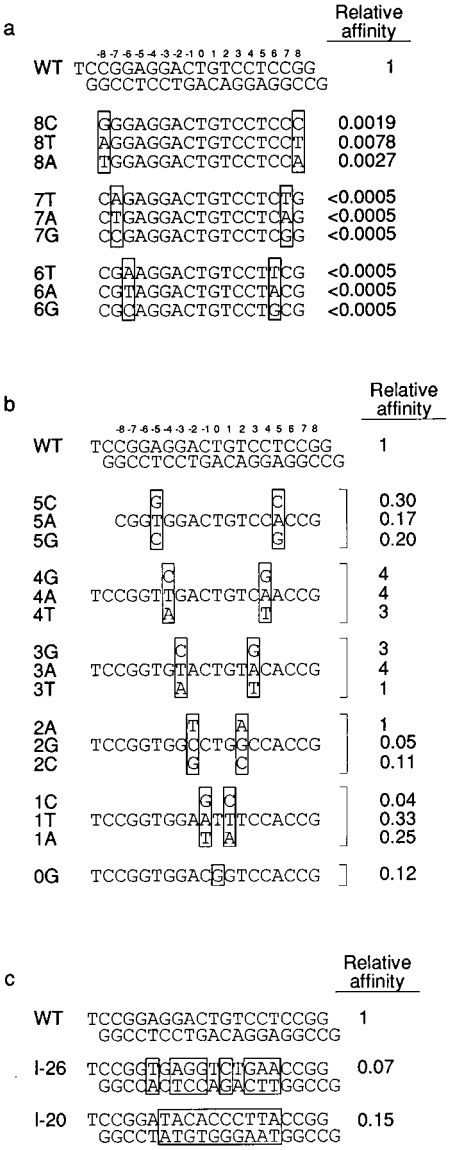

FIG. 3. Summary of the affinity of GAL4(1-100) for the WT GAL4 DNA site and mutant sites consisting of all possible doubly symmetric mutations of the GAL4 site. (a) Mutant sites with doubly symmetric mutations in positions 6 to 8. (b) Mutant sites with doubly symmetric mutations in positions 0 to 5. Affinity is expressed as the affinity of the protein for a mutant site relative to the affinity of the protein for the WT site. Experiments were performed by equilibrium competition gel shift DNA binding assay (Materials and Methods). The 17-bp GAL4 DNA binding site is numbered as follows. Nucleotides are numbered 0 to 8 from the central base in the 5′-to-3′ direction. Nucleotides in the 3′ direction of the central base are indicated by minus signs. The oligonucleotide bearing the WT site is shown in full. The other sites are identical to the WT site except for the mutations as noted. For simplicity, for these mutant sites, only the sense strand and only the GAL4 17-bp portion of the oligonucleotides are shown. Each mutant site with doubly symmetric mutations is named after the mutant nucleotide present in the positive half site. (c) Relative affinities of two mutant sites with multiple changes in the center 11 bp for GAL4(1-100).

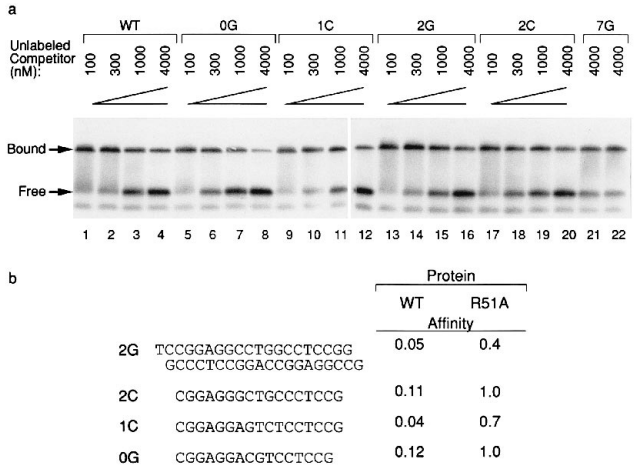

FIG. 4. GAL4(1-100)R51A mutant protein binds a GAL4 17-bp site specifically and does not have DNA sequence preferences in the center 11 bp between the palindromic CGG half sites of the GAL4 site. (a) DNA binding experiment showing that unlabeled WT competitor competes for binding with labeled WT site (at a concentration of 100 pM) for GAL4(1-100)R51A but a mutant site, 7G, does not. The experiment also shows that unlabeled mutant sites 0G, 1G, 2G, and 2C compete efficiently, as well or nearly as well as the WT site. DNA molecules bound to protein were separated from free DNA molecules by native gel electrophoresis. (b) Comparison of affinities of altered sites 2G, 2C, 1C, and 0G relative to WT site for GAL4(1-100) and the mutant R51A.

