Structure of HAP1-PC7 bound to DNA: implications for DNA recognition and allosteric effects of DNA-binding on transcriptional activation

Amanda K. Lukens, Daniel A. King and Ronen Marmorstein*

The Wistar Institute and The Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, USA

Received July 14, 2000; Revised and Accepted August 28, 2000 NDB/PDB no. 1QP9

ABSTRACT

HAP1 is a transcription factor in yeast whose DNA-binding domain has been implicated in directly affecting transcriptional activation. Two separate mutations in the DNA-binding domain, S63G (HAP1-PC7) and S63R (HAP1-18), retain wild-type binding affinity. However, HAP1-PC7 is transcriptionally silent while HAP1-18 shows highly elevated levels of transcription. We have determined the X-ray crystal structure of the DNA-binding domain of HAP1-PC7 bound to its DNA target, UAS CYC7, and compared it to the previously solved HAP1-wt and HAP1-18 complexes to UASCYC7. Additionally, we have quantitatively compared the DNA-binding affinity and specificity of the HAP1-PC7, HAP1-18 and HAP1-wt DNA-binding domains. We show that, although the DNA-binding domains of these three proteins bind UAS CYC7 with comparable affinity and specificity, the protein–DNA interactions are dramatically different between the three complexes. Conserved protein–DNA interactions are largely restricted to an internal DNA sequence that excludes one of the two conserved DNA half-sites of UASCYC7 suggesting a mode of recognition distinct from other HAP1 family members. Alternative protein–DNA interactions result in divergent DNA configurations between the three complexes. These results suggest that the differential transcriptional activities of the HAP1, HAP1-18 and HAP1-PC7 proteins are due, at least in part, to alternative protein–DNA contacts, and implies that HAP1–DNA interactions have direct allosteric effects on transcriptional activation.

INTRODUCTION

The HAP1 protein is a member of a family of over 80 fungal transcription factors that contain an N-terminal Zn2Cys6 binuclear cluster domain (Fig. 1a) for DNA recognition and an extreme C-terminal transcriptional activation domain (1–3). The X-ray crystal structures of DNA complexes with the DNA-binding domains of several members of this family, including GAL4 (1), PPR1 (4) and PUT3 (5), reveal that these proteins bind as homodimers to DNA targets containing two CGG half-sites. In each protein–DNA complex, N-terminal Zn2Cys6 domains make nearly identical DNA half-site contacts, while C-terminal coiled-coil and linker protein domains make DNA interactions to accommodate the protein-specific inter-base pair separation between DNA half-sites. Specifically, GAL4, PPR1 and PUT3 each bind DNA targets containing 11, 6 and 10 inter-half-site base pairs, respectively (6). GAL4 accommodates a 10 bp inter-half-site separation by using its coiled-coil and linker region to make electrostatic interactions with the backbone of the intervening bases (1); PPR1 uses a coiled-coil and linker region to create a hydrophobic core that aligns the two protein subunits to accommodate a 6 bp inter-half-site separation (4); and PUT3 inserts a linker strand into the DNA minor groove to accommodate a 10 inter base-pair separation (5).

The DNA-binding and transcriptional activation properties of the HAP1 protein are unusual among the proteins containing a Zn2Cys6 domain. First, HAP1 has a relatively promiscuous DNA-binding specificity which can accommodate the divergent DNA half-sites, CGG (UAS1 of the CYC1 gene) (7), CGC (UAS of the CYC7 gene, UASCYC7) (7), AGG (UAS of the CYB2 gene) (8) and TGG (UAS of the CTT1 gene) (9) (Fig. 1b). Moreover, selected mutations in the DNA-binding domain of HAP1 and the HAP1 DNA target sequence have negligible effects on DNA-binding affinity, while having dramatic effects on transcriptional activation levels (10) (Fig. 1a and c). Of particular interest in this regard are mutations at serine 63 just N-terminal to the first cysteine of the Zn2Cys6 domain of HAP1 (11). A mutation to arginine (HAP1-18 protein) results in a dramatic increase in transcriptional activation from the UASCYC7
sequence by 10–100-fold, while mutation to glycine (HAP1-PC7 protein) results in a complete loss of transcriptional activation (10,11).

Table 1. DNA-binding and transcription properties of HAP1 proteins

<table>
<thead>
<tr>
<th>Protein</th>
<th>$K_D$ (UAS$_{CYC7}$) (nM)</th>
<th>Specific/non-specific</th>
<th>In vivo transcription activation$^b$ (β-gal units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAP1-wt</td>
<td>45.5 ± 3.6</td>
<td>634 ± 13</td>
<td>1</td>
</tr>
<tr>
<td>HAP1-PC7</td>
<td>37.8 ± 2.3</td>
<td>826 ± 35</td>
<td>0.023</td>
</tr>
<tr>
<td>HAP1-18</td>
<td>26.9 ± 2.4</td>
<td>706 ± 20</td>
<td>12.6</td>
</tr>
</tbody>
</table>

$^a$ A target sequence for the p53 protein (5'-GTGACATGCTGAGCATGCTC) was used as a non-specific site for the HAP1 proteins.

