The yeast activator HAP1—a GAL4 family member—binds DNA in a directly repeated orientation

Li Zhang and Leonard Guarente

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02129 USA

The yeast transcriptional activator HAP1 contains a DNA-binding domain homologous to GAL4, PPR1, and related factors. By selecting random HAP1-binding sites, we found that HAP1, like GAL4, binds to two CGG triplets. Unlike GAL4, the CGGs in the HAP1 consensus are in a direct and not inverted orientation. Sites with inverted CGGs were not recovered, and mutations converting the direct repeat of CGGs to an inverted repeat greatly reduce HAP1-binding affinity. Also, the 6-bp spacer between the CGGs contains a consensus TA that is positioned asymmetrically. Dimethylsulfate protection patterns on six of these sites show protections and enhancements that also lie in a directly repeated orientation, suggesting that the two HAP1 DNA recognition domains of a HAP1 homodimer are oriented in a directly repeated configuration on the DNA. Moreover, substitution of the HAP1 dimerization domain with that of PPR1, which forms coiled-coils and dimerizes symmetrically, did not diminish the ability of the protein to bind selectively to a direct repeat. This result suggests that one DNA-binding domain of the HAP1 homodimer must be able to swivel 180° relative to the dimerization domain to make specific contacts with the second CGG triplet. Our results present a novel example of domain swiveling in one of the two identical subunits of a homodimer to accommodate specific DNA contacts to both CGG triplets of a direct repeat.

[*Key Words:* Transcriptional activator; HAP1; asymmetric binding; direct repeat; homodimer; domain swiveling]

Received April 13, 1994; revised version accepted July 11, 1994.

The yeast transcriptional activator HAP1 is a member of the GAL4 family whose DNA-binding domain contains a highly conserved zinc cluster domain and a coiled-coil dimerization domain (Johnston 1987; Creusot et al. 1988; Pfeifer et al. 1989; Zhang et al. 1993). Data from x-ray crystallography and other experiments show that the GAL4 dimer recognizes a symmetrical 17-bp sequence containing an inverted repeat of two CGG triplets separated by 11 nucleotides (CGGn11CCG; Carey et al. 1989; Marmorstein et al. 1992). The DNA-binding sequence of GAL4 contains a zinc cluster domain that recognizes CGG, a coiled-coil dimerization domain, and a linker that connects the two (Marmorstein et al. 1992). A GAL4 dimer binds to the inverted CGGs symmetrically. PUT3, PPR1, and LEU3, three other members of the GAL4 family (Friden and Schimmel 1987; Zhou et al. 1987; Siddiqui and Brandriss 1989; Roy et al. 1990; Reece and Ptashne 1993), all recognize a DNA site containing a rotationally symmetrical inverted repeat of two CGG triplets, separated by 10, 6, and 4 bp, respectively (Friden and Schimmel 1988; Reece and Ptashne 1993). Recent experiments show that PPR1 and PUT3 use their zinc cluster domains to recognize the CGG triplets like GAL4 (Reece and Ptashne 1993). Furthermore, the zinc cluster domains of GAL4, PUT3, and PPR1 are identical

in their ability to recognize CGG triplet DNA sequences. The linker and the start of the dimerization domain are responsible for determining the spacing of the CGG triplets of its preferred site (Reece and Ptashne 1993).

HAP1 is a 1483-residue protein containing a C6, GAL4-like zinc cluster DNA-binding domain (residues 1–148), a heme-responsive domain (residues 244–444), an acidic activation domain (residues 1309–1483), and a region of undetermined function (residues 445–1309) (Creusot et al. 1988; Pfeifer et al. 1989). The HAP1 zinc cluster domain is homologous to that of the GAL4 family (the proteins share 45% identity across this region), and the key residues in GAL4 responsible for specific base contacts are conserved in HAP1 (Johnston 1987; Marmorstein et al. 1992). Furthermore, the HAP1 dimerization domain contains a coiled-coil sequence similar to that within the dimerization domain of GAL4 (Zhang et al. 1993).

However, HAP1 DNA-binding specificity appears to differ from that of GAL4. HAP1 binds diverse DNA elements such as the upstream activating sequence 1 (UAS1) of *CYC1*, the UAS of *CYC7*, and the UAS of *CYT1* (Pfeifer et al. 1987; Prezant et al. 1987; Zitomer et al. 1987; Schneider and Guarente 1991). The UAS of

CYC7 contains a direct repeat of TCGCTATTA while the UAS of *CYT1* contains an inverted repeat of GCGGCCGG (Pfeifer et al. 1987; Schneider and Guarente 1991). These UASs share no apparent homology, and HAP1 transcriptional activity at UAS1 versus the UAS of *CYC7* differs significantly although the protein binds equally well to all the sites (Pfeifer et al. 1987; Kim and Guarente 1989). Thus, it is of great interest to determine the critical elements for HAP1 binding and how HAP1 binds to these diverse DNA elements. These studies will allow us to gain insights into molecular mechanisms governing differential transcriptional activation at various HAP1-binding sites.

Another feature of HAP1 that distinguishes it from other members of the GAL4 family is that its DNA-binding activity requires heme (Pfeifer et al. 1987). In the absence of heme, the heme regulatory domain represses DNA binding (Pfeifer et al. 1989) by preventing dimerization (Zhang et al. 1993). This process is mediated by the binding of one or more cellular factors to the heme regulatory domain in the absence of heme, thereby repressing its DNA-binding and transcriptional activity (Fytlovich et al. 1993; Zhang and Guarente 1994). Upon the addition of heme, this repression complex disassembles and HAP1 dimerizes and binds to DNA with high affinity.