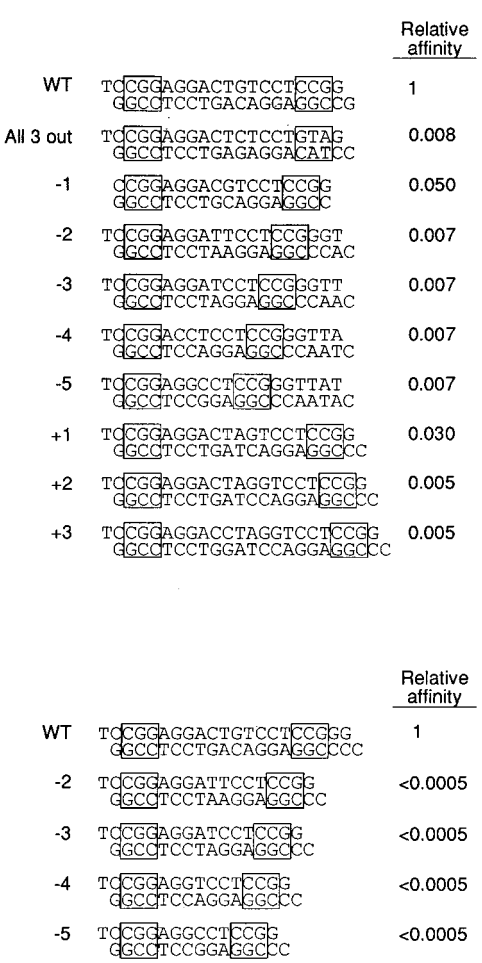

FIG. 5. (a) Summary of affinities of GAL4(1-100) for the WT GAL4 17-bp site, mutant sites with the palindromic CGG half sites separated by 6 bp (−5), 7 bp (−4), 8 bp (−3), 9 bp (−2), 10 bp (−1), 12 bp (+1), 13 bp (+2), and 14 bp (+3), and a mutant site with all three residues of one CGG half site completely mutated (all 3 out). (b) Relative affinities of GAL4(1-100) for the WT GAL4 17-bp site and spacing mutant sites contained in oligonucleotides shorter than those presented in panel a, so that for each CGG triplet, there is no triplet site 11 bp away (because of the length of the oligonucleotide). Under these conditions, a GAL4 dimer cannot simultaneously engage a CGG triplet specifically and any other triplet 11 bp away nonspecifically.

In contrast to mutant sites with doubly symmetric mutations in the CGG triplets at the ends of the 17-bp site, those with all possible doubly symmetric mutations in the middle of the site, at positions 0, 1, 2, 3, 4, and 5, bind GAL4 with relatively high affinity (Fig. 3b). This statement is especially true for sites bearing doubly symmetric mutations at positions 3, 4, and 5. These results are consistent with the observation that the protein does not contact bp 0 to 5 in the cocrystal structure.

Some mutant sites with doubly symmetric mutations in the center (positions 0, 1, and 2) bind GAL4 up to 26-fold less strongly than the WT site (Fig. 3b). These reductions are smaller than those that result from doubly symmetric mutations in the terminal CGG triplet sequences, but they are substantial compared with the effect of substitutions at position 5. The largest effects at positions 0 to 2 result from increases in the G/C content of the central sequence. We expect that such changes would be unfavorable for the way GAL4 constricts the minor groove at the center of the site.

Two DNA sites with multiple substitutions in the central 11 bp were constructed (31). Figure 3c shows that these two sites have relatively high affinity for GAL4(1-100). This result further confirms that the central 11 bp are not contacted by the protein.

The crystalline complex of GAL4(1-65)-DNA shows that the two Arg-51 residues participate in a network of hydrogen bonds, which anchors the dimer across the minor groove at the center of the site. This network includes phosphates 2, 3, −2, and −3, main- and side-chain nitrogens of the arginines, and serine 47. We imagine that in the absence of these contacts, conformational constraints on the DNA would be relieved and sequence preferences in the center would be diminished. We have therefore expressed the mutant protein GAL4(1-100)R51A, in which the Arg-51 residues are changed to alanine. This protein has significantly reduced affinity for a GAL4 site, but it does bind specifically. Unlike the WT protein, it fails to discriminate between the WT DNA site and altered sites with changes at bp 0, 1, or 2. The data are shown in Fig. 4a. Comparison of lanes 4 and 22 shows that the WT site is a much better competitor than the nonspecific 7G site. Figure 4b compares affinities of altered sites 0G, 2G, and 2C (relative to the WT site) for GAL4(1-100)R51A and WT GAL4(1-100). In contrast to the WT protein, the R51A mutant essentially fails to discriminate altered sites from the WT site.