$^b$ In vivo transcription activation levels are taken from Turcotte and Guarente (10) and normalized to 1 for HAP1-wt.

In order to begin to address the unusual DNA-binding and transcriptional activation properties of the HAP1 protein, we have recently reported the structure of the DNA-binding domain of wild-type HAP1 and mutant HAP1-18 bound to UAS$_{CYC7}$ (12,13). The HAP1/UAS$_{CYC7}$ complex reveals a dramatically asymmetric HAP1 homodimer that makes base-specific DNA contacts to two CGC half-sites and to inter-half-site bases, suggesting that inter-half-site sequences may play an especially important role in DNA-binding specificity by HAP1 (12). Surprisingly, a comparison of UAS$_{CYC7}$ complexes with HAP1 and HAP1-18 reveals that while the proteins are essentially superimposable, the DNA targets show significant structural differences with dramatically rearranged protein–DNA interfaces (13).

To further extend these previous studies we now report the structure of HAP1-PC7 bound to UAS$_{CYC7}$, and we quantitatively compare the DNA-binding properties of the isolated DNA-binding domains of HAP1-PC7, HAP1-18 and HAP1-wt. We find that, as previously suggested for the full-length proteins, the DNA-binding domains of HAP1-PC7, HAP1-18 and HAP1-wt show comparable dissociation constants for the UAS$_{CYC7}$ sequence, and also show comparable sequence binding specificity. Remarkably, despite their similar DNA-binding properties, we find that the HAP1-PC7/UAS$_{CYC7}$ structure shows a dramatically different protein–DNA interface than either the HAP1-18/UAS$_{CYC7}$ or HAP1-wt/UAS structures. Conserved protein–DNA
contacts to the inter-half-site bases and to one of the two
conserved DNA half-sites suggesting a mode of DNA recognition
for HAP1 that is distinct from other members of the Zn$_2$Cys$_6$
in family protein. Moreover, taking our results together with the
divergent transcriptional properties of the HAP1-PC7, HAP1-18
and HAP1-wt proteins suggest that alternative HAP1–DNA
interactions may have direct allosteric effects on transcriptional
activation.

MATERIALS AND METHODS

Coordinates for the HAP1-PC7/UASCYC7 complex

The atomic coordinates for the HAP1-PC7/UASCYC7 complex
have been submitted to the Protein Data Bank under accession
number 1QP9.

Protein and oligonucleotide preparation

The DNA-binding domains (residues 55–135) of wild-type
HAP1 (HAP1-wt) and HAP1-18 (bearing a Ser→Arg mutation
at residue 63) were overexpressed in bacteria and purified to
homogeneity as previously described (12,13). The DNA
sequence encoding the DNA-binding domain of HAP1-PC7
(residues 55–135, bearing a Ser→Gly mutation at position 63)
was amplified by PCR from plasmid pHAP1 using a 5′ primer
in which the sequence was designed to encode a glycine at
position 63 of HAP1. The fragment encoding HAP1-PC7 (55–135)
was cloned into the PRSET-A vector and transformed in
Escherichia coli strain BL21LysS for protein over-expression.
The HAP1-PC7 protein was over-produced and purified as
described for HAP1-wt (55–135) (12). All proteins were
concentrated to ~40 mg/ml and frozen at ~70°C in small aliquots
prior to use in crystallization or protein–DNA binding studies.

All oligonucleotides for crystallization were synthesized at
the University of Pennsylvania DNA synthesis facilities and
purified by reverse phase HPLC using a Dynamax 300 column
(Rainin) as previously described (14). Prior to crystallization,
DNA duplexes were prepared by mixing equimolar amounts of
complimentary strands and annealed by heating to 75°C and
slowly cooling to 4°C.

5′-Fluorescein-labeled oligonucleotides for in vitro protein–DNA
binding studies were synthesized and purified at the University
of Pennsylvania DNA synthesis facilities. The UAS$_{UCYC}$ DNA
duplex used for the binding studies was prepared by annealing
as described above and contained the sequence 5′-fluorescein-
ACCCCTCGTTATACCTCGTTATT (bases in bold indicate
DNA half-sites) complexed to a complimentary strand not
containing a fluorescein label. A non-specific control DNA site
used in the in vitro binding studies (harboring an idealized
binding site for the p53 protein) contained the sequence 5′-GTGA-
GCATGCTCGAGCATGCTC.