To compare the DNA-binding specificity of HAP1 with other members of the GAL4 family and to understand how HAP1 binds to diverse DNA elements, we decided to carry out a systematic study on HAP1 DNA binding. Our results show that like the other GAL4 family proteins, the HAP1 zinc cluster domain recognizes CGG triplets, but unlike the other GAL4 family proteins, the HAP1 site is an asymmetric, direct repeat of two CGGs instead of a symmetrical, inverted repeat of CGGs. The two DNA-binding domains (DBDs) of HAP1 dimers are positioned asymmetrically to make the same contacts at both CGG triplets. Furthermore, domain-swapping experiments show that one DBD of the HAP1 homodimer swivels 180° relative to the dimerization domain to maintain specific DNA-protein interactions with both CGG triplets.

Results

HAP1 homodimers selectively bind to asymmetric sites

To study HAP1-DNA interactions, we purified two GST-HAP1 fusion proteins containing residues 1–171 and 52–171 (Guan and Dixon 1991). The GST moiety was subsequently cleaved off the fusion proteins with thrombin. Both HAP1 fragments exhibited similar DNA-binding properties, with the longer fragment having slightly higher affinity toward DNA. Most experiments described below were carried out using the fragment 1–171. To select for optimal HAP1-binding sites, we employed a random oligonucleotide-polymerase chain reaction (PCR) selection technique (see Fig. 1; Pollock and Treisman 1990). Briefly, a 76-bp DNA fragment contain-

A 76 bp oligonucleotide with 26 random nucleotides was synthesized for the selection of HAP1 sites

Figure 1. An outline of the procedure for the selection of HAP1-binding sites.

ing 26 random nucleotides in the center bounded by restriction sites, and a primer site for PCR on each side, was synthesized and labeled with [³²P]dCTP nucleotides. Purified HAP1 fragment 1–171 was incubated with the synthesized DNA fragments containing random nucleotides, and bound DNA was separated from unbound DNA by gel electrophoresis. The bound DNA was then cut out from the dried gels and amplified directly by PCR. The amplified DNA was selected further by binding with purified HAP1 again. After four cycles of HAP1-binding selection and PCR amplification, the selected DNA was cloned into Bluescript KS+ vector (Stratagene) and single-stranded DNA was generated and sequenced. Figure 2 shows the alignment of all the selected sites along with the natural sites, including the UASs of *CYC1*, *CYT1*, *CTT1*, and *CYC7*. A consensus emerges from the 27 selected sites, CGGnnnTAnCGG (the number of appearance in the 27 sites for each conserved nucleotide is shown in Figure 2 by the bar graph). This consensus contains three elements, two CGG triplets and a 6-bp spacer with TA positioned asymmetrically within the spacer. This result shows that like GAL4, the HAP1 zinc cluster domain recognizes the CGG triplet, consistent with the high homology of the zinc clusters (Johnston 1987). Strikingly, HAP1 prefers an asymmetric site containing directly repeated CGG triplets rather than a symmetrical site.

All natural sites fit this consensus except the UAS of *CYC7*, which contains two copies of CGC rather than CGG. Furthermore, nine selected sites, including sites 2, 6, 9, 16, 18, 19, 20, 31, and 81, as well as the UAS of *CYC7*, contain a more specific TTAT sequence in the spacer (most other sites contain only a TA or T sequence). There are two copies of the TATTA sequence in the UAS of *CYC7*, which may provide additional interactions to compensate for the change of CGG → CGC.

Figure 2. Alignment of all the selected HAP1-binding sites and natural HAP1 sites including UAS1/CYC1, UAS/CYT1, UAS/CTT1, and UAS/CYC7. The consensus sequence is shown (*top*), and the frequency of appearance by each conserved nucleotide among the 27 selected sites is shown by the bar graph. The underlined nucleotides in each site are those that occur in the consensus. The TTAT sequence in the spacer is also underlined in sites 2, 6, 9, 16, 18, 19, 20, 31, 81, and the UAS/CYC7. The uppercase letters represent the nucleotides in the 26-bp random oligonucleotide region. The lowercase letters represent the flanking sequences outside the random region.

All selected sites were shown to bind to DNA specifically. Figure 3 shows the binding of some selected sites, 10 (lane 4), 23 (lane 6), 68 (lane 8), 75 (lane 10), 79 (lane 12), and 18 (lane 14), and the UAS1 of *CYC1* (lane 2) by purified HAP1 fragment 1–171. These sites bind to HAP1 with similar affinity except for site 18, which binds to HAP1 with significantly higher affinity. Because site 18 contains two perfect CGG repeats as well as the TTAT sequence in the spacer, it is expected to have the highest affinity for HAP1. To demonstrate further that HAP1 binds specifically to the region containing the consensus sequence, we carried out DNase I footprinting analysis. Figure 4 shows the DNase I footprinting patterns of HAP1 on sites 10 (cf. lanes 2 and 3 with 1), 18 (cf. lanes 5 and 6 with 4), and 23 (cf. lanes 8 and 9 with 7), as well as the UAS of *CYC7* (cf. lanes 11 and 12 with 10). HAP1 binds only to the region containing the consensus DNA element. The protected sequences are CCACGATCGTATCGGCATTAGCATA for site 10, ACACGACTTATCGGTCTGTC for site 18, GCGGTAGT-TACGGACATATC for site 23, and CTTCGCTAT-TATCGCTATTA for the UAS/*CYC7*, respectively. These results confirmed that all of the selected sites bind to HAP1 specifically.