We have also analyzed the consequence of altering the spacing of CGG half sites within the consensus GAL4 DNA site. Figure 5a shows that GAL4 has moderately reduced affinity for sites with the CGG half sites separated by 10 (−1 site) or 12 bp (+1 site) instead of the WT spacing of 11 bp but that it has very low affinity for spacing of 13 or 14 bp and spacings between 6 and 9 bp. We suggest that a GAL4(1-100) homodimer binds to the altered sites with various spacings between their palindromic CGG triplets by making specific contacts to one CGG triplet and nonspecific contacts to whatever triplet lies 11 bp away. A site (all 3 out [Fig. 5a]) that we specifically engineered so as to have all three residues of one half site completely

mutated is reduced in binding by about the same amount as any of the sites with altered spacing, except for ±1 (Fig. 5a).

Figure 5b shows that mutant sites with CGG triplets spaced by 6 to 9 bp embedded in short oligonucleotides such that there is no triplet sequence 11 bp away from one CGG sequence have very little affinity for GAL4(1-100). Thus, binding of one Zn cluster domain with no possible contact for the other is weaker than the detection limit of our experiment.

We have also analyzed the DNA sequence preference of another Zn_2Cys_6 binuclear cluster family member, the *S. cerevisiae* transcriptional activator protein PPR1. The PPR1 consensus site is reported as a twofold symmetric 16-mer with the sequence $TTCGGN_6CCGAA$ (17, 30). That is, the GAL4 and PPR1 recognize the same $CGGN_XCCG$ site, but each requires a distinct X.

PPR1(29-123), a fragment that was used in cocrystallographic studies (25), is a dimer in solution and binds to the high-affinity site from the *URA3* promoter (18, 30) (Fig. 4a) or a consensus site (Fig. 4b) with an apparent K_D of 5 nM. The 29 N-terminal residues of PPR1 were removed because they contain a high concentration of basic residues, and the strong positive charge prevents the protein-DNA complex from entering a gel under the conditions used. PPR1(1-123) and PPR1(29-123) exhibit identical dissociation constants and DNase I footprint patterns, as assayed by quantitative DNase I footprinting (data not shown).

Figure 6 shows that any DNA site containing a doubly symmetric mutation in the CGG triplet has low affinity for PPR1(29-123), while any site with a doubly symmetric mutation in the 6 bp between the palindromic CGG triplets has high affinity for PPR1(29-123). The highly conserved TT sequences just 5′ to the CGG in each half site can also be altered with little effect. Thus, only the CGG residues within the PPR1 DNA site are crucial for tight binding by PPR1, but unlike the case for GAL4, the sequence of the 6 bp between the two palindromic CGG half sites is completely irrelevant. This result is consistent with the crystal structure, which shows neither any base pair contacts in the center nor any sugar-phosphate contacts that might constrain the DNA conformation.

The consequences of altering the spacing of CGG half sites within the consensus PPR1 DNA site are shown in Fig. 7. PPR1 has low affinity for a GAL4 site with a spacing of 11 bp between the palindromic CGG half sites, as well as for sites in which the CGG half sites are separated by 5 (−1 site) or 4 (−2 site) bp. The cocrystal structure with DNA shows that the PPR1 linker folds into an antiparallel beta-ribbon and is not in contact with DNA (20). The linker from one subunit of the homodimer makes hydrophobic contacts with the dimerization domain, and additional hydrophobic interactions are observed between the dimerization and Zn domains. These structural features of PPR1 appear to hold the linker to this conformation, not allowing for flexibility.

In vitro binding experiments described above demonstrated that GAL4 binds a 16-mer or an 18-mer (with the CGG triplets separated by 10 or 12 bp, respectively, instead of the WT spacing of 11 bp) with roughly the same affinity, about 20- to 30-fold lower than that of the WT 17-mer. This reduction seems modest and suggests that in the cell, these may be target sites for GAL4. We have carried out experiments that show that an 18-mer is a fairly strong site in vivo but a 16-mer is a poor site. A *GAL1-lacZ* reporter gene bearing a 16-mer at −100 from TATA is completely inactive at physiological levels of GAL4. When GAL4 is overexpressed from the strong *ADH1* promoter, the 16-mer directs expression at about 1% of that found with a WT 17-mer under the same conditions (Fig. 8a). In contrast, a single 18-mer and the WT 17-mer, in the same

