Control of Hap1−DNA Site Recognition through the Interplay of Multiple Distinct Intermolecular Interactions

Lee L. Wang, Inga Denman, and Matthew Junker*

Department of Molecular and Cell Biology, University of Texas at Dallas, P.O. Box 830688, Richardson, Texas 75080-0688

Received March 25, 2004; Revised Manuscript Received July 5, 2004

ABSTRACT: Hap1 belongs to the Zn_2Cys_6 zinc binuclear cluster family of transcription factors that typically bind as dimers to symmetric DNA sites containing two CGG triplets separated by spacer DNA. The cluster domain binds CGG while an adjoining C-terminal linker and dimerization helix specifies the length of spacer DNA recognized. Hap1 is unusual in binding a direct repeat of CGG triplets, in contacting a TA in the spacer DNA, and in making direct dimer contacts between its cluster domains. Binding of Hap1 fragments to different DNA sites was tested to determine how these interactions control Hap1−DNA recognition. The spacer TA contacts were found to facilitate monomer binding of Hap1 to a single CGG. When the spacer-binding residues were deleted, binding was still specific for the direct repeat but was much weaker and appeared to require dimerization. When the dimerization helix and all subsequent C-terminal residues were deleted, the remaining linker, cluster domain, and spacer-binding residues still dimerized on DNA. The energy of this dimerization was comparable to that of the Hap1−spacer TA interaction. Moving the TA from the spacer to a position following the second CGG maintained Hap1 monomer binding but greatly weakened dimerization. This suggested that binding a TA after the second CGG triplet required a geometry that impaired dimerization with a Hap1 molecule on the first CGG. The geometric restraints for optimal TA binding and dimerization thus drive Hap1 selectivity for CGG direct repeat sites that contain an asymmetrically positioned spacer TA following the first CGG triplet.

The transcription factor Hap1 (heme activation protein 1) mediates heme-induced transcription of genes in the yeast *Saccharomyces cerevisiae*. Hap1 binds to DNA and activates transcription in a heme-dependent fashion, coordinating the expression of heme-containing proteins such as cytochromes with the availability of their required heme cofactor (*1*). Hap1 belongs to a large family of fungal transcription factors including GAL4, PPR1, and PUT3 that contain a Zn_2Cys_6 binuclear cluster within their DNA-binding domain. These proteins can be delineated at the primary sequence level into DNA-binding, dimerization, and transactivation regions. The zinc cluster forms the core of an ~40 residue domain within the DNA-binding region that recognizes CGG DNA sequences. The zinc cluster transcription factors typically bind as dimers to sites containing two CGG triplets separated by a specific length of spacer DNA. The sites are usually symmetric with the CGG triplets arranged as an inverted repeat (CGG-N_i-CCG; N is any nucleotide) or an everted repeat (CCG-N_i-CGG), although at least one member binds monomerically to single CGG sites (*2*−*5*). At least for those members that bind to inverted repeats, the structure of linker and dimerization domains immediately C-terminal to the zinc cluster domain specifies which length of spacer DNA is recognized (*2*).

Hap1 exhibits several unusual properties in DNA binding compared with the other zinc binuclear cluster transcription factors. Instead of binding to a symmetric repeat of CGG triplets, Hap1 binds to a direct repeat of CGG triplets. Hap1 also has a strong TA sequence preference in the spacer DNA, with a preferred binding site of 5′-CGG-N_3TAN-CGG-3′ (*6*). Domain-swapping experiments indicate that Hap1's specificity for the direct repeat sequence is controlled by its cluster domain and preceding residues, rather than by its C-terminal linker and dimerization helices (*7*). Hap1 is also unusual in binding to sites containing a direct repeat of CGC triplets separated by six base pairs with affinity similar to that of binding to sites containing CGG triplets. Transcriptional activation from the CGC sites is lower than from the CGG sites, and this appears to be solely dependent on the difference in the triplet sequence. This suggests that the DNA may allosterically regulate Hap1, a conclusion supported by studies of Hap1 variants containing mutations in the DNA-binding region (*8*−*10*).

Hap1's unusual DNA-binding properties likely result from a complex set of protein−protein and protein−DNA interactions that are evident in the crystal structure of a Hap1 fragment bound to a direct repeat of CGC (not CGG) triplets (*11*). Despite forming a nonsymmetric head-to-tail dimer on the DNA, the Hap1 zinc cluster domains bound each CGC triplet in a manner similar to that of the cluster domains of other transcription factors binding to CGG. Also like other zinc cluster transcription factors, dimer contacts were mediated by Hap1 C-terminal helices (Figure 1). However, other interactions were unique to Hap1. Residues N-terminal to the "upstream" Hap1 molecule (bound to the more 5′ CGC triplet) extended over the minor groove of the spacer DNA and made base contacts to the preferred TA sequence and electrostatic contacts to the phosphodiester backbone (Figure 1). Extensive dimer contacts also occurred among the two

* To whom correspondence should be addressed. Phone: (972) 883-2514. Fax: (972) 883-2409. E-mail: mjunker@utdallas.edu.

10.1021/bi049419i CCC: $27.50
Published on Web 10/05/2004

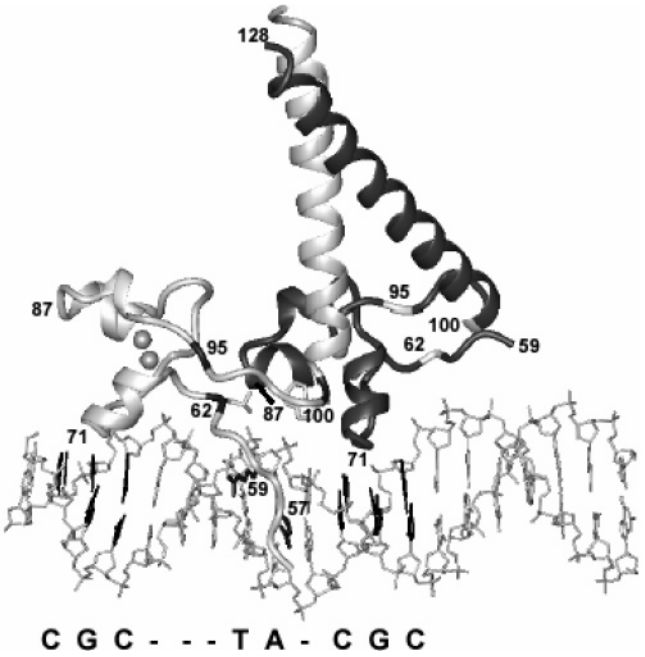

FIGURE 1: Representation of the crystal structure of a Hap1 fragment dimerized on a DNA site containing a direct repeat of CGC triplets (PDB code 1hwt) (*11*). The upstream Hap1 molecule is shown in white and the downstream Hap1 molecule in dark gray with distinct regions defined by stripes and residue numbers (62−95 for the zinc cluster domains, 96−100 for the linkers, and 101−128 for the dimerization helices). Residues 71 and 87 mark the ends of helices 1 and 2, respectively, within the cluster domains. The zinc atoms in the cluster domains are shown as gray spheres, and the CGC triplet bases are colored black. Side chains contacting the spacer TA and two dimerization side chains (Leu62 and Trp100) are shown for the upstream Hap1 molecule. The figure was rendered with the molecular graphics program MolMol (*21*).

cluster domains with the linker (especially Trp100) and N-terminal residue (Leu62) of the upstream Hap1 molecule (Figure 1). Furthermore, the N-terminus of the dimer helix of the upstream Hap1 molecule packed against the downstream cluster domain (Figure 1). While the sum of these interactions must control Hap1−DNA recognition, the relative contribution of each interaction is not known.

Multiple intermolecular interactions are often evident in the structures of transcription factor−DNA complexes (*12*). In addition, the large multiprotein complexes that assemble at gene promoters in vivo to regulate transcription are themselves stabilized by multiple protein−protein and protein−DNA contacts. Given its complexity in DNA-site recognition, Hap1 provides a model for probing how the interplay of multiple distinct interactions can control DNA recognition. The basis for Hap1−DNA site recognition was investigated by comparing the binding of different Hap1 fragments to different DNA sites. This allowed for systematically testing the effects of eliminating discrete interactions. The results demonstrate that a simultaneous optimization of distinct intermolecular interactions determines the DNA-binding specificity of Hap1, and that perturbations to individual interactions can modulate the Hap1 binding affinity and, very likely, its binding geometry.

MATERIALS AND METHODS

Recombinant Expression and Purification of Hap1 Fragments. Expression plasmids for the various Hap1 fragments were generated by subcloning from a copy of the full-length Hap1 gene in the plasmid pKP(SD5−Hap1) obtained from the laboratory of Dr. Leonard Guarente (MIT). The Hap1(1−146) and Hap1(62−149) fragments were expressed from pET-3b-derived plasmids. The expression plasmid for the Hap1(1−146) fragment was generated by inserting the portion of the Hap1 gene encoding residues 1−445 into pET-3b (*13*), followed by restriction cutting, filling in, and religation at the *Ava*I restriction site at the codon for residue 146. The resulting plasmid encoded Hap1 residues 1−146 followed by nine non-native residues, Asp-Arg-Glu-Gln-Gln-Pro-Val-Tyr-Gly. An analogous expression plasmid for Hap1(62−146) encoded the same nine non-native residues at the C-terminus as well as an N-terminal Met. The Hap1(1−99) and Hap1(54−99) fragments were expressed as fusions to the maltose-binding protein (MBP).[1] DNA encoding for the corresponding fragment of Hap1 was amplified from the Hap1 gene by PCR (polymerase chain reaction) and inserted into the pMAL-c2 plasmid (New England Biolabs). The sequence of all recombinant expression plasmids was verified by DNA sequencing.