The HAP1-binding site consensus shows that HAP1 recognizes a direct repeat of two CGGs with a 6-bp nnn-TAn spacer. This contrasts with PPR1, which recognizes a palindromic, inverted repeat of two CGGs with a 6-bp spacer (Fig. 7A, below; Roy et al. 1990; Reece and Ptashne 1993). To obtain further evidence that HAP1 recognizes a direct repeat and to demonstrate that the sequence TA in the spacer plays a role in HAP1 binding, we mutated the highest affinity HAP1-binding site, number 18, and tested the effect on HAP1-binding affinity. Briefly, we synthesized four oligonucleotides. The first contains wild-type sequence from site 18, CGGACTTATCCG (HC1). The second contains a G → C mutation in the second CGG, CGGACTTATCCG (M1), thereby generating an inverted repeat of CGGs. The third contains a T → C mutation in the conserved T (26 out of 27) in the spacer, CGGACTCATCCG (M2). The fourth combines the mutations in M1 and M2, CGGACTCATCCG (M3). We tested the HAP1-binding affinity of all the oligonucleotides by competition assays. Figure 5 shows the competition of HAP1 binding to the labeled UAS of *CYC7* by various unlabeled oligonucleotides. Lanes 2–5 show the titration of HAP1 binding to the UAS/*CYC7* by the same oligonucleotide, and lanes 7–10 show the titration by the wild-type HC1 oligonucleotide. Both oligonucleotides competed effectively for HAP1 binding as expected, and the HC1 competed even better than the UAS/*CYC7* oligonucleotide because of its higher affinity. Lanes 11–14 and lanes 15–18 show the titration patterns of M1 and M2 mutant oligonucleotides, respectively. It is clear that both mutants competed much less (~10-fold) effectively compared with the wild-type HC1 oligonucleotide. The combined mutant (M3) did not compete at all even at more than a 1000-fold excess to the labeled oligonucleotide, suggesting that it completely lost the ability to bind to HAP1.

Figure 3. Binding of the selected sites and the UAS1 of *CYC1* by HAP1 (residues 1–171). DNA-binding reactions and gel shift analysis were carried out using DNA containing the UAS1 of *CYC1* (lanes 1,2), site 10 (lanes 3,4), site 23 (lanes 5,6), site 68 (lanes 7,8), site 75 (lanes 9,10), site 79 (lanes 11,12), and site 18 (lanes 13,14) in the presence (lanes 2,4,6,8,10,12,14) or absence (lanes 1,3,5,7,9,11,13) of 20 nM HAP1. The positions of free and HAP1-bound DNA are indicated.

Similar results were obtained when HAP1 binding was carried out directly on these oligonucleotides (data not shown). These results show that HAP1 selectively recognizes a direct repeat of CGGs and that the conserved T in the spacer is critical for high-affinity HAP1 binding.

Two DNA-binding domains of a HAP1 dimer are positioned asymmetrically on the DNA

As discussed above, GAL4 binds to a palindromic inverted repeat of two CGGs and the two subunits are positioned symmetrically on the symmetric DNA sites (Marmorstein et al. 1992). We have shown that HAP1 selectively recognizes a direct repeat of CGGs, but it is unclear how a HAP1 homodimer recognizes an asymmetric site. Are the two zinc cluster DBDs positioned symmetrically on the DNA as in the case of GAL4, or are they positioned asymmetrically on DNA to make the same contacts at both CGGs?

To address these questions, we used dimethylsulfate (DMS) footprinting to study HAP1–DNA interactions with various natural and selected sites. Figure 6A shows the DMS footprinting patterns of HAP1 on the UAS1 of *CYC1*, the UAS of *CYT1*, and the selected site 18. When HAP1 binds to the UAS1 of *CYC1*, the first G of both CGG triplets is strongly protected on the top strand (solid circles, cf. lane 1 to 2 and 3) while there are no G contacts on the other strand (cf. lanes 4 and 5). The same occurs at the UAS of *CYT1* and site 18; the Gs in the CGG triplets of the top strand are strongly protected (cf. lanes 7 and 8 with 6 for site 18; cf. lane 10 with 9 for the UAS of *CYT1*), and there is no protection on the other strand (not shown). Although no DMS protection was observed on the bottom strand of these sites, DNase I footprinting showed that both strands of these sites were equally protected by HAP1 (see Fig. 4; data not shown). These results, summarized in Figure 6C, suggest that HAP1 makes comparable contacts at both CGGs. Note that on the UAS1 of *CYC1*, HAP1 binding creates two hypersensitive sites (open arrowheads) adjacent to the first CGG but not the second. This is attributable to the difference in the nucleotide sequences following the CGG. When a synthetic site containing the same sequences following the CGGs are probed, HAP1 binding

created the same protected and hypersensitive Gs near both CGG triplets (data not shown).

The footprinting patterns on the UAS of *CYC7* and site 75 (Figure 6B) give a better comparison of the nucleotide contacts. Both CGCs of *CYC7* are followed by TA, and both CGGs of site 75 are followed by AG. As shown in lanes 1 and 2 (Fig. 6B), at the CGGAG sequence on the

Figure 4. HAP1 binds specifically to the region that contains the consensus nucleotide sequences. Shown are the DNase I footprinting patterns of HAP1 (1–171) on site 10 (lanes 1–3), site 18 (lanes 4–6), site 23 (lanes 7–9), and the UAS of *CYC7* (lanes 10–12). The G + A digestion patterns (by formic acid) of DNA of sites 10, 18, and 23 are shown in the unnumbered lanes. The DNA used here was labeled on the bottom strand (Materials and methods). Approximately 500 nM of HAP1 was incubated with DNA in lanes 2,3,5,6,8,9,11, and 12. As controls, no HAP1 was included in the reactions in lanes 1,4,7, and 10.