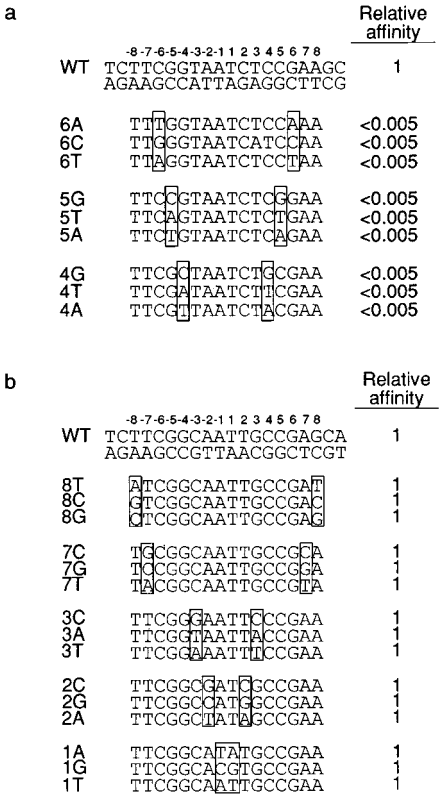

FIG. 6. Summary of the affinities of PPR1(29-123) for the WT PPR1 DNA site and mutant sites consisting of all possible doubly symmetric mutations of the PPR1 site. (a) Mutant site with doubly symmetric mutations in positions 4 to 6. The WT site is the high-affinity site from the *URA3* promoter. (b) Mutant sites with doubly symmetric mutations in positions 1 to 3 and positions 7 and 8. The WT site is the consensus PPR1 site. This site and the site from the *URA3* promoter have the same affinity for PPR1(29-123) (data not shown). Experiments were performed by equilibrium competition gel shift DNA binding assay (Materials and Methods). As in Fig. 3, only the WT site is shown in full. The mutant sites are identical to the WT site except for the mutations as shown. For simplicity, these sites are shown as 16-mers and only the sense strands are shown. The PPR1 DNA binding site is numbered as follows. Nucleotides are numbered 1 to 8 from the central base in the 5′-to-3′ direction. Nucleotides in the 3′ direction of the central base are indicated by minus signs. Only the sense strand is shown for simplicity. Also, each mutant site with doubly symmetric mutations is named after the mutant nucleotide present in the positive half site.

context, are induced by galactose between 1,300- and 2,400-fold at physiological levels of GAL4 (Fig. 8a). At physiological levels of GAL4, reporter genes with two copies of the 18-mer at −94, −191, or −150 from TATA are at least as active as two copies of the WT 17-mer site at the same distances from TATA (Fig. 8b and c), while reporter genes with two copies of a 16-mer at these distances from TATA are only weakly active (Fig. 8b and c). The activities of all these reporter genes are completely dependent on GAL4 (Fig. 8b and data not shown).

The 16-mer and 18-mer sites have about the same affinity for GAL4, yet they behave completely differently in vivo. One possible explanation for the difference between the two systems is that the in vitro experiments were done with a polypeptide fragment of GAL4, GAL4(1-100), and the in vivo experiments were done with full-length GAL4. We therefore used derivatives of a similar GAL4 fragment, GAL4(1-147), in the in vivo experiments. As shown in Fig. 9, the activities of GAL4(1-147)+I+II (22) and GAL4(1-147)+B17 (23) were qualitatively the same as that of full-length GAL4: they activated from a 16-mer site very weakly compared with their

		Relative affinity
WT PPR1 site 1	TCTTCGGTAATCTCCGAAGC AGAAGCCATTAGAGGCTTCG	1
WT PPR1 site 2	TTCGGTAACTCCGAA AAGCCATTGAGGCTT	1
-1 site	TTCGGTAACTCCGAA AAGCCATTGAGGCTT	<0.02
-2 site	TTCGGTACTCCGAA AAGCCATGAGGCTT	<0.02
All 3 out	TTCGGTAATCTGTAAA AAGCCATTAGACATTT	<0.01
GAL4 site	TCCGGAGGACTGTCCTCCGG GGCCTCCTGACAGGAGGCCG	<0.02

FIG. 7. Summary of affinities of PPR1(29-123) for the WT PPR1 site, mutant sites with the palindromic CGG half sites separated by 4 bp (−2), 5 bp (−1), and 11 bp (GAL4 site), and a mutant site with all three residues of one CGG half site completely mutant (all 3 out). Experiments were performed by equilibrium competition gel shift DNA binding assay (Materials and Methods).

activities on the 17-mer and 18-mer. GAL4(1-147)+I+II is as strong an activator on each of the templates tested. GAL4(1-147)+B17 is weaker than full-length GAL4 and not disproportionally stronger than GAL4 on a 16-mer. Control experiments show that GAL4(1-147) alone, lacking an activation domain, is inactive on all three templates. Thus, there appears to be an additional mechanism in the cell that prevents GAL4 activating from a 16-mer. It is possible that GAL4 has a much weaker affinity for a 16-mer in DNA wrapped in nucleosomes than for a 17-mer or 18-mer under the same condition. It has been reported that GAL4 has high affinity for a single 17-mer in nucleosomal DNA in vitro (34).