In vitro HAP1–DNA binding studies

The dissociation constants of wild-type and mutant HAP1
DNA-binding domains for a UAS$_{UCYC}$ DNA duplex site were
determined using a fluorescence polarization assay employing
a DNA duplex with a 5′-fluorescein label on one strand of the
DNA duplex. All binding reactions were carried out in 100 μl
of binding buffer (BB) containing 20 mM HEPES, pH 7.5,
30 mM NaCl, 10 μM Zn(OAc)$_2$ and 10 μg/ml BSA (carried
over from the protein dilutions which were made in the
presence of 100 mg/ml BSA) in borosilicate glass cuvetts and
at a temperature of 20°C. Each reaction was carried out in the
presence of 1 nM DNA duplex and the respective protein was
titrated between 0.2 nM and 4 μM protein dimer in 2-fold
dilutions. Each binding reaction was allowed to equilibrate for
15 min prior to fluorescence measurements using a BEACON
200 Fluorescence Polarization System. Each experiment
was performed in triplicate and the millipolarization value (mP)
was plotted against a log scale of the protein concentration.
The data were fitted to a non-linear dose–response logistical
transition [y = a0 + a1/(1 + x/a2)a3] using the Levenberg–
Marquardt algorithm within the SlideWrite software package,
where the a2 coefficient is the dissociation constant (Kd).

To determine the specific to non-specific binding ratio for
each of the HAP1 proteins, unlabeled competitor DNA (a
24 bp idealized p53 binding site) was titrated into BB buffer
containing 0.5 nM fluorescein-labeled UAS$_{UCYC}$ DNA and
1 μM of the respective protein dimer. In the absence of
competitor DNA, this concentration of the respective HAP1
protein was saturating for UAS$_{UCYC}$ DNA binding as determined
by the respective millipolarization values. The competitor
DNA was titrated from 25 nM to 1.25 μM in 2-fold increments,
each titration was performed in triplicate. Each binding
reaction was allowed to equilibrate for 15 min prior to fluores-
cence measurements using a BEACON 200 Fluorescence Polarization
System. The mP value was plotted against a log scale of the
competitor DNA concentration and the data were fitted to a non-linear dose–response logistical transition. The
specific to non-specific binding ratio was calculated by
dividing the concentration of competitor DNA required to
displace one-half of the specific protein–DNA complex by
the amount of total specific DNA used in the titration (0.5 nM).

Crystallization and structure determination of the
HAP1-PC7/UASCYC7 complex

Crystals of HAP1-PC7 (55–135) bound to a consensus blunt-
ended 20 bp oligonucleotide were obtained by vapor diffusion
at 20°C, using 2 μl hanging drops containing 0.3 mM protein,
0.4 mM DNA duplex, 25 mM MES (2-morpholinoethane-
sulfonic acid) pH 5.6, 5.5% PEG 400 (polyethylene glycol
average molecular mass = 400 kDa), 5 mM MgSO$_4$
and 100 mM KCl equilibrated over a reservoir containing twice
the concentration of buffer, precipitating reagent and salts. Crystals
grew in several days to a typical size of 0.4 × 0.2 × 0.1 mm$^2$
in the space group P2$_1$2$_1$2$_1$, with a = 85.9, b = 90.9 and c = 96.5 Å,
and contain two protein–DNA complexes per asymmetric unit
(Table 2).

Data from crystals of the HAP1-PC7/UASCYC7 complex
were collected at 110 K at the A1 beamline of the Cornell High
Energy Synchrotron Source (CHESS) using a charged couple
device (CCD) detector and the data were reduced and pro-
cessed with DENZO and SCALEPACK (15). The structure
was solved to 3.0 Å by molecular replacement using the coordinates
of the HAP1-18(S63R)–DNA complex as a search model. In the
absence of residue 63 of HAP1-PC7, the rotation and translation
functions using the program AMORE (16) yielded two solutions
that were related by non-crystallographic symmetry (NCS). The
two NCS-related complexes were refined with rigid body
movements using the program X-PLOR (17). Prior to further
refinement, a randomly selected 10% of the data was set aside
for the free data set while the remaining data were used for refinement. The starting model was subjected to one round of positional refinement using strict NCS restraints at 3.0 Å with subsequent model building adjustments into electron density maps generated with Fourier coefficients $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ using the program O (18). Electron density maps showed strong density for all regions except for the DNA of the B-molecule. Due to the poor density of the B-DNA, the entire B-DNA model was removed prior to further refinement using the maximum likelihood function in the program CNS (19). Within the CNS program, an anisotropic B-factor refinement was applied to the reflection file from 8 to 3 Å along with a bulk solvent correction. After one round of positional refinement followed by grouped B-factor refinement, maps generated with Fourier coefficients $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ at 3.0 Å showed density for the omitted DNA of the B-molecule. All nucleotides were built into a $1\sigma$ difference map with the exception of three nucleotides from one strand which showed no clear density. Iterative refinement using positional, simulated annealing (20) or torsion angle dynamics (21) protocols as implemented in CNS, followed by model building in O in resolution steps extending to 2.8 Å was carried out. At the later stages of refinement, NCS restraints were gradually released, individual B-factors were adjusted and water molecules were built into regions that showed strong F_o–F_c peaks that made stereochemically feasible hydrogen bonds. The correctness of the model was checked against simulated annealing omit maps (22) over the entire structure by omitting 10 residues at a time, and the model was adjusted appropriately. A Ramachandran plot shows no residues in disallowed regions. A final round of refinement at 2.8 Å resolution resulted in a model with good refinement statistics and geometry (Table 2). The current model still has many regions of the B-molecule’s DNA with high B-factors and poor $2|F_o| - |F_c|$, $|F_o| - |F_c|$ and omit maps. For this reason, the DNA target of the B-molecule is considered disordered and has not been analyzed in this study.