Figure 5. The effect of mutations in the consensus sequence on HAP1-binding affinity. HAP1 binding to the radiolabeled UAS/CYC7 oligonucleotide was assayed by gel shift and competed by the unlabeled UAS/CYC7 oligonucleotide (lanes 2–5, the wild-type site 18 (HC1) (CGGACTTATCGG, lanes 7–10), mutant sites M1 (CGGACTTATCCG, lanes 11–14), M2 (CGGACTCATCCG, lanes 15–18), and M3 (CGGACTCATCCG, lanes 19–22). As controls, lanes 1 and 6 did not contain any unlabeled oligonucleotides. Unlabeled oligonucleotide of 10 (lanes 2,7,11,15,19), 50 (lanes 3,8,12,16,20), 250 (lanes 4,9,13,17,21), and 1250 ng (lanes 5,10,14,18,22) was included in the DNA-binding reactions in the presence of 50 nM HAP1 (1–171).

top strand of site 75, HAP1 binding produces a strong protection at the first G (solid circle) and a hypersensitive site at the last G (open arrowhead) on both copies of CGGAG (see also Fig. 6C). This strongly suggests that the two DBDs of HAP1 make contacts with both copies in the same polarity. The appearance of hypersensitive G adjacent to CGGs in the UAS1/CYC1 and site 75 suggests that the DNAs in these sites are distorted, and as a result, the second G in the CGGs are not well protected, unlike those in the UAS/CYT1 and site 18. In the case of the UAS/CYC7, most interactions occur at the TAGCG sequence on the bottom strand (Figure 6B); the Gs of CGCs on the top strand are only weakly protected (not shown). HAP1 binding creates a hypersensitive A (open arrowhead), a strong protection at the first G (solid circle) followed by a weak protection at the last G (solid circle) on both TAGCGs (Fig. 6B, cf. lane 4 with 3; summarized in Fig. 6C). This strongly suggests that the two DBDs of HAP1 dimers are positioned in the same orientation when contacting the two copies of GCGAT of the UAS/CYC7. In summary, the footprinting experiments strongly suggest that the two DBDs of a HAP1 dimer are positioned in tandem when contacting its DNA sites.

One DBD of the HAP1 homodimer swivel 180° relative to the dimerization domain

It is well established how homodimers bind to palindromic symmetrical DNA elements. Well-studied examples include *Escherichia coli* λ repressor (Jordan and Pabo 1988), the yeast activator GAL4 (Marmorstein et al. 1992), and the mammalian steroid hormone receptors (Luisi et al. 1991). We consider two possible models for how HAP1 homodimers may differ from the other GAL4 family proteins, such as GAL4 and PPR1, and bind asymmetrically to a direct repeat. One possibility involves an unusual dimerization domain such that two monomers bind to each other via different dimerization interfaces (Fig. 7B, model I), thereby, positioning the two DBDs in a directly repeated orientation. By this model, the HAP1 dimerization domain must contain more than one

dimerization interface. This is possible, as the HAP1 dimerization domain is homologous to that of myosin heavy chains (Zhang et al. 1993), which possess multiple dimerization interfaces and polymerize into oligomers (McLachlan and Karn 1982).

A second possibility is that HAP1 dimerizes like GAL4 but that the HAP1 linker rotates freely, allowing asymmetric interactions between the DBDs to position the DBDs in a directly repeated orientation (Fig. 7B, model II). Furthermore, interaction between the DBD and the TA in the spacer may also stabilize this asymmetric binding. By model II, however, we must conclude that HAP1 DBDs cannot position themselves in a rotationally symmetrical orientation, because none of the selected sites contained inverted CGGs and mutations converting the direct repeat to an inverted repeat greatly diminish HAP1-binding affinity.

To distinguish between these two possibilities, we substituted the dimerization domain of HAP1 with that of PPR1, which is known to contain coiled-coils and dimerize symmetrically like that of GAL4 (Reece and Ptashne 1993; R. Marmorstein and S. Harrison, pers. comm.). If the dimerization domains of HAP1 dimerize asymmetrically to bind to a direct repeat of CGGs as proposed in model I, the substitution by the symmetric dimerization domain of PPR1 should abolish its preference for direct repeats. However, if the linker can rotate freely as suggested in model II, the substitution of the dimerization domain should not affect its ability to bind to DNA. The fusion protein containing the zinc cluster domain and linker of HAP1 and the dimerization domain of PPR1 (Fig. 8, H–P) was expressed in *E. coli* under the control of the T7 polymerase and purified by ion exchange chromatography (Materials and methods). Strikingly, the fusion binds to the HAP1 site as well as or better than HAP1 (cf. lanes 2 and 3). The fusion may bind better than HAP1 because the PPR1 dimerization domain is stronger than that of HAP1 as suggested by cross-linking experiments (data not shown). Furthermore, the fusion protein selectively binds to the HAP1 site but not the PPR1 site and forms a complex with the same mo-