DISCUSSION

Using an in vitro equilibrium competition DNA binding assay, we have shown that the palindromic CGG triplets at the ends of the GAL4 17-bp consensus DNA site are essential for tight binding by GAL4(1-100) and that the identities of the 11 bp between the palindromic CGG triplets are less important. These results are in agreement with previously published data on the *Kluyveromyces lactis* GAL4 homolog, LAC9 (10). They are also in accord with the structure of GAL4-DNA complex (24), which reveals base-pair-specific contacts to CGG triplet sequences at the ends of the DNA site.

We found that GAL4(1-100) has a modest (10- to 30-fold) reduction in affinity for several mutant sites with doubly symmetric mutations in the center at bp 0, 1, and 2. Marmorstein et al. (24) noted that in the GAL4-DNA structure, the minor groove is constricted by 3 Å relative to average B-form DNA in the center of the site. The constriction appears to be imposed by multiple hydrogen bonds to the phosphate backbone at positions 2, 3, −2, and −3 from Arg-51. This arginine residue also contributes to the apparent rigidity of the amino-terminal element and of the dimerization element (residues 50 to 65). It donates a hydrogen bond to the main-chain carbonyl of serine 47 on the partner chain, thereby straddling the base of the coil-coil and fixing the position of the guanidinium group. This structure straddles the minor groove and requires a defined spacing for the contacted phosphates. We suggest that the effects of changes at bp 0, 1, and 2 can be explained by the crystallographically observed minor groove constriction, which can occur more readily with some DNA sequences than with others. Consistent with this interpretation is an observation that a mutant protein, GAL4(1-100)R51A, does not discriminate sites mutated in the innermost 5 bp from the WT site. Examination of naturally occurring GAL4 sites (reviewed in reference 28) shows that of the central 11 bp, positions 0, 1, and 2 are the least variable. This sequence requirement for base pairs whose functional groups are uncontacted is reminiscent of the interaction between bacteriophage 434 repressor and Cro with their operators (19). In that case, there is a strong

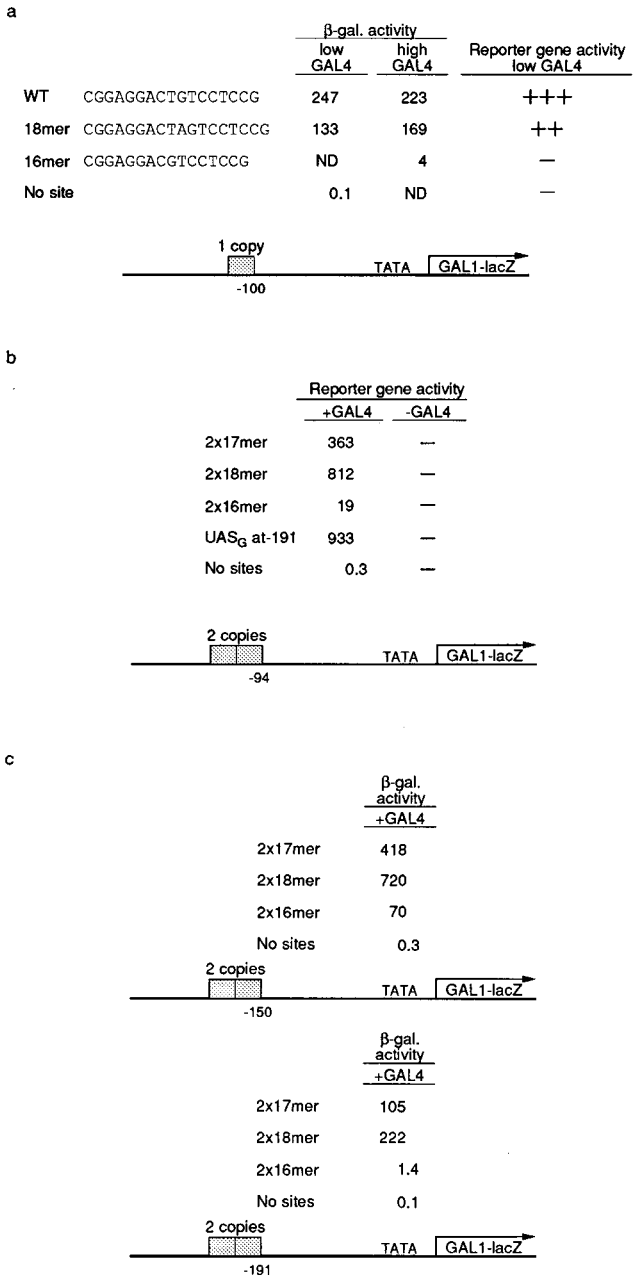

FIG. 8. Summary of in vivo experiments showing activities of GAL4 on *GAL1-lacZ* reporter genes bearing WT or mutant GAL4 sites with the palindromic CGG half sites separated by 10 bp (16-mer) or by 12 bp (18-mer). Activity indicated as low GAL4 or +GAL4 is the level of GAL4 protein in the cell expressed from the native GAL4 gene in the yeast chromosome under inducing conditions (galactose, glycerol, and ethanol as carbon sources); high GAL4 is the level of GAL4 protein with the gene encoded on an ARS-CEN plasmid expressed from the *ADH1* promoter; −GAL4 represents activity in assays done under repressing, noninducing condition (glucose as the sole carbon source). β-gal., β-galactosidase.