The Hap1(1−146) and Hap1(62−146) fragments were expressed in *Escherichia coli* BL21-DE3 cells grown in Luria−Bertani medium. Hap1(1−146) was induced for 4 h at 37 °C with 0.3 mM IPTG (isopropyl β-D-thiogalactopyranoside) and 50 μM $ZnSO_4$, and Hap1(62−146) was induced overnight at 25 °C with 0.03 mM IPTG and 50 μM $ZnSO_4$. Harvested cells were lysed by sonication in 20 mM Tris, 200 mM KCl, 5 mM βME (β-mercaptoethanol), and 1 mM PMSF (phenylmethylsulfonyl fluoride) at pH 8 (lysis buffer). The lysates were clarified by centrifugation and applied to a hydroxylapatite column from which the proteins were eluted with a potassium phosphate gradient at pH 6. Pooled fractions were applied to a Q-Sepharose column equilibrated with 20 mM Tris, 25 mM NaCl, and 5 mM βME at pH 8 (buffer Q). The proteins eluted in the flow-through and were loaded onto an S2 column on a BioRad Biologic system equilibrated with 20 mM MOPS and 5 mM βME at pH 6.5 from which the proteins were eluted by a NaCl gradient. After purification, the Hap1(1−146) and Hap1(62−146) appeared as single bands on Coomassie-stained SDS−PAGE gels.

The Hap1(54−99) and Hap1(1−99) fragments were expressed as MBP fusions in *E. coli* BL21 cells grown at 37 °C and induced with 0.3 mM IPTG and 50 μM $ZnSO_4$. Harvested cells were lysed by sonication in lysis buffer, and the lysates were clarified by centrifugation and applied either to an SP-Sepharose column [MBP−Hap1(1−99)] or to a Q-Sepharose column [MBP−Hap1(54−99)] equilibrated with buffer Q and developed with a NaCl gradient. MBP−Hap1-containing fractions were applied to an amylose column equilibrated with buffer Q containing 150 mM NaCl from which the fusions were eluted with 10 mM maltose. The Hap1 fragments were cleaved from MBP by factor X_a proteolysis. The MBP−Hap1(1−99) digest was passed over a Mono Q column (Amersham-Pharmacia) equilibrated with Q buffer from which Hap1(1−99) was collected in the flow-through and further purified by chromatography on a Mono

[1] Abbreviations: EDTA, ethylenediaminetetraacetic acid; MBP, maltose-binding protein; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis.

S column equilibrated with buffer Q. The MBP−Hap1(54−99) digest was chromatographed on a Mono S column to purify Hap1(54−99).

Prior to DNA-binding experiments, the Hap1 fragments were dialyzed into 10 mM Tris, 50 mM KCl, and 2 μM $ZnCl_2$. Protein concentrations were determined by direct measurement of absorbance at 280 nm using calculated extinction coefficients (*14*) or by determining the Cys concentration (six Cys/Hap1 molecules) by denaturing the proteins in 6 M guanidine−HCl, 30 mM Tris, 10 mM NaCl, 10 mM EDTA, and 1 mM 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) at pH 7.

DNA-Binding Experiments and Analysis. The binding of Hap1 fragments to DNA was monitored using the polyacrylamide gel electrophoretic mobility shift assay. For the binding reactions, purified Hap1 fragments were mixed with 1.5 nM ^{32}P-end-labeled duplex DNA and 1 μg of poly(dI·dC) −poly(dI·dC) (Amersham-Pharmacia) as nonspecific DNA in DNA-binding buffer containing 10 mM Tris, 50 mM KCl, 1 mM dithiotreitol (DTT), and 6% glycerol at pH 7.5 in a total volume of 20 μL. All labeled duplexes contained 24 base pairs based on the direct repeat site sequence 5′-CACACGGACTTATCGGTCTGTCAG-3′. The duplexes also contained four nucleotide overhangs at each end of the complementary strand (GATC at the 5′ end and AGCT at the 3′ end). The binding reactions were incubated for 15 min prior to electrophoresis at a constant 250 V on 15% polyacrylamide gels (75:1 acrylamide/bisacrylamide). The gel and running buffers were TB/2 (44.5 mM Tris, 44.5 mM boric acid). EDTA was omitted from the buffers to avoid chelation of zinc from the Hap1 fragments. The binding reactions and electrophoresis were performed at 22 °C. The labeled DNAs on the gels were detected by autoradiography and phosphorimaging.

For analysis of the gel-shift binding data, phosphorimage band intensities were quantified using the ImageQuant software (Molecular Dynamics). The fraction of DNA shifted in each band was calculated by dividing the integrated intensity of the band by the sum of the integrated intensities of all bands present in the same lane of the gel, including the free DNA. The titrations to the direct repeat site were fit using a model in which Hap1 bound to two sites on a single DNA duplex with dissociation constants K_H for a higher affinity site and K_L for a lower affinity site. For the model, Hap1 dimerized cooperatively on the DNA with dissociation constant K_C. The partition function for this model was $1 + [P]/K_H + [P]/K_L + [P]^2/(K_C K_H K_L)$, where [P] was the protein concentration. The fitting was performed using the Marquardt−Levenberg algorithm as implemented in SigmaPlot 3.0.

Light Scattering and Native Gel Electrophoresis. Multiangle light scattering was measured on a DAWN EOS multiangle light scattering detector in series with an Optilab refractive index detector for determination of protein concentration (Wyatt Technology Corp.). Approximately 200 μg of Hap1(1−146) was chromatographed on a Superdex75 gel filtration column equilibrated with 20 mM Tris and 300 mM NaCl at pH 8, and the column elution was passed directly through the light scattering instrument. The 300 mM NaCl prevented interaction of Hap1 fragments with the Superdex75 resin and did not affect the pattern of band shifts when tested in the gel-shift assay. Molar masses were calculated using the ASTRA software (Wyatt Technology). Native gel electrophoresis was performed on a 15% polyacrylamide gel with gel and running buffers of 10 mM Tris and 50 mM KCl. Protein samples were loaded in the DNA-binding reaction buffer. The gel was run at 40 mA constant current with opposite polarity of standard SDS−PAGE gels. Methylene blue was used as a tracking dye in a separate lane.

RESULTS

To compare the contributions of different intermolecular interactions to Hap1−DNA recognition, DNA-binding experiments were carried out using Hap1 fragments and DNA sites illustrated in Figure 2A. The DNA sites were based on an optimal Hap1 direct repeat sequence of CACACGGACTTATCGGTCTGTCAG identified by in vitro selection (*6*). This site contained the preferred TA sequence in the spacer DNA (underlined), positioned as N_3TA after the upstream CGG triplet, and had no similarly positioned TA following the downstream CGG. The largest fragment, Hap1(1−146), contained Hap1 residues 1−146, which included the DNA-binding and dimerization elements identified in the X-ray crystal structure of the Hap1(56−135)−DNA complex (*11*). A shorter Hap1(62−146) fragment began two residues before the first cluster domain Cys and eliminated all DNA-binding residues N-terminal to the zinc cluster domain (Figure 2A, an initiating Met substituted for residue 61). Hap1(1−99) retained the N-terminal residues but eliminated the dimerization helix and the only linker residue (Trp100) that directly participates in dimerization (Figure 2A). Hap1(54−99) eliminated all residues preceding the DNA-spacer-binding residues as well as the dimerization helix (Figure 2A).

The Hap1 protein fragments were purified following recombinant expression in *E. coli*. In a previous cross-linking study, a Hap1(55−148) fragment showed only weak, if any, dimerization in solution (*7*). Consistent with this, multiangle light scattering showed Hap1(1−146) to be mostly monomeric. Hap1(1−146) eluted as a single major peak near 10 mL on a Superdex75 gel filtration column, monitored by changes in the refractive index (Figure 2B). The calculated molar mass across this peak, along with the 90° light scattering and refractive index signals, is shown in Figure 2C. Across most of the peak the calculated molar mass was near 20 kDa, consistent with the 17.4 kDa Hap1(1−146) being primarily monomeric. The average molar mass approached 30 kDa near the leading edge of the peak, suggesting some Hap1(1−146) could have been dimeric. However, in native gel electrophoresis at the lower concentrations used for DNA-binding experiments, both Hap1(1−146) and Hap1(1−99) (with the dimerization helix deleted) migrated as single bands, suggesting essentially single oligomeric states (Figure 2D).