### RESULTS AND DISCUSSION

**DNA-binding properties of HAP1-PC7, HAP1-wt and HAP1-18**

Using an electrophoretic mobility shift assay (EMSA), Turcotte and Guarente reported that the HAP1-PC7 and HAP1-18 proteins bound to a UAS_CYC7 DNA target with dissociation constants comparable to the wild-type HAP1 protein (HAP1-wt) (10). These studies were carried out in the context of the full-length HAP1 protein and showed that transcriptional activation was strictly dependent on the C-terminal

---

**Table 2. Crystallographic data for the HAP1-PC7(55–135)/UAS_CYC7 complex**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit cell dimensions</td>
<td>(a = 85.9), (b = 90.8), (c = 96.5) Å, (\alpha = \beta = \gamma = 90^\circ)</td>
</tr>
<tr>
<td>Space group</td>
<td>(P2_12_12_1)</td>
</tr>
<tr>
<td>Resolution</td>
<td>2.8 Å</td>
</tr>
<tr>
<td>Measured reflections</td>
<td>27,006</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>14,650</td>
</tr>
<tr>
<td>Completeness</td>
<td>99.6%</td>
</tr>
<tr>
<td>Rsym</td>
<td>7.8%</td>
</tr>
<tr>
<td><strong>Refinement parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Resolution range</td>
<td>8.0–2.5 Å</td>
</tr>
<tr>
<td>Asymmetric unit</td>
<td>two HAP1-PC7 dimer/DNA Mr = 38,619</td>
</tr>
<tr>
<td>Final model</td>
<td></td>
</tr>
<tr>
<td>Number of non-hydrogen atoms</td>
<td>4,119</td>
</tr>
<tr>
<td>Number of water molecules</td>
<td>96</td>
</tr>
<tr>
<td>Number of zinc atoms</td>
<td>9</td>
</tr>
<tr>
<td>Average B-factors (Å²)</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>33.690</td>
</tr>
<tr>
<td>DNA (A molecule)</td>
<td>42.220</td>
</tr>
<tr>
<td>Water molecules</td>
<td>37.360</td>
</tr>
<tr>
<td>R-factor (16,303)</td>
<td>25.500</td>
</tr>
<tr>
<td>R_m. (10%)</td>
<td>29.400</td>
</tr>
<tr>
<td>R.m.s. deviation bond lengths (Å)</td>
<td>0.009</td>
</tr>
<tr>
<td>R.m.s. deviation bond angles (°)</td>
<td>1.529</td>
</tr>
<tr>
<td>R.m.s. deviation between NCS (Cα)</td>
<td>0.583</td>
</tr>
</tbody>
</table>

\(^a\)R-factor = \(\Sigma_{hk} |F_{P_{H}}| ± |F_{P_{C}}| - \Sigma_{hk} |F_{P_{H}}| ± |F_{P_{C}}| / \Sigma_{hk} |F_{P_{H}}| ± |F_{P_{C}}|\), for centric reflections \(\Sigma_{hk} |F_{O_{H}}| ± |F_{O_{C}}| - \Sigma_{hk} |F_{O_{H}}| ± |F_{O_{C}}| / \Sigma_{hk} |F_{O_{H}}| ± |F_{O_{C}}|\).
transcriptional activation domain of HAP1 (residues 1307–1483). To better quantitate and compare the DNA-binding affinity and specificity of the HAP1-PC7, HAP1-18 and HAP1-wt DNA-binding domains we carried out fluorescence polarization studies. As shown in Figure 2a and quantitated in Table 1, each of the HAP1 DNA-binding domains bind the UAS_CYC7 DNA target with comparable affinities of 45.5, 37.8 and 26.9 nM for the HAP1-wt, HAP1-PC7 and HAP1-18 proteins, respectively. Moreover, using a DNA target for the p53 protein as a competitive inhibitor reveals that a comparable amount of p53 binding site was required to displace each of the three HAP1 complexes, illustrating that each of the HAP1 proteins had similar DNA-binding specificities (Fig. 2b and Table 1). These results are consistent with the previous studies using larger constructs of the HAP1-wt, HAP1-PC7 and HAP1-18 proteins, which showed that the three proteins bind the UAS_CYC7 DNA sequence with comparable affinities (10,11). The observation that HAP1-wt constructs that extend beyond the minimal 81 amino acid DNA-binding domain (residues 55–135) bind DNA with greater affinity than the minimal DNA binding domain (23,24) suggests that regions outside the DNA-binding domain also influences overall DNA-binding affinity. Nonetheless, our data taken together with the previous data, suggest that the DNA-binding domains of HAP1-wt, HAP1-PC7 and HAP1-18, like the full-length proteins, bind the UAS_CYC7 DNA sequence with comparable affinity and specificity, and that regions outside the DNA-binding domain have similar influences on the overall DNA-binding properties of these three proteins.