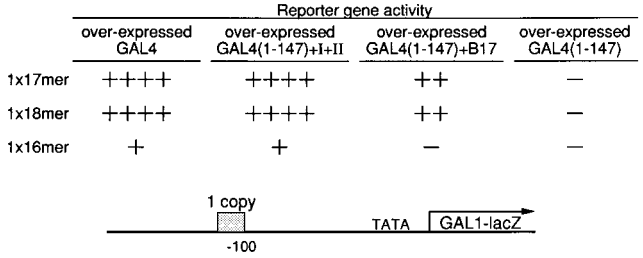

FIG. 9. Summary of in vivo experiments showing activities of GAL4 or GAL4 derivatives on *GAL1-lacZ* reporter genes bearing WT or mutant GAL4 sites with the palindromic CGG half sites separated by 10 bp (16-mer) or by 12 bp (18-mer). "Over-expressed" is the level of protein expressed from the *ADH1* promoter with the gene encoded on ARS-CEN plasmid.

preference in the center of the operator for T/A base pairs over G/C base pairs. In the case of GAL4, the preference of T/A base pairs over G/C base pairs is also true for positions 0 and 2 but not for position 1.

Is it possible that parts of the GAL4 protein missing in our experiment contacts the middle bases? One line of evidence against this idea is that purified full-length GAL4 protein binds with roughly the same affinity as GAL4(1-147) to a near consensus site in vitro (16). Furthermore, the contacts observed in the crystalline GAL4(1-65)-DNA complex account for the results of chemical and enzymatic footprinting experiments using GAL4(1-147) (5) or full-length GAL4 (8, 9). It is also possible that the remainder of GAL4 protein is required to restrict the spacing requirement further. We tested full-length GAL4 or GAL4(1-147) plus various acidic activation domains in vivo and found that GAL4 full-length protein has a spacing requirement similar to that of a smaller polypeptide fragment similar to the one used in our in vitro studies.

PPR1 is a protein that is closely related to GAL4, but only in the short stretch of amino acid residues that form the Zn_2Cys_6 binuclear cluster (reviewed in reference 25). Using the same equilibrium competition assay, we demonstrated that the palindromic CGG triplets at the ends of the PPR1 12-bp site are crucial for tight binding by PPR1(29-123) and that all DNA sites bearing mutations in the inner six positions between the palindromic CGG triplets have very high affinity for PPR1. The two TT residues just 5′ of the CGG triplet, although highly conserved, are not mutationally sensitive. This study suggests that PPR1 recognizes the 12-bp sequence $CGGN_6CCG$ and not the 16-bp sequence $TTCGGN_6CCGAA$ as previously proposed (18, 30). The reason that these two T residues are conserved is not clear; perhaps they form part of a site for another protein.

Although GAL4 and PPR1 both bind DNA sites which contain inverted CGG sequences, they each have a distinct preference for the number of base pairs separating the half sites. For GAL4, the preferred separation of CGG half sites is 11 bp. Sites with separations of 10 or 12 bp are reduced in affinity by 20- to 30-fold, while sites with more drastic separations are reduced by significantly more. The preferred half-site spacing between palindromic CGG half sites for PPR1 is 6 bp. DNA sites with spacings of 4, 5, or 11 bp bind PPR1 poorly. In the case of GAL4, the crystalline complex with DNA shows that a nine-amino-acid linker between the Zn domain and the dimerization segment is in an extended conformation and contacts the phosphate backbone of the DNA (24). It appears that this linker has the flexibility to accommodate a 16-bp (−1) site or a 18-bp (+1) site but not sites that have spacing between the CGG sequences of less than 10 bp or more than 12 bp. PPR1, on the other hand, does not tolerate incorrect spacing between the CGG sequences. The cocrystal structure with DNA shows that the PPR1 linker folds into an antiparallel beta-ribbon and is not in contact with DNA (25). The linker from one subunit of the homodimer makes hydrophobic contacts with the dimerization domain, and additional hydrophobic interactions are observed between the dimerization and the Zn domains. These structural features of PPR1 appear to hold the linker in this conformation, diminishing its flexibility and its capacity to adjust to sites of different length.