Hap1 Residues N-Terminal to the Cluster Domain Stabilize Monomeric Binding to DNA. Within the Hap1 dimer−DNA complex, residues 55−59 preceding the upstream Zn_2Cys_6 cluster domain bound the spacer DNA between the pair of CGC triplets, making base contacts to the spacer TA and phosphodiester contacts (Figure 1). The importance of these contacts was initially tested by comparing the ability of Hap1(1−146) and Hap1(62−146) to discriminate between

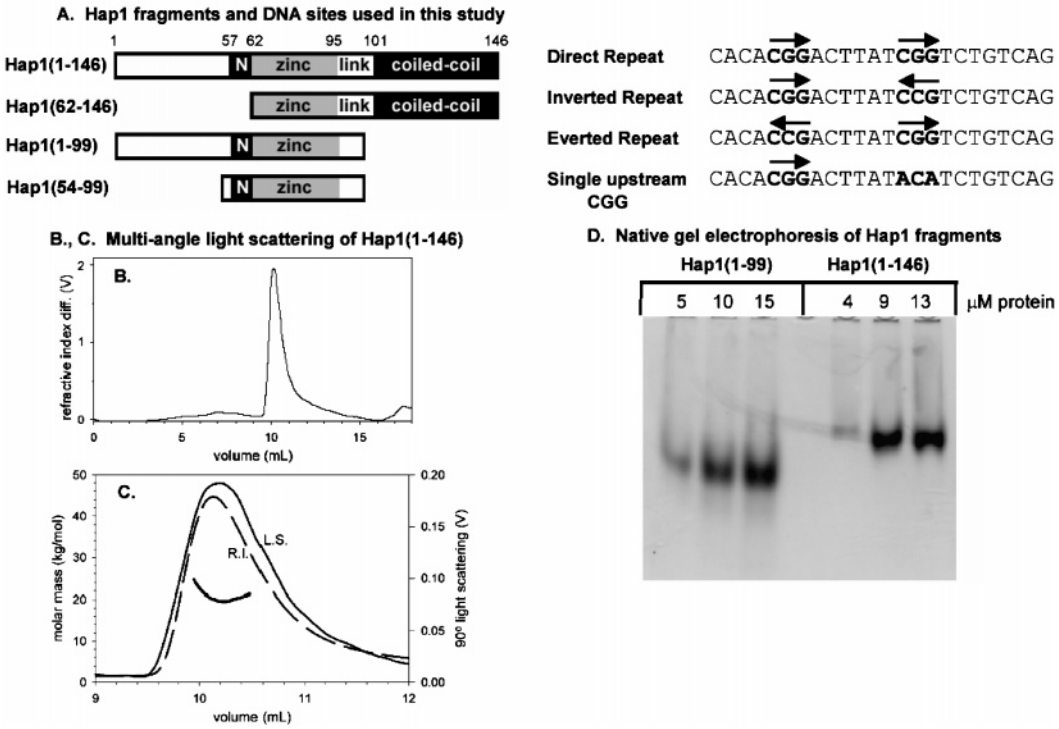

FIGURE 2: (A) Schematic representation of the Hap1 constructs and many of the DNA sites used in this study. The Hap1 domains as found in the crystal structure (*11*) are labeled "N" for the N-terminal residues that contact the spacer DNA, "zinc" for the zinc cluster domain, "link" for the linker domain, and "coil-coiled" for the dimerization helix. (B, C) Multiangle light scattering of Hap1(1−146). Panel B shows the refractive index difference signal for Hap1(1−146) eluting from a Superdex75 gel filtration column. Panel C shows the 90° light scattering signal (L.S., solid line), the refractive index difference signal (R.I., dashed line), and the calculated molar mass (dots) across the main refractive index peak at 10 mL in panel B. (D) Native electrophoretic gel of different concentrations of Hap1(1−99) (lanes 1−3) and of Hap1(1−146) (lanes 5−7) stained with Coomassie dye.

the canonical Hap1 direct repeat site with sites containing either a symmetric "inverted repeat" of CGG triplets (CGG-N_6-CCG) or a symmetric "everted repeat" of CGG triplets (CCG-N_6-CGG) (listed in Figure 2A). Symmetric repeats are typical of sites bound by other zinc cluster transcription factors, with CGG-N_6-CCG matching the sequence recognized by PPR1.

Gel shifts of DNA binding by Hap1(1−146) were very similar to those reported for Hap1(56-146) (*7*). As seen in Figure 3A, a titration of Hap1(1−146) to the direct repeat site caused DNA shifts consistent with DNA complexes containing either one or two bound Hap1 molecules, while a titration to the inverted repeat site showed only a shift for a single bound Hap1 molecule. In the inverted repeat, the orientation of the downstream CGG was reversed (Figure 2A). Reversing instead the upstream CGG to form the everted repeat greatly weakened all Hap1(1−146)−DNA complexes, both those containing a single Hap1 molecule and those containing two (Figure 3B and below). Since the everted repeat contained an unaltered downstream CGG, this CGG must be bound by Hap1 much more weakly than the upstream CGG that was followed by the N_3TA sequence. Consistent with this, Hap1(1−146) bound to a site containing only the upstream CGG triplet (downstream CGG changed to ACA) in a manner very similar to that of the inverted repeat (Figure 3B), and it bound to a site containing only the downstream CGG triplet in a manner very similar to that of the everted repeat (below). Underscoring the importance of the N_3TA sequence following the upstream CGG, changing the spacer TA to CT resulted in much weaker Hap1(1−146) binding (Figure 3C). A similar weakened binding was reported for the HAP(56−148) fragment when only the T in the spacer TA sequence was changed to C (*7*).

The binding titrations demonstrated that Hap1(1−146) bound cooperatively to the direct repeat site. At 4 μM, Hap1 formed complexes of both one and two bound Hap1 molecules on the direct repeat site (Figure 3A), but only complexes containing one Hap1 molecule on the site containing a single upstream CGG, and no complexes on the everted repeat site that contained an unaltered downstream CGG (Figure 3B). Thus, at 4 μM, Hap1 binding to the downstream CGG required an appropriately oriented upstream CGG, as found in the direct repeat. In all likelihood, binding of a Hap1 molecule to the upstream CGG stabilized the binding of a second Hap1 molecule to the downstream CGG through cooperative protein−protein interactions. This cooperativity in Hap1−DNA binding had been previously reported (*7*) and was likely mediated by the extensive dimer interface apparent in the Hap1−DNA crystal structure (*11*). This interface contained both coiled-coil and cluster domain−cluster domain contacts (Figure 1). However, the relative contribution of each of these contacts to cooperative dimerization was not known.

Compared with Hap1(1−146), the Hap1(62−146) fragment with all N-terminal residues deleted bound much more weakly to the direct repeat site (Figure 3A). For example, only a small fraction of the direct repeat site duplex was shifted by 4 μM Hap1(62−146) compared with >50% shifted by 4 μM Hap1(1−146) (Figure 3A). In addition, Hap1(62−146) showed only a single shifted band compared with the two bands apparent with Hap1(1−146). This single band likely represented a dimer complex of Hap1(62−146)

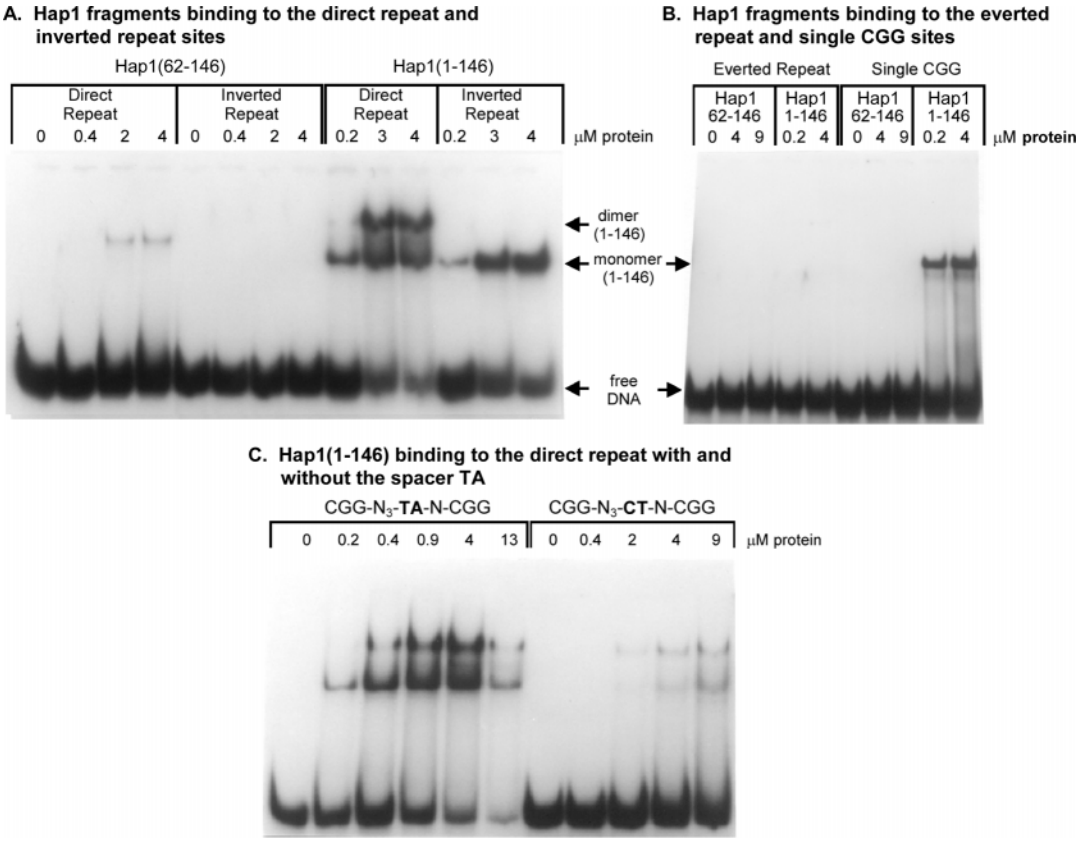

FIGURE 3: DNA gel shifts of Hap1(62−146) and Hap1(1−146) binding to different DNA duplexes. (A) DNA gel shifts of Hap1(62−146) (lanes 1−8) and Hap1(1−146) (lanes 9−14) binding to the direct and inverted repeat site duplexes. (B) DNA gel shifts of Hap1(62−146) (lanes 1−3 and 6−8) and Hap1(1−146) (lanes 4−5 and 9−10) binding to the everted repeat and single upstream CGG sites. (C) DNA gel shifts for Hap1(1−146) binding to a direct repeat site with the spacer TA (CACACGGACTTATCGGTCTGTCAG) and without the spacer TA (CACACGGACTCTTCGGTCTGTCAG).

since its mobility was between those of the monomer and dimer bands of the Hap1(1−146), as expected for its total molecular weight (Figure 3A). The much weaker binding of Hap1(62−146) indicated that the missing N-terminal residues provided substantial affinity in Hap1−DNA binding that was especially critical for monomeric binding. In addition, while they might have contributed to dimerization in Hap1(1−146), the N-terminal residues were not required for dimerization in Hap1(62−146).