Overall structure of the HAP1-PC7–DNA complex

The HAP1-PC7 DNA-binding domain and the UAS_CYC7 DNA target used in the structure determination are shown in Figure 1a and b (along with their numbering schemes). Since the HAP1 dimer is bound asymmetrically to the DNA site each protein subunit (with its associated DNA half-site) is designated ‘A’ and ‘B’, respectively. The complex crystallizes in space group P2_1_2_1_2, with two complexes per asymmetric unit (Table 2). The DNA of one of the two protein–DNA complexes of the unit cell is principally disordered with relatively high atomic B-factors and is therefore not analyzed here (see Materials and Methods). A schematic of the structure is shown in Figure 1c. Like the wild-type HAP1/UAS_CYC7 complex (12), the HAP1-PC7 DNA-binding domain is bound as an asymmetric dimer to the DNA site aligning each Zn_2Cys_6 domain (residues 64–93) with one of the two CGC half-sites in the major groove of DNA. Each Zn_2Cys_6 domain is connected to a short symmetrical coiled-coil (residues 114–128) via an asymmetrically disposed linker region (residues 94–113). Sequences directly N-terminal to the Zn_2Cys_6 domains form extended strands (residues 55–63) that sit over the minor groove to the right of each Zn_2Cys_6 domain. Protein segments that stabilize the asymmetric disposition of the dimer on DNA include the linker and N-terminal arm of protein subunit A and the Zn_2Cys_6 domain of protein subunit B.

Comparison of DNA complexes with HAP1-PC7, HAP1 and HAP1-18

A comparison of the DNA-binding domain of the HAP1-PC7 mutant with wild-type HAP1 and HAP1-18 reveals nearly identical conformations for the proteins with a r.m.s. deviation between Cα atoms of 0.590 and 0.667 Å, respectively (Fig. 3a). In contrast to the structural conservation between the proteins, the respective bound DNA targets show significant localized differences (Fig. 3b–d). These localized DNA differences can be grouped into three regions: a left side containing base pairs 0A–6A and harboring the A DNA half-site, a central ‘core DNA’ region containing base pairs 7A–3B and harboring inter-half-site sequences and the B DNA half-site, and a right side containing base pairs 4B–10B and harboring sequences directly flanking the B DNA half-site (Fig. 1b). Comparison of the right DNA segment reveals a large degree of inherent
Figure 3. Comparison of UAS\textsubscript{CYC7} complexes with HAP1, HAP1-18 and HAP1-PC7. (a) Alignment of the complexes showing the entire UAS\textsubscript{CYC7} DNA target and the Cα trace of the protein dimers. The HAP1, HAP1-18 and HAP1-PC7 complexes are shown in blue (dark and light shade for the two complexes in the asymmetric unit cell), red and yellow, respectively. The location of residue 63 is indicated by a green ball on each protein subunit. The protein subunits are numbered according to the discrete domains labeled in Figure 1 and the half-sites are highlighted in gray. Junction points for the three segments of the DNA (left, central and right) are indicated. (b) Alignment and color scheme as in (a) comparing the UAS\textsubscript{CYC7} DNA structure of HAP1-PC7 and both crystallographically unrelated UAS\textsubscript{CYC7} structures of HAP1. (c) Alignment and color scheme as in (a) comparing the UAS\textsubscript{CYC7} DNA structure of HAP1-18 and both crystallographically unrelated UAS\textsubscript{CYC7} structures of HAP1. (d) Alignment and color scheme as in (a) comparing the UAS\textsubscript{CYC7} DNA structures of HAP1-PC7 and HAP1-18.
flexibility. This is supported by the observation that the two DNA duplexes in the unit cell of the HAP1/UASCYC7 complex show structural divergence which also appears to extend to the HAP1-18 and HAP1-PC7 DNA complexes (12) (Fig. 3). In contrast, the central core DNA segment (bases 7A–3B) superimpose remarkably well between all three complexes. The left segment of the DNA harbors structural differences between the three complexes that appears to be a function of the protein bound. This is supported by the observation that the two HAP1/UASCYC complexes in the unit cell have nearly identical structure in this region of the DNA (12).