It is noteworthy that in vivo, GAL4 has an absolute requirement for sites with spacing between the palindromic CGG triplets of greater than 10 bp and less than 13 bp—a 16 bp (−1) site is a poor site in vivo, whereas an 18-bp (+1) site, with about the same affinity for GAL4(1-100) in vitro as the 16-bp site, is nearly as strong a site as the 17-bp WT site. Similar results were obtained by another group (35). Apparently, an additional mechanism exists in the yeast cell that restricts the spacing requirement of GAL4.

It also appears that a GAL4(1-100) homodimer binds cooperatively to a DNA site that contains one CGG sequence and any triplet sequence 11 bp away. Such a site is significantly reduced in affinity for GAL4, about 150-fold less than WT. We imagine that in this instance, GAL4 makes specific contacts to one half site and nonspecific contacts to the mutant half site.

There are at least 12 fungal transcriptional regulatory proteins containing domains with a Zn_2Cys_6 binuclear cluster. The DNA binding sites for some of the other members of the family are not well characterized. Not all of the proteins with known DNA sites have sites each with palindromic CGG half sites separated by a distinct spacing. GAL4, PPR1, PUT3, and LAC9 are the only ones for which such a site has been established. The regions of these four proteins equivalent to the DNA recognition domain of GAL4 are extremely similar in amino acid sequence (Fig. 1a). Lysines 17 and 18 of GAL4, which make the specific contacts to CGG, are conserved except in PUT3, in which they are replaced by Arg and His, respectively (Fig. 1a). The Arg and His in PUT3 could engage in contacts homologous to those made by the lysyl 17 and 18 residues in GAL4. In fact, GAL4(1-100)K18H binds to a GAL4 site with high affinity (21a; unpublished observation). Thus, we reemphasize the proposal (29, 38) that the determinant of specificity for these proteins does not lie in the DNA recognition domain alone, which is required to contact the CGG directly, but also involves structural elements that determine the preferred spatial relationship between the two Zn_2Cys_6 domains in a dimer bound to the correct DNA sequence. In GAL4 and PPR1, the structural elements are the linker and the part of the dimerization element just C terminal to the linker, as suggested by genetic data (6, 18, 38) and confirmed by domain swap experiments (29) and structural studies (24, 25). These elements adopt notably different conformations in these two proteins. There are proteins with Zn cluster domains whose sites are not CGG half sites separated by a distinct number of base pairs. We imagine that the Zn domains of these proteins may contact DNA differently or that they are used for other purposes.

ACKNOWLEDGMENTS

We are grateful for stimulating discussion by members of the Ptashne laboratory.

This work was funded by grants from the NIH to M.P. and S.C.H.

REFERENCES

1. **Baleja, J. D., R. Marmorstein, S. C. Harrison, and G. Wagner.** 1992. Solution structure of the DNA-binding domain of Cd_2-GAL4 from *S. cerevisiae*. Nature (London) **356:**450–453.

2. **Barberis, A., J. Pearlberg, N. Simkovich, S. Farrell, P. Reinagel, C. Bambad, G. Sigal, and M. Ptashne.** 1995. Contact with the Pol II holoenzyme suffices for gene activation. Cell **81:**359–368.
3. **Bram, R., and R. Kornberg.** 1985. Specific protein binding to far upstream activating sequences in polymerase II promoters. Proc. Natl. Acad. Sci. USA **82:**42–47.
4. **Bram, R., N. Lue, and R. Kornberg.** 1986. A GAL family of upstream activating sequences in yeast: roles in both induction and repression of transcription. EMBO J. **5:**603–608.
5. **Carey, M., H. Kakidani, J. Leatherwood, F. Mostashari, and M. Ptashne.** 1989. An amino-terminal fragment of GAL4 binds DNA as a dimer. J. Mol. Biol. **209:**423–432.
6. **Corton, J. C., and S. A. Johnston.** 1989. Altering DNA-binding specificity of GAL4 requires sequences adjacent to the zinc finger. Nature (London) **340:**724–727.
7. **Fried, M., and D. M. Crothers.** 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. **9:**6505–6525.
8. **Giniger, E., and M. Ptashne.** 1987. Transcription in yeast activated by a putative amphilpathic alpha helix linked to a DNA binding unit. Nature (London) **330:**670–672.
9. **Giniger, E., S. Varnum, and M. Ptashne.** 1985. Specific DNA binding of GAL4, a positive regulatory protein of yeast. Cell **40:**767–774.
10. **Halvorsen, Y.-D. C., K. Nandabalan, and R. C. Dickson.** 1991. Identification of base and backbone contacts used for DNA sequence recognition and high-affinity binding by LAC9, a transcription activator containing a C6 zinc finger. Mol. Cell. Biol. **11:**1777–1784.
11. **Harrison, S. C.** 1991. A structural taxonomy of DNA-binding domains. Nature (London) **353:**715–719.
12. **Himmelfarb, H., J. Pearlberg, D. Last, and M. Ptashne.** 1990. GAL11P: a yeast mutation that potentiates the effect of a weak GAL4-derived activator. Cell **63:**1299–1309.
13. **Ito, H., Y. Fukuda, K. Murata, and A. Kimura.** 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153:**163–168.
14. **Johnson, A. D.** 1979. Mechanism of action of the lambda Cro protein. Ph.D. thesis. Harvard University, Cambridge, Mass.
15. **Johnston, M.** 1987. A model fungal gene regulatory mechanism: the *GAL* genes of *Saccharomyces cerevisiae*. Microbiol. Rev. **51:**458–476.
16. **Kang, T., T. Martins, and I. Sadowski.** 1993. Wild-type GAL4 binds cooperatively to the GAL1-10 UAS_G *in vitro*. J. Biol. Chem. **268:**9629–9635.
17. **Keegan, L., G. Gill, and M. Ptashne.** 1986. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science **231:**699–704.
18. **Keegan, W. P.** 1989. Studies on the DNA-binding of the yeast GAL4 protein. Ph.D. thesis. Harvard University, Cambridge, Mass.
19. **Koudelka, G. B., S. C. Harrison, and M. Ptashne.** 1987. Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and cro. Nature (London) **326:**886–888.
20. **Kraulis, P. J., A. R. C. Raine, P. L. Gadhavi, and E. D. Laue.** 1992. Structure of the DNA-binding domain of zinc-GAL4. Nature (London) **356:**448–450.
21. **Kunkel, T.** 1985. Rapid and efficient site-specific mutations without phenotypic selection. Proc. Natl. Acad. Sci. USA **82:**488–492.