Despite the weaker binding to the direct repeat site, Hap1-(62−146) still exhibited a preference for the direct repeat site over the inverted repeat site, for which no band shift was observed (Figure 3A). No shift of the inverted repeat site was detected even when the Hap1(62−146) concentration was increased to 20 μM (not shown). Hap1(62−146) also did not bind to the everted repeat or to the single upstream CGG sites (Figure 3B). This suggested that dimerization required a direct repeat orientation of CGG triplets.

Hap1 N-Terminal and Cluster Domain Residues Are Sufficient for Dimerization. A previous study showed that a chimeric protein containing the Hap1 cluster domain and six N-terminal residues (Hap1 residues 56−95) fused to the linker and dimerization helix of PPR1 maintained a DNA-binding specificity for the Hap1 direct repeat site. This indicated that the Hap1 cluster domain and N-terminal residues were sufficient to specify binding to the Hap1 direct repeat site, perhaps by contributing to dimerization (*7*). Another study showed that Hap1 fragments missing dimerization helix residues 105−135 still dimerized on DNA although these fragments contained additional C-terminal residues predicted to form dimer helices (*15*). To directly test whether and by how much these residues alone (without other dimerization elements) contribute to Hap1−DNA site discrimination, we compared the binding of the Hap1(1−146) fragment with that of Hap1(1−99), for which the last linker residue (Trp100) and all additional C-terminal residues were deleted. If Hap1 dimerized in the same manner as other zinc cluster transcription factors, Hap1(1−99) was expected to lose all cooperativity and show only monomer binding to the direct repeat site.

Figure 4 shows titrations of Hap1(1−146) and Hap1(1−99) binding to the direct repeat DNA-binding site. Like Hap1(1−146), Hap1(1−99) caused two shifted DNA bands, indicating two distinct protein−DNA complexes (Figure 4). Consistent with these being complexes containing one and two bound Hap1(1−99) molecules, each of the two shifted bands migrated faster than the corresponding band observed with the larger Hap1(1−146) as seen in the last lane of Figure 4B (a comparison is also shown in Figure 4A with Hap1(1−99) in the last lane).

Hap1(1−99) binding and Hap1(1−146) binding to the inverted and everted repeat sites and to the single CGG sites were also compared (Figure 5A,B). In all cases the pattern of Hap1(1−99) binding was much more similar to that of Hap1(1−146) than that of Hap1(62−146) seen in Figure 3. Both Hap1(1−99) and Hap1(1−146) showed relatively high affinity, monomer-only binding to the inverted repeat, and much weaker binding to the everted repeat. Also like Hap1-

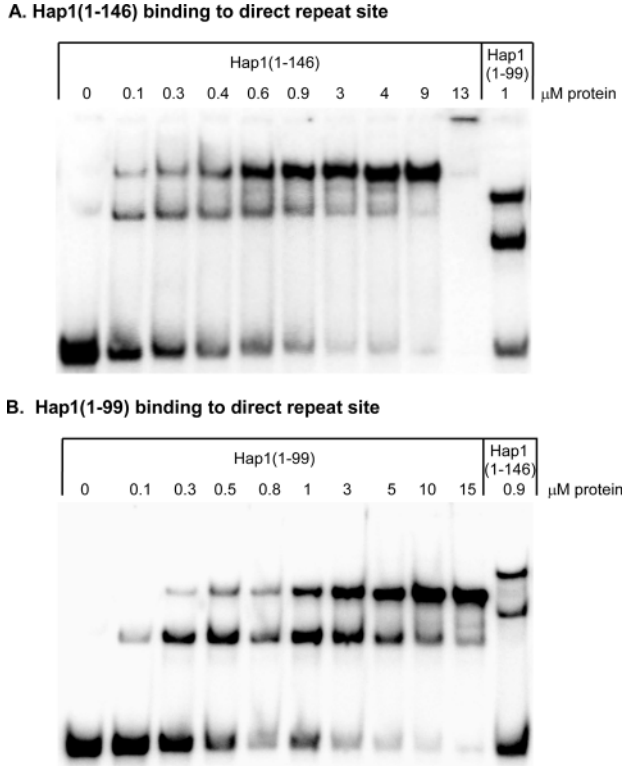

FIGURE 4: DNA gel shifts for Hap1(1−146) (A) and Hap1(1−99) (B) binding to the direct repeat binding site, CACACGGACTTATCGGTCTGTCAG. For reference, the last lane in the Hap1(1−146) titration gel in (A) shows binding by Hap1(1−99) and the last lane in the Hap1(1−99) gel in (B) shows binding by Hap1(1−146). In lane 10 in (A) much of the DNA was trapped in the well. This was periodically observed at the higher concentrations of Hap1 fragments.

(1−146), Hap1(1−99) bound similarly well to the inverted repeat and to the single upstream CGG (with N_3TA) sites (Figure 5A,C), and it bound weakly to the everted repeat and to the single downstream CGG (no N_3TA) sites (Figure 5B,D). Changing the spacer TA to CT greatly weakened Hap1(1−99) binding as it did for Hap1(1−146), verifying the importance of the Hap1−TA interaction (Figure 6). Only a small amount of the CT-containing site was shifted by 7 μM Hap1(1−99), while >50% of the TA-containing site was shifted by 5 μM HAP(1−99) (lanes 1 and 6 in Figure 6). Thus, both Hap1 fragments had greater affinity for the upstream CGG followed by N_3TA than for the downstream CGG with no TA.

As with Hap1(1−146), Hap1(1−99) binding to the direct repeat site was cooperative, with the binding of a Hap1(1−99) molecule to the upstream CGGN_3TA enhancing the binding of a second Hap1(1−99) molecule to the downstream CGG without the N_3TA. For example, at 1 μM, Hap1(1−99) bound as a monomer to the upstream-only CGG site but did not bind at all to the downstream-only CGG site (Figure 5C,D). When the upstream and downstream CGGs were combined in the direct repeat site, 1 μM Hap1(1−99) shifted much of the DNA as a complex with two bound Hap1 molecules, indicating that binding occurred to the downstream as well as to the upstream CGG (Figure 4). Thus, despite the elimination of all C-terminal dimerization helices, Hap1(1−99) retained the ability to bind cooperatively as a dimer to the direct repeat site. The dimerization most likely arose from within residues 1−95 (N-terminal residues and cluster domain) since linker residues 96−99 did not directly participate in the dimer interface in the Hap1−DNA crystal structure.

Hap1(1−99) and Hap1(1−146) did show some differences in DNA binding. While Hap1(1−146) bound to the direct repeat over the same concentration range as Hap1(1−99), it showed slightly greater cooperativity as evidenced by higher ratios of dimer to monomer complex for similar amounts of DNA shifted (Figure 4). This was verified by quantitating the band intensities as described below. The greater cooperativity likely resulted from additional dimerization contacts made by the C-terminal linker and helix residues in Hap1(1−146). A second difference was that while Hap1(1−146) and Hap1(1−99) bound to the direct repeat site over similar ranges of protein concentration, Hap1(1−146) consistently bound more weakly to the single CGG, the inverted repeat, and the everted repeat sites than did Hap1(1−99) (Figure 5). Since Hap1(1−146) was primarily monomeric in solution (Figure 2), the C-terminal linker and helix residues of Hap1(1−146) may have been unstructured in the monomeric state and impeded monomer binding. Hap1(1−146) also differed from Hap1(1−99) in showing very weak dimer formation on the everted repeat and the single downstream sites, while Hap1(1−99) showed only monomer binding to these sites (Figure 5B,D). This could reflect weak nonspecific binding or weak dimerization mediated by the C-terminal linker and helix domains.