The majority of conserved interactions between the three protein–DNA complexes map to the central core DNA segment (Fig. 4a and b). These interactions co-localize to protein segments of HAP1-PC7 that form the asymmetric dimerization interface: the linker and N-terminal arm of the A subunit, and the Zn2Cys6 domain of the B subunit. Specifically, Arg57A and Asn58A from the A N-terminal arm contact Thy8A (O2 and O4') and Cyt1B (Pi), respectively. The Nε nitrogen of Trp100A of the A subunit’s linker and the guanidinium group of Arg70B of the B Zn2Cys6 domain donate hydrogen bonds to the phosphate of Thy9A while making Van der Waals interactions with each other at the asymmetric dimerization interface between subunits. Within the Zn2Cys6 domain of subunit B, the carbonyl of Arg70B accepts a hydrogen bond from the N4 nitrogen of Cyt1B, and Lys71B makes Van der Waals interactions with the C5 and C6 positions of Cyt1B and the C5 position of Gua3'B of the B DNA half-site.

In comparing HAP1-PC7 to HAP1-18, the most striking differences are the dramatically rearranged protein–DNA contacts nucleated by a single amino acid change from serine to glycine (HAP1-PC7) or to arginine (HAP1-18) (Fig. 4). This is especially surprising since the three proteins show comparable DNA-binding affinity and specificity (Fig. 2 and Table 1). Strikingly, the structural differences are largely mediated by the DNA sequences flanking the structurally conserved central DNA core region (bases 7A–3B). Some of the most dramatic differences are seen at the left DNA segment which contains the A DNA half-site (bases 0A–6A) (Fig. 4c). Only weak, limited interactions are seen between the A Zn2Cys6 binuclear cluster and the A CGC half-site in the HAP1-PC7 complex (Fig. 3c). Specifically, only Van der Waals interactions are observed between Lys71A with Cyt1A and Val72A and Lys73A with Gua3'A. This is in contrast to the multiple interactions found in the wild-type and HAP1-18 DNA complexes (Fig. 4b and c). In addition to these conserved Van der Waals interactions between residues Lys71A and Lys73A and bases Cyt1A and Gua3'A, respectively, the wild-type and HAP1-18 structures show supplemental hydrogen bonding interactions. In the wild-type complex, the Nε atom of Lys71A makes three interactions in the major groove to the N7 nitrogen and O6 oxygen of Gua2A and the O6 oxygen of Gua3'A. In the HAP1-18 complex, main chain carbonyls of Arg70A and Lys71A accept hydrogen bonds from the N4 nitrogens of cytosines 1A and 2'A, respectively. Additionally, Arg63A of HAP1-18 donates a hydrogen bond to the N7 nitrogen of Gua3'A of the A DNA half-site. Amazingly, no hydrogen bonds at the A half-site are found in common between the three complexes. Interestingly, this region of DNA colocalizes to the part of the DNA that appears to adopt a variable structure that is dependent on the protein species bound (Fig. 2).

Another region of the three protein–DNA complexes that shows highly divergent protein–DNA interactions maps to DNA sequences of the right DNA segment (bases 5B–10B) (Fig. 4d). As noted above, this region of the DNA shows the greatest degree of inherent flexibility between the three complexes. These variable interactions are almost exclusively mediated by the N-terminal arm region of the B protein subunit. In the wild-type HAP1 complex, Arg59B contacts the phosphate oxygen of Ade6B (Fig. 4d), and in the HAP1-18 complex this arginine flips by nearly 180° to make a hydrogen bond to the phosphate oxygen of Ade6'B (Fig. 4d). In the HAP1-PC7 complex, Arg59B points away from the DNA and, instead, residues 55–57 make extensive base and backbone interactions through bases 5B–10B.

**Implications for the mode of DNA recognition by HAP1**

In light of the similar DNA-binding properties of the HAP1-wt, HAP1-PC7 and HAP1-18 proteins, a comparison of the DNA complexes with HAP1, HAP1-18 and HAP1-PC7 have implications for the mode of DNA-binding by HAP1. Specifically, the extent of conserved contacts to the central core DNA segment, which includes the B DNA half-site, demonstrates that this region of the DNA is critical for DNA recognition by HAP1. Consistent with the importance of this region of the DNA is the relatively high degree of structural superposition of this region in the three protein–DNA complexes (Fig. 3), its high degree of sequence conservation between known HAP1 binding sites (Fig. 1b), and its mutational sensitivity (25, 26). The asymmetric dimerization interface of HAP1, which colocalizes to the central DNA core region, appears to play a particularly important role in DNA recognition (Fig. 3a). In particular, the regions of HAP1 that stabilize the dimer, the linker and N-terminal arm of the A protein subunit, and the Zn2Cys6 domain of the B protein subunit, mediate conserved DNA interactions between the three protein–DNA complexes. Moreover, crosslinking, gel filtration studies and various protein chimeras suggest that HAP1 exists primarily as a monomer in solution and that cooperative interactions between the DNA and protein promotes HAP1 dimerization on DNA (12, 27). Together, the asymmetric dimerization region of HAP1 and the central core DNA segment play a prominent and interrelated role on DNA recognition by HAP1.