Figure 6. DMS protection footprinting analysis on various HAP1 sites. (A) DMS footprinting patterns of HAP1 on the UAS1/CYC1, site 18, and the UAS/CYT1. DNA containing the UAS1/CYC1, site 18, and the UAS/CYT1 was incubated with 500 nM HAP1 (lanes 2,3,5,7,8,10) prior to DMS treatment. As controls, lanes 1, 4, 6, and 9 show the DMS cleavage patterns of DNA containing the UAS/CYC1 (top and bottom strand), site 18, and the UAS/CYT1, respectively, in the absence of HAP1. Note that only the bottom strand of the UAS/CYC1 is shown here as an example because the DMS digestion patterns of the bottom strand of site 18 and UAS/CYT1 did not exhibit any difference in the presence or absence of HAP1. The strongly protected nucleotides by HAP1 are labeled by solid circles, and the hypersensitive nucleotides induced by HAP1 are labeled by open arrowheads. (B) DMS footprinting patterns of HAP1 on site 75 and the UAS/CYC7. DNA containing site 75 and the UAS/CYC7 was incubated with 500 nM HAP1 (lanes 1–171) (lanes 2, 4) prior to DMS treatment. As controls, lanes 1 and 3 show the DMS cleavage patterns of DNA containing site 75 and the UAS/CYC7, respectively, in the absence of HAP1. The weakly and strongly protected nucleotides by HAP1 are labeled by open and solid circles, respectively. The hypersensitive nucleotides induced by HAP1 are labeled by open arrowheads. (C) Summary of DMS footprinting data in A and B. Shown are DMS footprinting results on the UAS of CYC7, the UAS1 of CYC1, the UAS of CYT1, site 18, and site 75. The weakly protected, strongly protected, and hypersensitive nucleotides are labeled by open circles, solid circles, and open arrowheads, respectively.

bility as HAP1, suggesting that the fusion protein maintains the DNA-binding properties of HAP1. These results show that HAP1 asymmetric binding to a direct repeat of CGG triplets is not attributable to asymmetric dimerization. Rather, one DBD of the HAP1 homodimer must swivel 180° to align the two DBDs asymmetrically on the DNA.

Discussion

In this report we have shown that HAP1 dimers recognize a direct repeat of two CGG triplets in contrast to the other members of the GAL4 family, which bind to inverted repeats of CGG triplets (see Fig. 7A). The two DBDs of HAP1 dimers make the same contacts with both CGGs, suggesting that the two DBDs are also positioned asymmetrically according to the polarity of the

DNA sites. Furthermore, our domain-swapping experiment suggests that one DBD of the HAP1 homodimer is able to swivel 180° to contact the direct repeat of two CGGs.

Asymmetric binding: the DNA elements

The HAP1 DNA-binding site is different from that of other GAL4 family proteins in two aspects (Carey et al. 1989; Reece and Ptashne 1993). First, the DNA sites of GAL4, PPR1, and PUT3 contain a palindromic inverted repeat of two CGGs, whereas HAP1 contains a direct repeat (see Fig. 7A). The orientation of the second CGG in the HAP1 site is crucial for binding, as a mutation (M1) that changes a direct repeat to an inverted repeat greatly diminishes HAP1-binding affinity. Second, the spacer of the HAP1 sites possesses no symmetry and

A

GAL4	→ CGG n11 ←
PUT3	CGG n10 CCG
PPR1	CGG n6 CCG
LEU3	CGG n4 CCG
HAP1	→ CGG n6 ←

B

I

CGGNNNTANCGG

II

CGGNNNTANCGG

highly selective binding to direct repeats of CGG (Fig. 8). This result allows us to exclude the possibility that asymmetric binding is attributable to asymmetric dimerization. Rather, the linker must be able to rotate freely to position the two DBDs in a directly repeated orientation (Fig. 7B, model II). Additional energy is likely provided to stabilize the two DBDs in such an orientation. The energy could come from the protein-DNA interactions involving the TA nucleotides, as discussed above. Another source of energy can come from possible asymmetric interactions between the two DBDs (Fig. 7B, model II). From the x-ray crystallographic structures of GAL4 and PPR1 [Marmorstein et al. 1992; R. Marmor-

Substitution of the HAP1 dimerization domain with the PPR1 symmetric dimerization domain preserved the

DNA Binding

	HAP1 site	PPR1 site
HAP1	+	-
PPR1	-	+
H-P	+	-

DNA Binding Assay

lane	HAP1 site			PPR1 site			protein
	P	H-P	H	P	H-P	H	
1							bound
2							
3							
4							
5							
6							

stein and S. Harrison, pers. comm.), the zinc cluster domains appear to be quite compact and rigid. Thus, it is not likely that the HAP1 zinc cluster domain can make additional DNA–protein or protein–protein interactions except for contacting the CGG triplet. On the contrary, the linkers are long and flexible and the HAP1 linker is 5 residues longer than that of GAL4 or PPR1. Previous studies have demonstrated that the GAL4 family proteins distinguish DNA sites of different length via the linkers (Reece and Ptashne 1993). The GAL4 linker contacts the phosphate backbone in the spacer between the two symmetric CGG triplets (Marmorstein et al. 1992). It is possible that the HAP1 linker makes asymmetric protein–protein interactions or interacts with TA in the spacer to stabilize the two DBDs in a directly repeated orientation. Whatever the exact sequence involved, the mechanism of recognition of direct repeats by HAP1 is to swivel one of the DBDs of the dimer about the linker.