Hap1 Residues 1−53 Do Not Significantly Contribute to DNA Binding. In the crystal structure of the Hap1(56−135)−DNA complex, residues 57−60 preceding the Zn_2Cys_6 cluster domain were responsible for making the extensive contacts to the spacer DNA. In addition to these residues, the Hap1(1−146) and Hap1(1−99) fragments also contained the preceding residues 1−55. To determine if residues 1−55 contributed to Hap1−DNA binding and/or dimerization, a fragment containing Hap1 residues 54−99 was generated and its binding was compared with that of Hap1(1−99). As seen in Figure 7, Hap1(54−99) bound to the direct repeat, the single upstream CGG, and the single downstream CGG duplexes in a manner very similar to that of Hap1(1−99) (compare with Figures 4 and 5). The DNA shifts were consistent with the formation of monomer and dimer complexes on the direct repeat site but only monomer complexes on the single CGG duplexes. As expected given the smaller size of Hap1(54−99), the monomer and dimer complexes showed greater mobility than the monomer complex of HAP(1−99). Like Hap1(1−99), Hap1(54−99) bound the upstream CGG duplex with much greater affinity than the downstream CGG duplex (Figure 7A). When both were tested at 2 μM protein, Hap1(54−99) and Hap1(1−99) shifted similar proportions of DNA into monomer and dimer complexes, indicating similar binding affinities (Figure 7B; by phosphorimage quantitation, 33% of the total DNA was shifted as monomer complexes by both proteins, while 54% [Hap1(54−99)] and 59% [Hap1(1−99)] were shifted into dimer complexes). Thus, Hap1 residues 1−53 did not contribute significantly to Hap1 monomer or dimer binding to DNA.

Relative Strengths of Hap1 Intermolecular Interactions. To estimate the relative strengths of the Hap1 dimerization and the Hap1−DNA interactions, the gel-shift data were quantified by phosphorimaging and analyzed by a simple model in which Hap1 molecules bound a high-affinity CGG

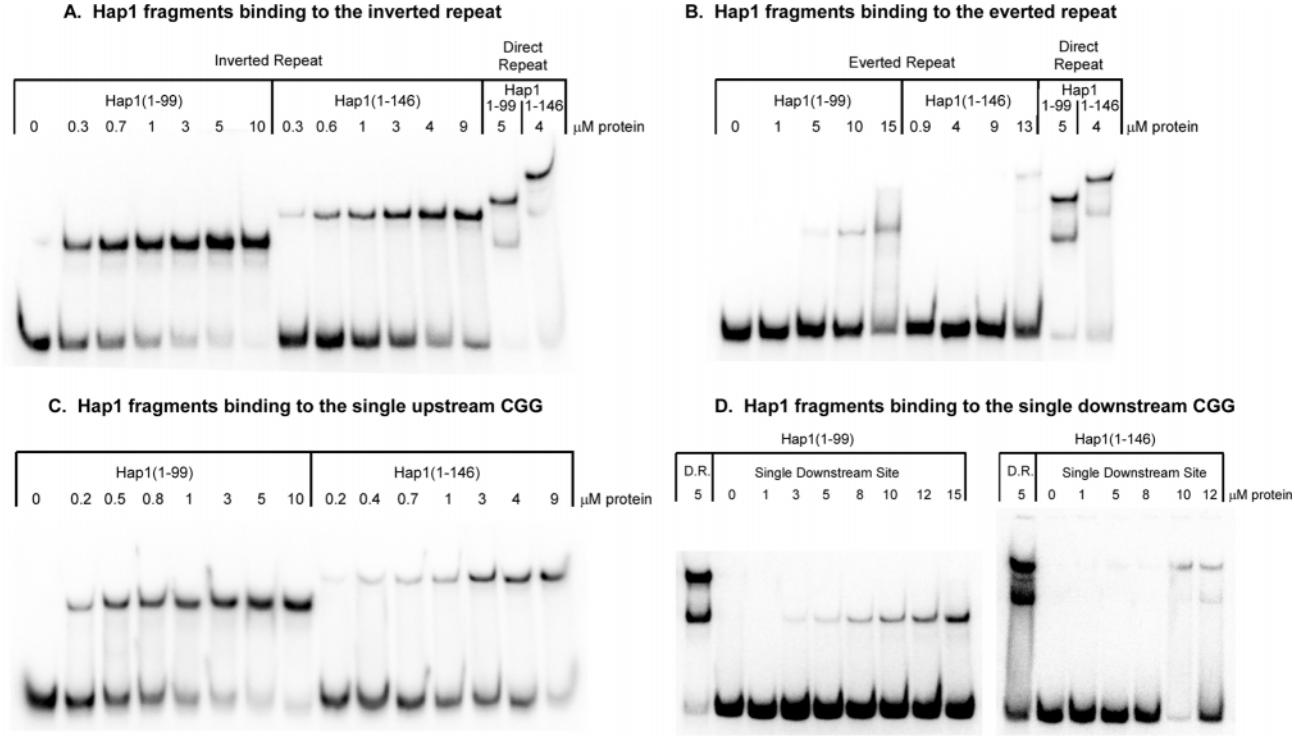

FIGURE 5: Comparisons of Hap1(1−99) and Hap1(1−146) binding to different DNA sites. (A) Hap1(1−99) (lanes 1−7) and Hap1(1−146) (lanes 8−13) binding to the inverted repeat site. Lanes 14 and 15 show binding of each Hap1 fragment to the direct repeat site for reference. (B) Hap1(1−99) (lanes 1−5) and Hap1(1−146) (lanes 6−9) binding to the everted repeat site. Lanes 10 and 11 show binding of each Hap1 fragment to the direct repeat site for reference. (C) Hap1(1−99) (lanes 1−8) and Hap1(1−146) (lanes 9−15) binding to the single upstream CGG site. (D) Hap1(1−99)and Hap1(1−146) binding to the single downstream CGG site, with binding to the direct repeat (D.R.) shown in lanes 1 for reference.

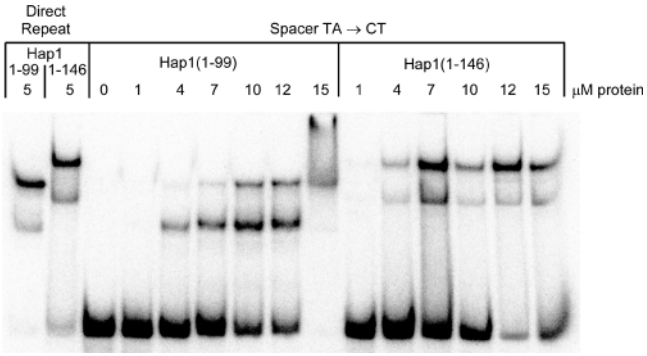

FIGURE 6: Hap1(1−99) and Hap1(1−146) binding to the direct repeat site containing no spacer TA. The spacer TA was changed to CT (CACACGGACTCTTCGGTCTGTCAG). Lanes 1 and 2 show Hap1(1−99) and Hap1(1−146) binding to the direct repeat site for reference. DNA retention in the well in lane 9 may have resulted from sample precipitation.

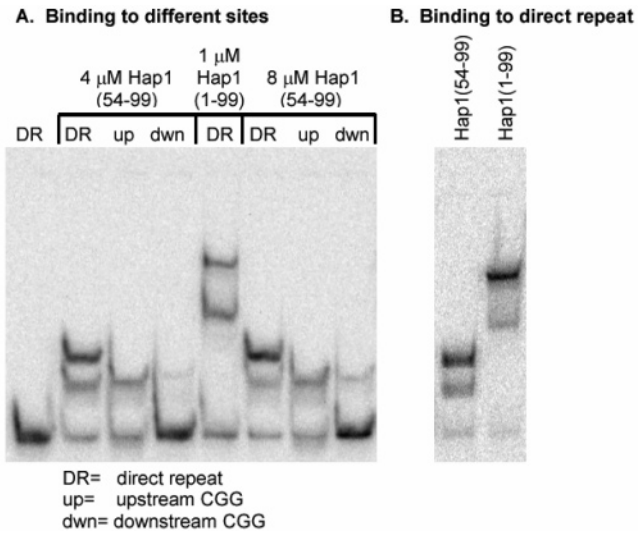

FIGURE 7: Comparison of DNA binding by Hap1(54−99) and Hap1(1−99). (A) DNA gel shifts for Hap1(54−99) binding to the direct repeat (DR), the single upstream CGG (up), and the single downstream CGG (dwn) sites. Lane 5 shows binding by Hap1-(1−99) to the direct repeat site for reference. (B) DNA gel shifts for 2 μM Hap1(54−99) and 2 μM Hap1(1−99) binding to the direct repeat.

(CGGN$_3$TA) with dissociation constant K_H and a low-affinity CGG (CGG with no N$_3$TA) with dissociation constant K_L and dimerized cooperatively on a direct repeat with dissociation constant K_C (see the Materials and Methods). Only relative differences in dissociation constants could be determined since the gel shifts do not represent true solution equilibrium conditions, and since the weak binding to the low-affinity CGG sites precluded accurate quantitation. All analyses were based on experiments employing a constant 0.05 μg/μL nonspecific DNA (~2000-fold excess in base pairs compared to the labeled DNA). In the absence of nonspecific DNA, 1 μM Hap1(1−146) shifted essentially all the direct repeat site duplex (not shown), consistent with a 46 nM K_D reported for Hap1(55−135) binding in solution to a direct repeat in the absence of nonspecific DNA (*10*). The 0.05 μg/μL nonspecific DNA used in these experiments was sufficient to block Hap1 shifts of an 18 base pair duplex containing no CGG or CGC (sequence CTTGACTGACT-GATCAAG, data not shown).