The relative lack of conservation of HAP1–DNA contacts to the A DNA half-site suggests that this half-site does not play a major role in DNA recognition by HAP1. This is supported by the observation that HAP1 binds DNA targets in vivo that contain a divergent A DNA half-site (Fig. 1b). For example, while the UASCYC target used in this study contains two CGC half-sites, the UAS1 sequence contains two CGG half-sites (found in most members of the Zn2Cys6 family), and the UAS sequences of CYB2 and CT1 contain AAG and TGG half-sites (8, 9), respectively, on the A DNA half-site. Interestingly, all but the UASCYC target has a CGG half-site as the B half-site, correlating with the relative importance of this half-site in the central region (bases 7A–3B) for DNA recognition by HAP1. The reduced importance of the two DNA half-sites in HAP1 recognition of UASCYC is in contrast to the mode in which the other proteins containing a Zn2Cys6 domain, such as GAL4, PPR1 and PUT3 recognize DNA. In the DNA complexes with
each of these proteins, the symmetrically disposed Zn2Cys6 domains of the respective protein dimers make symmetrical DNA contacts with two CGG DNA half-sites (1,4,5). Also in these cases, the DNA half-sites are the most mutationally sensitive positions of the DNA target, whereby the identity of the inter half-site bases play a relatively minor role in DNA-binding specificity in these proteins (5,28). As noted previously, UASCYC7 includes two CGC half-sites, unlike the optimal CGG half-sites found for many Zn2Cys6 binuclear cluster family members, including many other HAP1 targets sites (Fig. 1b). Therefore, the reduced importance of the DNA half-sites of UASCYC7 for HAP1 binding may reflect the presence of the non-optimal CGC half-sites. Nonetheless, the extensive and specific interface found between the protein and DNA in each of the three HAP1/DNA complexes compared here indicates an inherent versatility in HAP1 binding relative to other GAL4 family members. This versatility may be able to compensate for mutations at the half-sites when bases separating the half-sites are favorable for binding. Whether HAP1 would make DNA interactions that are more typical to other proteins of the Zn2Cys6 binuclear cluster family if the half-sites contained a CGG sequence (such as in UAS1) requires a structure determination of HAP1 bound to DNA containing CGG half-sites. Nonetheless, our data illustrate how the conserved Zn2Cys6 domain of HAP1 can be used to recognize DNA by a mechanism that is quite distinct from its structurally related proteins.

Implications for the direct involvement of DNA binding by HAP1 on transcription

Mutation of position 63 of the HAP1 protein, just N-terminal to the first cysteine of the Zn2Cys6 domain, has dramatic effects on the transcriptional activation properties of HAP1 (10,11). Two of the most unusual phenotypes are exhibited by the HAP1-18 (Ser→Arg substitution) and HAP1-PC7 (Ser→Gly substitution) proteins. While both proteins bind the UASCYC7 site with near wild-type affinities, HAP1-18 shows elevated levels of transcriptional activation (10–100-fold), while HAP1-PC7 is transcriptionally silent. In addition, several other HAP1 mutations have been identified (K54E, R55G, P61Q, H94D, Q98L, A101E, T127I, L128R) (10) (Fig. 1a) that have minor effects on DNA binding and relatively dramatic effects on transcriptional activation. A mapping of each of these mutations onto the HAP1-PC7/DNA structure reveal that
they are not localized to a protein surface, and all but two of them map to residues of the HAP1 protein that appear to mediate, either directly or indirectly, DNA interactions (Fig. 1c). This observation led us to previously postulate that protein–DNA interactions may have direct allosteric effects on transcriptional regulation (12). This hypothesis was further supported by the structure of a HAP1-18/UASCYC7 complex which revealed a dramatically rearranged protein–DNA interface relative to the HAP1/UASCYC7 complex (13).

This conclusion has been further supported and extended by analysis of the HAP1-PC7/UASCYC7 complex reported here. Specifically, the HAP1-PC7/DNA complex reveals protein–DNA contacts that are significantly rearranged relative to both the HAP1– and HAP1-18–DNA complexes (Fig. 4). Significantly, while many contacts to the central region of the DNA target (bases 7A–3B) are conserved between the three DNA complexes, contacts to the bases flanking this central region and including the A DNA half-site are highly divergent. These structural correlations suggest that while the internal sequences of the UASCYC7 site (central DNA segment) play an important role in both sequence binding specificity and transcriptional activation, protein contacts to DNA bases directly flanking this internal sequence (bases 0A–6A of the left DNA segment and including the A DNA half-site and bases 4B–10B of the right DNA segment) have additional effects on the transcriptional regulatory properties of HAP1.