Although it has been shown that heterodimers of retinoid X receptor (RXR) with retinoic acid receptor (RAR), thyroid hormone receptor (TR), or vitamin D receptor (VDR) selectively bind to directly repeated hormone response elements (HREs) asymmetrically (Kurokawa et al. 1993; Mader et al. 1993; Perlmann et al. 1993), the molecular basis for this binding preference by the receptors is not yet entirely clear. While one study suggested that the preference may be attributable to asymmetric interactions between the zinc fingers of the receptors (Perlmann et al. 1993), another suggested it is because of an orientation of the DNA-binding and dimerization domain in RXR that is opposite to their relative orientation in TR, RAR, or VDR (Kurokawa et al. 1993). In any case, the mechanism of recognition of direct repeats is likely to differ from the HAP1 example. First, RXR binds its sites as a heterodimer with TR, RAR, or VDR. Thus, asymmetry may be established by structural differences between the subunits. Second, the structure of the DNA recognition region of the receptors, two separate Cys-4 zinc fingers and flexible helices allowing weak dimerization, is completely different from that of the GAL4 family.

HAP1 zinc cluster domain is flexible enough to accommodate the change of CGG to CGC triplets

The UAS of *CYC7* contains a direct repeat of CGC triplets rather than CGG triplets (see Fig. 2). When binding to the UAS/*CYC7*, the HAP1 DNA-binding domain makes strong contacts with the GCG triplets on the strand opposite to the conserved T, instead of the CGC triplets (see Fig. 6C). If we infer that HAP1 is oriented by the TATT sequences in the UAS/*CYC7* site, these GCGs lie on the opposite strand from the CGGs contacted in most HAP1 sites. The key residues in GAL4 for making specific base contacts are all conserved in HAP1 (Marmorstein et al. 1992). Whereas the x-ray crystal structures of GAL4–DNA complexes and PPR1–DNA complexes suggest that this class of zinc cluster is compact and quite rigid (Marmorstein et al. 1992; R. Marmorstein and S. Harrison, pers. comm.), the ability of

HAP1 to recognize the *CYC7* site might indicate that this zinc cluster is sufficiently flexible to accommodate the change in DNA sequence in *CYC7*. Furthermore, this could indicate that HAP1 is forced into a novel conformation when bound to UAS/*CYC7*. An altered conformation may explain why HAP1 activation at *CYC7* sites is relatively poor compared with other sites (Kim et al. 1990). The HAP1–18 mutation in the DNA-binding domain greatly increases the activity of HAP1 at *CYC7* without altering its affinity for the site (Creusot et al. 1988; Kim and Guarente 1989; Pfeifer et al. 1989). Perhaps the Cys-63 → Arg change overcomes the limitation to activation imposed by the conformation of HAP1 at *CYC7*.

Materials and methods

Expression and purification of HAP1 fragments and HAP1–PPR1 chimeric proteins

To construct GST–HAP1 fusion expression vectors, HAP1 DNA containing coding sequences for residues 1–171 and 52–171 was inserted into the pGEX-KT vector (Guan and Dixon 1991) at the *Bam*HI site in-frame, respectively. *E. coli* cells bearing the fusion vectors were induced with 1 mM IPTG and lysed by passing through a French press as described (Guan and Dixon 1991). After centrifugation and filtering, the cell extracts were passed through glutathione–Sepharose columns (Pharmacia). After extensive washing, thrombin was added to the columns followed by incubation at room temperature for 1 hr. The HAP1 fragments were subsequently washed off the columns and stored in 25 mM HEPES (pH 7.5), 10% glycerol, 10 μ M ZnOAc₂, 2 mM dithiothreitol (DTT). Both fragments 1–171 and 52–171 were shown to bind DNA as dimers (Zhang et al. 1993).

The HAP1 fragment 56–148, PPR1, and the fusion H–P (Fig. 8) were all constructed with PCR (oligonucleotide sequences available upon request). Amplified DNA fragments were cloned into the T7 expression vector pRSET A (Invitrogen) cut with *Nde*I and *Eco*RI. The plasmids were subjected to DNA sequence analysis to ensure that the correct fusion was made and no mutation occurred. The plasmids were transformed into *E. coli* strain BL21–(DE3)pLysS (Invitrogen), and cells were grown at 37°C in the presence of 50 μ g/ml of ampicillin until an OD₆₀₀ of 0.6 was reached. Then, cells were induced with 1 mM IPTG in the presence of 100 μ M ZnOAc₂. After 2.5 hr, the cells were collected by centrifugation and lysed by passing through a French press. The proteins were then purified to 90% purity from the extract by using S-Sepharose Fast Flow column chromatography (Pharmacia) as described (Marmorstein et al. 1992; Reece and Ptashne 1993). The HAP1 56–148 and PPR1 29–123 were purified and provided by R. Marmorstein.

HAP1-binding site selection

The basic idea for selecting HAP1-binding sites is outlined in Figure 1. We followed the protocol described by Pollock and Treisman (1990) with a few modifications. Briefly, we synthesized three oligonucleotides: Random sequence oligonucleotide H76, 5'-CAGGTCAGTTCAGCGGATCCAGCAC(A/G/C/T)-26GTGGCGAATTTCAGTGCAACTGCAGC-3'; primer F, 5'-GCTGCAGTTGCACTGAATTCGCCAC-3'; and primer R, 5'-CAGGTCAGTTCAGCGGATCCAGCAC-3'. Each oligonucleotide was gel purified prior to use. Oligonucleotide H76 was first rendered double stranded by DNA synthesis using *Taq*