The monomeric binding of Hap1(1−99) to the single upstream CGG and the inverted repeat sites in Figure 5A,C

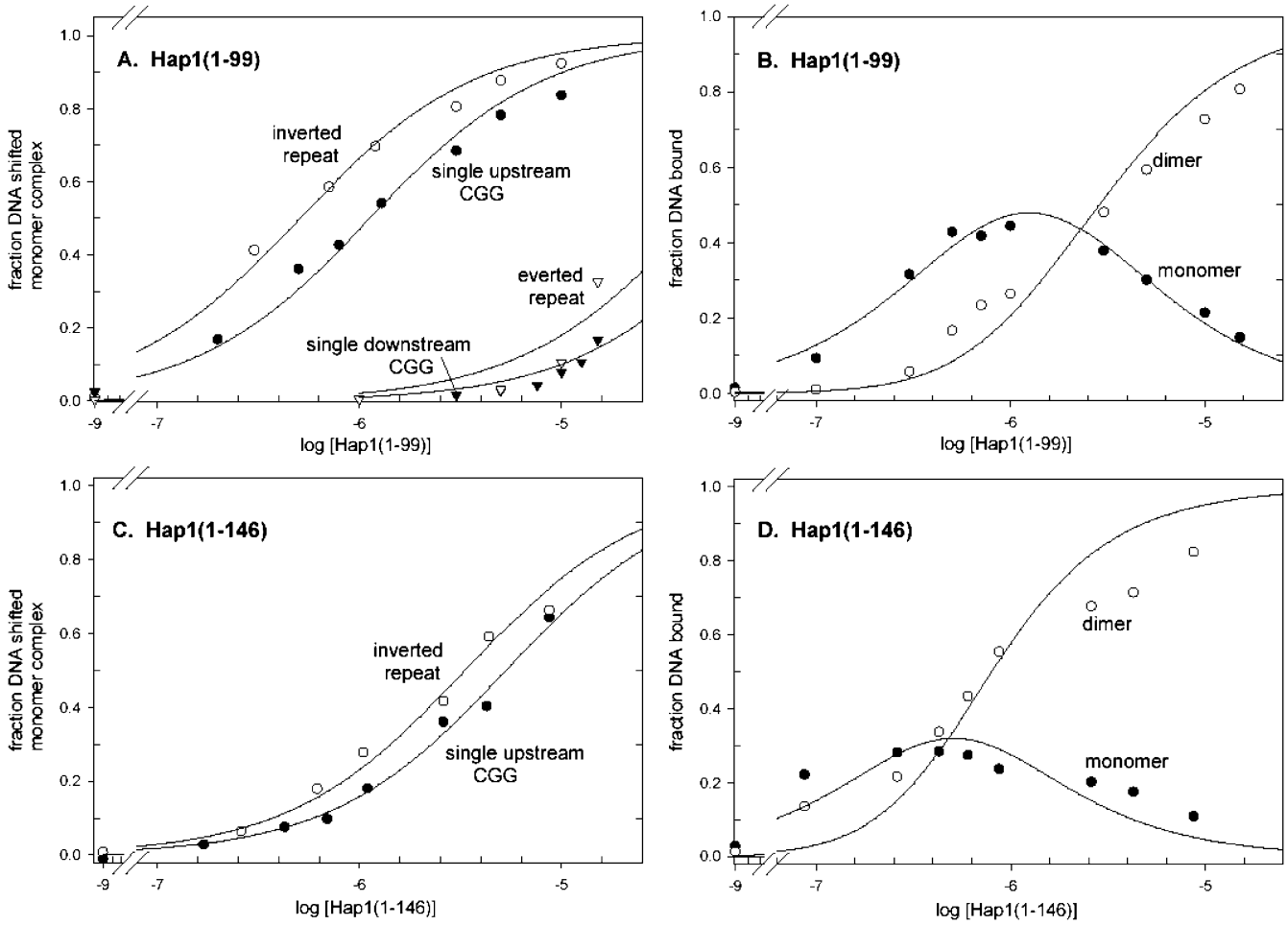

FIGURE 8: Fits of Hap1(1−99) and Hap1(1−146) DNA gel-shift data by binding models. (A) Fits of Hap1(1−99) binding to single CGG, inverted repeat, and everted repeat sites. The upstream single CGG site data are from Figure 5C and were fit with $K_H = 1.1$ μM, the inverted repeat data are from Figure 5A and were fit with $K_H/(1 + K_H/K_L) = 0.51$ μM (the average dissociation constant for two CGGs), the single downstream CGG site data are from Figure 5D and were fit with $K_L = 100$ μM, and the everted repeat data are from Figure 5B and were fit with $K_L = 50$ μM. Due to weak binding, only the last two points were used to determine K_L for the single downstream CGG and everted repeat sites. (B) Fit of Hap1(1−99) binding to the direct repeat site. The data are from Figure 4 and were fit with $K_H = 0.68$ μM and $K_C = 0.023$ with K_L constrained to 100 μM. (C) Fits of Hap1(1−146) binding to the single upstream CGG and inverted repeat sites. The single upstream CGG site data are from Figure 5C and were fit with $K_H = 5.3$ μM, and the inverted repeat data are from Figure 5A and were fit with $K_H/(1 + K_H/K_L) = 3.4$ μM. (D) Fit of Hap1(1−146) binding to the direct repeat site. The data are from Figure 4 and were fit with $K_H = 0.54$ μM and $K_C = 0.0019$ with K_L constrained to 250 μM.

were fit by single-site binding isotherms with $K_H = 1.1 \pm 0.1$ and 0.51 ± 0.04 μM, respectively (Figure 8A), suggesting a dissociation constant of 0.5−1 μM. Formally, the inverted repeat fit provided $1/K = 1/K_H + 1/K_L$ since two CGGs were present, but K_L was $\gg K_H$ (below). The single downstream CGG and the everted repeat site data (Figure 5B,D) were too weak to accurately fit K_L, so K_L was estimated to be ∼50−100 μM using the percent DNA shifted at the highest Hap1(1−99) concentrations. Even with the uncertainty in K_L, the spacer TA clearly enhanced monomeric CGG binding by Hap1(1−99) on the order of 100−200-fold, corresponding to values of $\Delta\Delta G°$ [$= -RT \ln(K_L/K_H)$] of −2.7 to −3.1 kcal/mol.

Hap1(1−99) binding to the direct repeat site in Figure 4B was fit by the two-CGG model with $K_H = 0.68 \pm 0.05$ μM and $K_C = 0.023 \pm 0.002$ when K_L was constrained to 100 μM (Figure 8B). The data for the monomer and dimer complexes were fit together to obtain the best overall fit. K_H was within the range of 0.5−1.1 μM determined from the upstream single CGG and inverted repeat site data. Although K_L was less well determined, the product $K_C K_L$ was well determined since changes in K_L were compensated by opposite changes in K_C in the fitting to generate identical curves (this was because $K_L \gg K_H$). The best fit for $K_C K_L$ was 2.3 μM, which was only ∼3-fold greater than $K_H = 0.68$ μM ($K_C K_L/K_H = 3.4$, Table 1), indicating that the energy of Hap1(1−99) dimerization on the direct repeat largely compensated for the absence of the TA in the downstream CGG site. $K_C K_L/K_H$ only varied from 2.2 to 4.0 for K_H

Table 1: Dissociation Constants for HAP1 Fragment Binding to the Direct Repeat Site

	dissociation constants[a]			
fragment	K_H (μM)	K_L (μM)	K_C	$K_C K_L$ (μM)
Hap1(1−99)	0.68 ± 0.05	(100)	0.023 ± 0.002	2.3 ± 0.2
Hap1(1−146)	0.54 ± 0.15	(250)	0.0019 ± 0.0004	0.48 ± 0.12

[a] K_H and K_L are dissociation constants for Hap1 binding to a high-affinity CGG-containing site and a low-affinity CGG-containing site, respectively. K_C is a dissociation constant for Hap1 cooperative dimerization on the DNA. K_L was constrained to the value in parentheses on the basis of fits of the single downstream CGG and the everted repeat binding data.

constrained to 0.5−1.1 μM. In terms of free energy, the dimerization of Hap1(1−99) provided a $\Delta G^\circ = -RT \ln(1/K_C)$ of approximately −2.2 kcal/mol of stabilization.

Fits of Hap1(1−146) DNA-binding data followed the same trends as those of Hap1(1−99). Binding to the single upstream CGG and that to the inverted repeat duplexes in Figure 5A,C were fit with $K_H = 3.4-5.3$ μM (Figure 8C), weaker than the 0.5−1.1 μM for Hap1(1−99). An upper limit of $K_L \geq 250$ μM, also weaker than the 50−100 μM for Hap1(1−99), was estimated from the amount of monomer complex on the everted repeat at 13 μM Hap1(1−146) (Figure 5D). With K_L constrained to 250 μM, Hap1(1−146) binding to the direct repeat in Figure 4 was best fit with $K_H = 0.54 \pm 0.15$ μM and $K_C = 0.0019 \pm 0.0004$ (Figure 8D). $K_C K_L$ was a constant 0.48 μM for different constrained values of K_L. Unlike for Hap1(1−99), the K_H determined from the Hap1(1−146) direct repeat data was not consistent with the upstream single CGG and inverted repeat titrations (0.54 μM vs 3.4−5.3 μM), suggesting greater complexity in DNA binding. Nonetheless, Hap1(1−146) appeared to exhibit greater cooperativity than Hap1(1−99) as evidenced by smaller K_C and $K_C K_L$ (Table 1). For $K_L = 250$ μM, the ΔG° [$= -RT \ln(1/K_C)$] for Hap1(1−146) dimerization was −3.7 kcal/mol compared with −2.2 kcal/mol for Hap1(1−99).