21a. **Liang, S., and M. Ptashne.** Unpublished observation.

22. **Ma, J., and M. Ptashne.** 1987. Deletion analysis of GAL4 defines two transcriptional activating segments. Cell **48:**847–853.
23. **Ma, J., and M. Ptashne.** 1987. A new class of yeast transcriptional activators. Cell **51:**113–119.
24. **Marmorstein, R., M. Carey, M. Ptashne, and S. Harrison.** 1992. DNA recognition by GAL4: structure of a protein/DNA complex. Nature (London) **356:**408–414.
25. **Marmorstein, R., and S. C. Harrison.** 1994. Crystal structure of a PPR1-DNA complex: DNA recognition by proteins containing a Zn_2Cys_6 binuclear cluster. Genes Dev. **8:**2504–2512.
26. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
27. **Pearlberg, J.** 1995. Studies of the yeast transcription factor GAL11. Ph.D. thesis. Harvard University, Cambridge, Mass.
28. **Ptashne, M.** 1992. A genetic switch, 2nd ed. Cell Press, Cambridge, Mass., and Blackwell, Palo Alto, Calif.
29. **Reece, R. J., and M. Ptashne.** 1993. Determinants of binding-site specificity among yeast C_6 zinc cluster proteins. Science **261:**909–911.
30. **Roy, A., F. Exinger, and R. Losson.** 1990. *cis*- and *trans*-acting regulatory elements of the yeast *URA3* promoter. Mol. Cell. Biol. **10:**5257–5270.
31. **Ruden, D. M.** 1990. Transcriptional activation by acidic activators in yeast. Ph.D. thesis. Harvard University, Cambridge, Mass.
32. **Sadowski, I., B. Bell, P. Broad, and M. Hollis.** 1992. GAL4 fusion vectors for expression in yeast or mammalian cells. Gene **118:**137–141.
33. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
34. **Taylor, I. C. A., J. L. Workman, T. J. Schuetz, and R. E. Kingston.** 1991. Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains. Genes Dev. **5:**1285–1298.
35. **Vashee, S., H. Xu, S. A. Johnston, and T. Kodadek.** 1993. How do Zn_2Cys_6 proteins distinguish between similar upstream activation sites? J. Biol. Chem. **268:**24699–24706.
36. **Webster, N., J. R. Jin, S. Green, M. Hollis, and P. Chambon.** 1988. The yeast UAS_G is a transcriptional enhancer in human HeLa cells in the presence of GAL4 trans-activator. Cell **52:**169–178.
37. **West, R. W., Jr., R. R. Yocum, and M. Ptashne.** 1984. *Saccharomyces cerevisiae GAL1-GAL10* divergent promoter region: location and function of the upstream activating sequence UAS_G. Mol. Cell. Biol. **4:**2467–2478.
38. **Witte, M. M., and R. C. Dickson.** 1990. The C6 zinc finger and adjacent amino acids determine DNA-binding specificity and affinity in the yeast activator proteins LAC9 and PPR1. Mol. Cell. Biol. **10:**5128–5137.
39. **Wood, K.** 1989. Transcriptional synergism by region IIA of the GAL4 protein. Undergraduate thesis, Harvard University.