polymerase and purified on nondenaturing 8% polyacrylamide gels. HAP1-binding reactions (25 μ l) were carried out in 5% glycerol, 4 mM Tris (pH 8), 40 mM NaCl, 4 mM MgCl₂, 10 mM DTT, 1 μ g of poly[d(I-C)]/[d(I-C)], 10 μ M ZnOAc₂, and 300 μ g/ml of BSA. The reaction mixtures were incubated at room temperature for at least 1 hr before loading onto 4% polyacrylamide gels in 0.5 $\times$ TBE buffer (at 4°C) as described (Zhang et al. 1993). The HAP1-bound DNA was cut out from dried gels and used directly for PCR amplification. The PCR reactions (50 μ l) contained 150 ng of each primer F and R, 5 μ Ci of [³²P]dCTP (3000 Ci/mMole), 20 μ M unlabeled dCTP, 50 μ M each dATP, dTTP, dGTP, and 2 units of *Taq* polymerase in the *Taq* polymerase reaction buffer supplied by Boehringer Mannheim Biochemical. Amplification was at 94°C for 1 min, 62°C for 1 min, 72°C for 1 min (18 cycles). Amplified DNA was purified on 8% polyacrylamide gels before use. HAP1-binding selection was carried out for four cycles. HAP1 fragment 1–171 (500 nm) was used for the first selection. The amount of HAP1 fragment used was decreased gradually in subsequent binding selections. Approximately 2 nm HAP1 fragment was included in the last selection.

After the last cycle of selection and amplification, the amplified DNA was purified and digested with *Eco*RI and *Bam*HI. The DNA was ligated with Bluescript KS+ vector (Stratagene), cut with *Eco*RI and *Bam*HI, and transformed into *E. coli* TG1 cells. Single-stranded DNA was prepared by using helper phage (Vieira and Messing 1987) and sequenced by the chain termination method (Bankier et al. 1987).

Footprinting analysis

The DNA used in footprinting analysis was ~200 bp generated by PCR amplification using single-stranded Bluescript DNA containing a HAP1 site as template. Two M13 sequencing primers, –20 primer (5'-GTAAAACGACGGCCAGT-3') and the reverse primer (5'-AACAGCTATGACCATGAT-3'), were used in the PCR reactions. To end-label the DNA, one of the primers was labeled at the 5' end with [γ -³²P]ATP and T4 polynucleotide kinase prior to PCR reactions. The amplification was at 94°C for 1 min, 42°C for 2 min, and 72°C for 1.5 min (18 cycles). The labeled DNA was purified on nondenaturing 8% polyacrylamide gels prior to use.

The DNA-binding reactions (20 μ l) were carried out in 5% glycerol, 4 mM Tris (pH 8), 40 mM NaCl, 4 mM MgCl₂, 10 mM DTT, 3 μ g of salmon sperm DNA, 10 μ M ZnOAc₂, and 300 μ g/ml of BSA. Approximately 0.01 pmole of labeled DNA was incubated with various amount of HAP1 as indicated. For DNase I footprinting (Galas and Schmitz 1978), 10⁻⁴ mg/ml of DNase I was added to each reaction for 1 min and the reactions were terminated by the addition of 50 mM EDTA followed by phenol extraction and ethanol precipitation. For DMS footprinting, 0.15% freshly made DMS was added to the reactions for 1 min and the reactions were terminated by the addition of β -mercaptoethanol. Piperidine cleavage was carried out as described (Maxam and Gilbert 1980). After DNase I or DMS treatment the cleaved DNA was analyzed on 6% polyacrylamide/urea sequencing gels and autoradiographed.

DNA mobility shift assay

The DNA-binding reactions were carried out as described above in footprinting analysis. The radiolabeled UAS/CYC7 probe was prepared as described (Turcotte and Guarente 1992). The labeled DNA used in Figure 3 was generated by PCR reactions with labeled primers by PCR amplification as described above. The reaction mixtures were loaded onto 4% polyacrylamide gels in 0.5 $\times$ TBE, and electrophoresis was carried out at 4°C.

Acknowledgments

We are grateful to Drs. R. Marmorstein, R. Reece, and E. Rebar for critical reading of the manuscript. We thank Drs. R. Marmorstein and S. Harrison for providing purified PPR1 and communication of unpublished data. We thank Dr. R. Pollock for discussions and suggestions regarding the random oligonucleotide selection technique. This work is supported by a National Institutes of Health grant (GM30454 to L.G.) L.Z. was supported previously by a fellowship from the Jane Coffin Childs Memorial Fund for Medical Research.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Bankier, A.T., K.M. Weston, and B. Barrell. 1987. Random cloning and sequencing by the M13/dideoxy nucleotide chain termination method. *Methods Enzymol.* **155**: 51–93.
- 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.
- Creusot, F., J. Verdiere, M. Gaisne, and P.P. Slonimski. 1988. CYP1 (HAP1) regulator of oxygen-dependent gene expression in yeast. *J. Mol. Biol.* **204**: 263–276.
- Friden, P. and P. Schimmel. 1987. LEU3 of *Saccharomyces cerevisiae* encodes a factor for control of a group of leucine-specific genes. *Mol. Cell. Biol.* **7**: 2708–2717.
- . 1988. LEU3 of *Saccharomyces cerevisiae* activates multiple genes for branched-chain amino acid biosynthesis by binding to a common decanucleotide core sequence. *Mol. Cell Biol.* **8**: 2690–2697.
- Fytlovich, S., M. Gervais, C. Agrimonti, and B. Guiard. 1993. Evidence for an interaction between the CYP1(HAP1) activator and a cellular factor during heme-dependent transcriptional regulation in the yeast *Saccharomyces cerevisiae*. *EMBO J.* **12**: 1209–1218.
- Galas, D. and A. Schmitz. 1978. DNase I footprinting: A simple method for the detection of protein–DNA binding specificity. *Nucleic Acids Res.* **5**: 3157–3170.
- Guan, K.L. and J.E. Dixon. 1991. Eukaryotic proteins expressed in *Escherichia coli*: An improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Anal. Biochem.* **192**: 262–267.
- Johnston, M. 1987. A model fungal gene regulatory mechanism: The GAL genes of *Saccharomyces cerevisiae*. *Microbiol. Rev.* **51**: 458–476.
- Jordan, S.R. and C.O. Pabo. 1988. Structure of the lamda complex at 2.5 Å resolution: Details of the repressor-operator interactions. *Science* **242**: 893–899.
- Kim, K.S. and L. Guarente. 1989. Mutations that alter transcriptional activation but not DNA binding in the zinc finger of yeast activator HAP1. *Nature* **342**: 200–203.
- Kim, K.S., K. Pfeifer, L. Powell, and L. Guarente. 1990. Internal deletions in the yeast transcriptional activator HAP1 have opposite effects at two sequence elements. *Proc. Natl. Acad. Sci.* **87**: 4524–4528.
- Kurokawa, R., V.C. Yu, A. Naar, S. Kyakumoto, Z. Han, S. Silverman, M.G. Rosenfeld, C.K. Glass. 1993. Differential orientations of the DNA-binding domain and carboxy-terminal dimerization interface regulate binding site selection by nuclear receptor heterodimers. *Genes & Dev.* **7**: 1423–1435.
- Luisi, B.F., W.X. Xu, Z. Otwinowski, L.P. Freedman, K.R. Yamamoto, and P.B. Sigler. 1991. Crystallographic analysis of