Asymmetry in Hap1 Dimer Formation on the Direct Repeat. In the crystal structure of Hap1 bound to the CGC direct repeat site, the two Hap1 fragments in the dimer do not make identical DNA contacts. The upstream Hap1 fragment makes TA base contacts to the neighboring N_3TA sequence with its N-terminal residues, while the downstream Hap1 fragment does not, despite the downstream presence of an identical N_3TA sequence (Figure 1). We wished to determine if an N_3TA sequence following the downstream CGG in a direct repeat site could stabilize Hap1 dimer formation if no N_3TA were present in the upstream CGG. As seen in Figure 9, removing the upstream spacer TA in the direct repeat site greatly weakened Hap1(1−146) DNA binding, as described above. Inserting the TA after the downstream CGG almost fully restored binding to the direct repeat (Figure 9A). Essentially identical results were obtained with Hap1(1−99) (Figure 9B; compare with the original upstream TA site in Figure 4 and with the no-TA site in Figure 6). These results indicate that a Hap1 molecule can make base contacts to an N_3TA sequence following the downstream CGG and cooperatively stabilize binding of another Hap1 molecule to the upstream CGG. This is the converse for Hap1 binding to the original direct repeat, where binding of a Hap1 molecule to an upstream CGGN_3TA stabilized binding of a second Hap1 molecule to the downstream CGG without the N_3TA.

For both Hap1(1−146) and Hap1(1−99) the cooperativity in dimer formation was not fully restored in binding to the CGG-N_6-CGG-N_3TAN site compared with the original CGG-N_3TAN-CGG-N_6 site. As seen in Figure 9A, the ratio of dimer to monomer complexes at 0.9 and 4.3 μM Hap1(1−146) was much lower for the downstream N_3TA-containing site than for the upstream N_3TA-containing site. A similar reduction in the ratio of dimer to monomer complexes was observed with Hap1(1−99) (compare Figure 9B with Figure 4). The Hap1(1−99) titration to the site containing the downstream N_3TA sequence in Figure 9B was best fit by the two-CGG-binding model with $K_H = 1.0$ μM and $K_C = 0.16$ when K_L was constrained to 100 μM (Figure 9C). The monomeric binding affinities K_H and K_L were consistent with those for the direct repeat site although the high-affinity site was now downstream and the low-affinity site was upstream. However, $K_C = 0.16$ was much larger (weaker dimerization) than the $K_C = 0.022$ determined for the original direct repeat site. Thus, moving the N_3TA sequence from the upstream to the downstream CGG triplet destabilized dimerization. This could occur if the TA interaction stabilized the dimer interface of the Hap1(1−99) molecule bound to the upstream CGG triplet or if it forced the downstream Hap1(1−99) molecule to adopt an altered position over the DNA that weakened its contacts with the upstream-bound Hap1 molecule (see the Discussion). In either case the contributions to DNA recognition

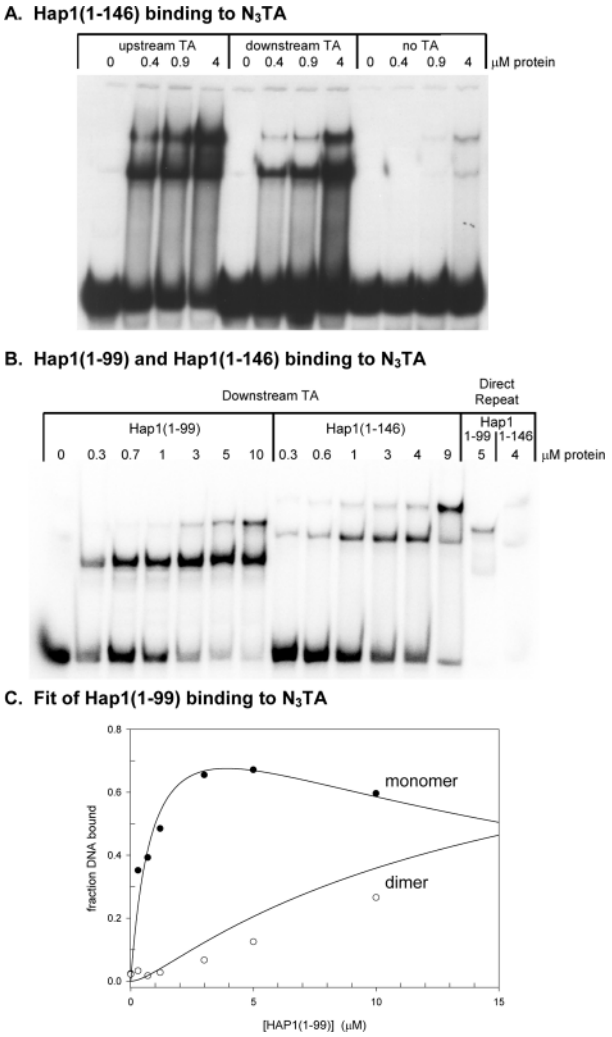

FIGURE 9: DNA gel shifts for Hap1(1−99) and Hap1(1−146) binding to direct repeat sites containing different positions of the N_3TA sequence. (A) Gel shift for Hap1(1−146) binding to the direct repeat containing an upstream TA only (in the spacer; lanes 1−4), a downstream TA only (lanes 5−8), or no TA (lanes 9−12). (B) Hap1(1−99) (lanes 1−7) and Hap1(1−146) (lanes 8−13) binding to the direct repeat containing a downstream TA only. Lanes 14 and 15 show binding to the site containing the upstream TA only. (C) Fit of the Hap1(1−99) titration to the downstream TA site (data in (B)) with the direct repeat site binding model. The curve is for $K_H = 0.95$ μM, $K_C = 0.16$, and $K_L = 100$ μM. The sequences for the binding sites were CACACGGACTTATCGGTCTGTCAG for the upstream TA, CACACGGACTCTTCGGTCTTACAG for the downstream TA, and CACACGGACTCTTCGGTCTGTCAG for the no-TA sites.

made by the N_3TA contacts and the dimer contacts are not completely independent but show some interdependence.

DISCUSSION

The level of expression of genes in eukaryotes is largely determined by the composition of large multiprotein complexes formed at the genes' promoters. These complexes include transcription factors whose presence depends on cellular conditions and signaling pathways extending beyond the nucleus. The complexes are assembled through multiple protein−protein and protein−DNA interactions so that the overall level of transcription of a gene depends on the summation of multiple distinct macromolecular interactions. The binding of Hap1 to its DNA site provides a simple model for quantifying how the multiple individual interactions observed in complex protein−DNA assemblies contribute to overall stability and DNA site specificity of multiprotein−DNA complexes. Hap1 exhibits protein−base contacts in both the major and the minor grooves of the DNA, nonspecific base contacts to the phosphodiester backbone, and protein−protein contacts in dimerization. Hap1 also provides a model for understanding how the interplay of multiple interactions can affect the geometry of the complex, which in turn can further modulate the level of transcription. By comparing the binding of different Hap1 fragments to different DNA-binding sites, this study provides insight into how individual macromolecular interactions contribute to DNA sequence recognition and DNA-binding stability.

Hap1−Spacer DNA Contacts. The minor groove spacer contacts by residues preceding a cluster domain are one of the distinguishing features of the Hap1−DNA crystal structure. This interaction is similar to that observed for homeodomains in which major groove contacts by a helix−turn−helix motif are augmented with minor groove contacts by extended N-terminal residues (see, e.g., refs *16* and *17*). For Hap1, this minor groove interaction is especially critical for stabilizing monomer binding. Deleting all residues preceding the cluster domain in the Hap1(62−146) fragment abolished all monomer binding. The only DNA shift observed with Hap1(62−146) was consistent with a dimer complex, suggesting that removing the N-terminal residues effectively turned Hap1 into a conventional zinc cluster protein that depended on dimerization for DNA association. Furthermore, Hap1(62−146) bound DNA much more weakly than Hap1(1−146), indicating that the spacer DNA contacts contributed substantial energy for binding DNA of ∼3 kcal/mol. Much of this energy must come from base contacts (vs phosphodiester contacts) since changing the spacer TA to CT was sufficient to greatly weaken Hap1(1−146) and Hap1(1−99) DNA binding. Consistent with the crystal structure of the Hap1−DNA complex, the critical N-terminal residues must lie within residues 54−61 since deleting residues 1−53 in Hap1(54−99) did not significantly alter monomer and dimer binding compared with that of Hap1(1−99).

Hap1−Hap1 Dimer Contacts. One surprise in the Hap1−DNA binding analysis was finding that the cluster domain−cluster domain interaction observed in the crystal structure was sufficient to dimerize Hap1 on DNA. Although previous studies suggested this interaction could contribute to dimerization, the Hap1 fragments used either included a dimerization domain from another zinc cluster protein (PPR1) or had helix residues 105−135 deleted but contained residues following 135 that were predicted to contain dimerization helices (*7*, *15*). In this study the Hap1(1−99) and Hap1(54−99) fragments contained only four linker residues (residues 96−99) after the cluster domain but still dimerized almost as well as the Hap1(1−146) that contained the C-terminal dimerization helix. The fact that Hap1(62−146) also dimerized despite having all residues preceding the cluster domain deleted suggested that the N-terminal residues preceding the cluster domain were not required for dimerization. In the X-ray crystal structure, the interface formed by Hap1 residues 62−99 comprised approximately 150 Å^2 of largely nonpolar buried surface. The analysis of Hap1(1−99) binding to the direct repeat site indicated that this interaction provided ∼2.2 kcal/mol of stability, or about 2/3 the stability provided by the N-terminal residues interacting with the spacer TA bases. The magnitude of the stability was close to a predicted 3.75 kcal/mol assuming 25 cal/(mol·Å^2) of buried nonpolar surface (*18*). The cluster domain residues that participate in this interface are not conserved within the family of Zn_2Cys_6 binuclear cluster proteins, underscoring the importance of this interface in specifying Hap1's recognition for a direct repeat site with the six base pair spacer.

In addition to the cluster domain−cluster domain interaction, the crystal structure of Hap1 bound to the CGC direct repeat site also showed dimer contacts mediated by linker and helix residues C-terminal to the cluster domain (Figure 1). The linker and C-terminal helix in Hap1 did contribute to dimerization as Hap1(1−146) exhibited greater cooperativity than Hap1(1−99) in binding to the CGG direct repeat site corresponding to an additional ∼1.5 kcal/mol of dimerization energy. This 1.5 kcal/mol represents a lower limit for the linker−helix contribution to Hap1 binding since it may include an unfavorable reduction in entropy from structural ordering of the linker and C-terminal helix residues upon Hap1(1−146) dimerization.

Interplay between Hap1−Spacer DNA and Hap1 Dimer Contacts. The Hap1−spacer DNA and the Hap1−Hap1 dimer contacts were not completely independent of each other but were energetically coupled. Moving the N_3TA sequence from the upstream CGG to the downstream CGG maintained high-affinity monomer binding by Hap1(1−146) and Hap1(1−99) but greatly reduced the cooperativity for dimer formation (Figure 9). Optimal dimerization of Hap1 required an N_3TA sequence that specifically followed the upstream CGG triplet in the spacer DNA.

The coupling between spacer TA binding and dimerization may have resulted from the TA interaction affecting the cluster domain orientation on the DNA. In the crystal structure of Hap1 bound to the CGC direct repeat, both CGCs were followed by N_3TA but only the upstream Hap1 molecule bound the TA (*11*). Compared with the downstream cluster domain−CGC complex, the upstream domain tilted by ∼20° from near the end of its first helix (residue 71 in Figure 1) toward the spacer DNA. When overlaid by their bound CGCs, the Cα atoms of residues 75−91 of the two domains differed by >3 Å (>5 Å for Lys86 and Thr87). Many of these residues were in the dimer interface (Figure 1). For the downstream cluster domain, the more upright orientation may have enhanced dimerization but prevented optimal TA binding. The Hap1 variant Ser63Gly bound a

CGC direct repeat with a geometry similar to that of wild-type Hap1 and did contact the downstream TA (*10*). However, it made only a single contact with a more N-terminal Arg55 instead of the multiple spacer TA contacts by wild-type Arg57 and Arg59.

When the N_3TA sequence was shifted from the upstream CGG (in the spacer DNA) to the downstream CGG, Hap1-(1−146) and Hap1(1−99) both recognized the TA and bound more strongly to the downstream CGG. This could have caused the cluster domain to tilt further downstream and away from an upstream Hap1 molecule. Simultaneously, an upstream Hap1 molecule with no spacer TA to contact would have been less stabilized in adopting the tilt toward the downstream cluster domain. The affinity for monomeric Hap1(1−146) and Hap1(1−99) binding would have been unchanged, as observed, while the weaker cooperativity in dimer formation would have resulted from reduced dimer contacts or from strain introduced into the DNA to maintain the dimer interface with a downstream TA contact.

In native Hap1-binding sites in yeast the N_3TA is more highly conserved in the spacer DNA than after the downstream CGG triplet. Among seven Hap1-binding sites from the CYC1, CYC7, CTT1, CYB2, and CYT1 genes (including two possible orientations for the CYT1 gene) (*19*, *20*), the spacer T and A are present six and five times, respectively, in the spacer but only two and one time, respectively, after the downstream CGG. In addition, in vitro selection identified a preferred consensus Hap1-binding sequence of CGG-N_3TAN-CGG, with little sequence preference exhibited for nucleotides downstream of the second CGG triplet (*6*). Thus, the role of the spacer TA in optimizing Hap1(1−146) and Hap1(1−99) dimer formation may be fundamental to Hap1 site recognition in vivo.

The optimal assembly of the Hap1 dimer depends on a mutual reinforcement of distinct intermolecular interactions. Since the N-terminal residue−TA interaction and the Hap1 dimer contacts are of comparable magnitudes, the overall stability and geometry of the assembly are sensitive to changes in either interaction, such as the location of the TA interaction. Changes in geometry are likely as important as changes in stability, as alternate geometries of Hap1−DNA association have been implicated as causing different levels of transcriptional activation (see, e.g., refs *8*−*10*). In the large multi-transcription-factor complexes found at gene promoters the complexity in the interplay of these intermolecular interactions is vastly magnified. The results with Hap1 underscore how the large number of these interactions allows for the broad ranges of transcriptional modulation needed by living systems.

ACKNOWLEDGMENT

We thank Katrina Grape for generating the pMAL−Hap1-(1−99) expression plasmid, and Prashant Mishra in the laboratory of Dr. Rama Ranganathan, Department of Pharmacology, Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, for his expert assistance in the multiangle light scattering measurements. This paper is dedicated to the memory of Joseph E. Coleman, Department of Molecular Biophysics and Biochemistry, Yale University, in whose laboratory this study was initiated.

REFERENCES

1. Zhang, L., and Hach, A. (1999) Molecular mechanism of heme signaling in yeast: the transcriptional activator Hap1 serves as the key mediator, *Cell. Mol. Life Sci. 56*, 415−426.
2. Schwabe, J. W. R., and Rhodes, D. (1997) Linkers made to measure, *Nat. Struct. Biol. 9*, 680−683.
3. Hellauer, K., Rochon, M.-H., and Turcotte, B. (1996) A novel DNA binding motif for yeast zinc cluster proteins: the Leu3p and Pdr3p transcriptional activators recognize everted repeats, *Mol. Cell. Biol. 16*, 6096−6102.
4. Lenouvel, F., Nikolaev, I., and Felenbok, B. (1997) *In vitro* recognition of specific DNA targets by AclR, a zinc binuclear luster activator different from other proteins of this class, *J. Biol. Chem. 272*, 15521−15526.
5. Akache, B., Wu, K., and Turcotte, B. (2001) Phenotypic analysis of genes encoding yeast zinc cluster proteins, *Nucleic Acids Res. 29*, 2181−2190.
6. Zhang, L., and, and Guarente, L. b. (1994) The yeast activator HAP1-a GAL4 family member-binds DNA in a directly repeated orientation, *Genes Dev. 8*, 2110−2119.
7. Zhang, L., and Guarente, L. (1996) The C6 zinc cluster dictates asymmetric binding by HAP1, *EMBO J. 15*, 4676−4681.
8. Pfeifer, K., Prezant, T., and Guarente, L. (1987) Yeast HAP1 activator binds to two upstream activation sites of different sequence, *Cell 49*, 19−27.
9. King, D. A., Zhang, L., Guarente, L., and Marmorstein, R. (1999) Structure of HAP1-18-DNA implicates direct allosteric effect of protein- DNA interactions on transcriptional activation, *Nat. Struct. Biol. 6*, 22−7.
10. Lukens, A. K., King, D. A., and Marmorstein, R. (2000) Structure of HAP1-PC7 bound to DNA: implications for DNA recognition and allosteric effects of DNA-binding on transcriptional activation, *Nucleic Acids Res. 28*, 3853−3863.
11. King, D. A., Zhang, L., Guarente, L., and Marmorstein, R. (1999) Structure of a HAP1-DNA complex reveals dramatically asymmetric DNA binding by a homodimeric protein, *Nat. Struct. Biol. 6*, 64−71.
12. Patikoglou, G., and Burley, S. K. (1997) Eukaryotic transcription factor-DNA complexes, *Annu. Rev. Biophys. Biomol. Struct. 26*, 289−325.
13. Rosenberg, A. H., Lade, B. N., Chui, D., Lin, S.-W., Dunn, J. J., and Studier, F. W. (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase, *Gene 56*, 125−135.
14. Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins, *Biochemistry 6*, 1948−1954.
15. Hach, A., Hon, T., and Zhang, L. (2000) The coiled coil dimerization element of the yeast transcriptional activator Hap1, a Gal4 family member, is dispensable for DNA binding but differentially affects transcriptional activation, *J. Biol. Chem. 275*, 248−254.
16. Kissinger, C. R., Beishan, L., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8Å resolution: a framework for understanding homeodomain-DNA interactions, *Cell 63*, 579−590.
17. Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J., and Wüthrich, K. (1993) Determination of the nuclear magnetic resonance solution structure of an *Antennapedia* homeodomain: DNA complex, *J. Mol. Biol. 234*, 1084−1097.
18. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry Part 1: The conformation of biological molecules*, W. H. Freeman and Co., New York.
19. Zitomer, R. S., and Lowry, C. V. (1992) Regulation of gene expression by oxygen in *Saccharomyces cerevisiae*, *Microb. Rev. 50*, 1−11.
20. Schneider, J. C., and Guarente, L. (1991) Regulation of the yeast *CYT1* gene encoding cytochrome *c*1 by HAP1 and HAP2/3/4, *Mol. Cell. Biol. 11*, 4934−4942.
21. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51−55.

BI049419I