- the interaction of the glucocorticoid receptor with DNA. *Nature* **352**: 497–505.
- Mader, S., J. Chen, Z. Chen, J. White, P. Chambon, and H. Gronemeyer. 1993. The patterns of binding of RAR, RXR and TR homo- and heterodimers to direct repeats are dictated by the binding specificities of the DNA binding domains. *EMBO J.* **12**: 5029–5041.
- Marmorstein, R., M. Carey, M. Ptashne, and S.C. Harrison. 1992. DNA recognition by GAL4: Structure of a protein-DNA complex. *Nature* **356**: 408–414.
- Maxam, A. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**: 497–559.
- McLachlan, A.D. and J. Karn. 1982. Periodic charge distributions in the myosin rod amino acid sequence match cross-bridge spacings in muscle. *Nature* **299**: 226–231.
- Perlmann, T., P.N. Rangarajan, K. Umesono, R.M. Evans. 1993. Determinants for selective RAR and TR recognition of direct repeat HREs. *Genes & Dev.* **7**: 1411–1422.
- Pfeifer, K., T. Prezant, and L. Guarente. 1987. Yeast HAP1 activator binds to two upstream activation sites of different sequence. *Cell* **49**: 19–27.
- Pfeifer, K., K.S. Kim, S. Kogan, and L. Guarente. 1989. Functional dissection and sequence of yeast HAP1 activator. *Cell* **56**: 291–301.
- Pollock, R. and R. Treisman. 1990. A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res.* **18**: 6197–6204.
- Prezant, T., K. Pfeifer, and L. Guarente. 1987. Organization of the regulatory region of the yeast *CYC7* gene: Multiple factors are involved in regulation. *Mol. Cell. Biol.* **7**: 3252–3259.
- Reece, R.J. and M. Ptashne. 1993. Determinants of binding-site specificity among yeast C6 zinc cluster proteins. *Science* **261**: 909–911.
- 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.
- Schneider, J.C. and L. Guarente. 1991. Regulation of the yeast *CYT1* gene encoding cytochrome c1 by HAP1 and HAP2/3/4. *Mol. Cell. Biol.* **11**: 4934–4942.
- Siddiqui, A.H. and M.C. Brandriss. 1989. The *Saccharomyces cerevisiae* PUT3 activator protein associates with proline-specific upstream activation sequences. *Mol. Cell Biol.* **9**: 4706–4712.
- Turcotte, B. and L. Guarente. 1992. HAP1 positive control mutants specific for one of two binding sites. *Genes & Dev.* **6**: 2001–2009.
- Vieira, J. and J. Messing. 1987. Production of single-stranded plasmid DNA. *Methods Enzymol.* **153**: 3–11.
- Zhang, L. and L. Guarente. 1994. HAP1 is nuclear but is bound to a cellular factor in the absence of heme. *J. Biol. Chem.* **269**: 14643–14647.
- Zhang, L., O. Bermingham-McDonogh, B. Turcotte, and L. Guarente. 1993. Antibody-promoted dimerization bypasses the regulation of DNA binding by the heme domain of the yeast transcriptional activator HAP1. *Proc. Natl. Acad. Sci.* **90**: 2851–2855.
- Zhou, K., P.R. Brisco, A.E. Hinkkanen, G.B. Kohlhaw. 1987. Structure of yeast regulatory gene *LEU3* and evidence that *LEU3* itself is under general amino acid control. *Nucleic Acids Res.* **15**: 5261–5273.
- Zitomer, R.S., J.W. Sellers, D.W. McCarter, G.A. Hastings, P. Wick, and C. Lowry. 1987. Elements involved in oxygen regulation of the *Saccharomyces cerevisiae* *CYC7* gene. *Mol. Cell. Biol.* **7**: 2212–2220.

The yeast activator HAP1--a GAL4 family member--binds DNA in a directly repeated orientation.

L Zhang and L Guarente

Genes Dev. 1994, **8**:

Access the most recent version at doi:[10.1101/gad.8.17.2110](https://doi.org/10.1101/gad.8.17.2110)

References

This article cites 33 articles, 16 of which can be accessed free at:
<http://genesdev.cshlp.org/content/8/17/2110.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).