These structural correlations agree well with mutagenesis studies of the HAP1 DNA targets, UASCYC7 and UAS1 (25). Specifically, selected mutations in the DNA sequences of the UASCYC7 and UAS1 sites have relatively minor effects on HAP1 binding in vitro, but have dramatic effects on transcriptional activation in vivo. For example, a Cyt→Gua change at position 3B (at the edge of the central core DNA region) of the UASCYC7 site is near wild-type for DNA binding but is up about 6-fold for transcriptional activation. In addition, multiple mutations including positions Cyt1A and Thy4B (outside the core DNA region), and positions Cyt13A and Thy7B (outside the core DNA region) are down in binding by only about 3-fold but have nearly background levels of transcriptional activation. Significantly, other mutations in the same positions that result
in even more reduced DNA-binding affinity have transcriptional activation levels that are significantly above background levels. Moreover, multiple mutations within another HAP1 target, the UAS1 site, that include positions 4B and 6B (outside the core DNA region of the corresponding UASCYC7 site) bind the UAS1 site at levels slightly better than wild-type, but with reduced transcriptional activation levels (25).

Our structural results also have implications for HAP1-mediated transcriptional activation through the UAS1 site of the CYC1 gene. While the wild-type HAP1 protein binds both UASCYC7 and UAS1 sequences with equal affinity, HAP1 activates transcription 10–100-fold higher at UAS1 relative to UASCYC7, regardless of promoter context (7). The relative lack of sequence homology between the two sites (Fig. 1b), combined with the structural observation that relatively subtle protein substitutions, and presumably also DNA substitutions, can result in dramatically altered protein–DNA interfaces, implies that the activation differences between the two sites can be attributed to alternative hydrogen bonding patterns at UAS1 and UASCYC7. Biochemical data also support this argument. Methylation protection assays of wild-type HAP1 bound to both targets show different protection patterns (7). Additionally, bases contacted differentially at UASCYC7 by the wild-type and two mutant proteins are not conserved at UAS1. Most strikingly, two CGG half-sites are found at UAS1 relative to the CGC half-sites found at UASCYC7. These half-sites have been implicated to have strong effects on the activity on HAP1 bound to the two targets (25). Specifically, mutating either CGC half-site at UASCYC7 to the UAS1-like CGG half-site dramatically raises transcriptional activity without changing the DNA-binding properties of wild-type HAP1. Additionally, these same mutations revert the phenotypes of HAP1-18 and other mutants that specifically effect transcriptional activation, to those mediated by the UAS1 sequence. Moreover, mutating the TGG and AGG half-sites of CYB2 and CTT1 to CGG dramatically raises the transcriptional activity of HAP1 at these DNA targets.

It is tempting to speculate that the A DNA half-site of the HAP1 target plays an especially important role in transcriptional activation since its conformation appears to be dependent on the protein species to which it is bound and mutations of this half-site have significant effects on transcriptional activation levels (25). However, further structural studies of other HAP1-PC mutants would be required to support this hypothesis. Although we observe a correlation between the levels of transcriptional activation by HAP1-PC mutants and altered protein–DNA contacts by these mutants, the precise mechanism by which HAP1–DNA interactions impart effects on transcriptional activation is not known. A possible scenario would involve specific recognition of DNA structural properties or recognition of a protein–DNA interface by a putative cofactor protein for transcriptional activation. Whether this cofactor functions through bridging communication between HAP1 and the basal transcription machinery or through modifying histone–DNA interactions awaits further biochemical and structural analysis. Nonetheless, the studies presented here suggest that HAP1–DNA interactions have direct allosteric effects on transcriptional activation which involves minor protein rearrangements of the DNA-binding domain but relatively significant DNA structural adjustments.

CONCLUSION

The HAP1, HAP1-18 and HAP1-PC7 homodimers each bind the UASCYC7 site with comparable DNA affinities and specificities, yet display distinct transcriptional activation properties from this site. HAP1-18 mediates transcriptional activation 10–100-fold more strongly than HAP1 while HAP1-PC7 is transcriptionally silent. A comparison of the structures of these proteins bound to the same UASCYC7 DNA reveals that conserved DNA contacts are largely restricted to a central DNA sequence harboring one of two conserved DNA half-sites and are made by regions of the protein that are associated with promoting the dramatically asymmetric conformation of the HAP1 homodimer. These results suggest that HAP1 binds DNA by a novel mechanism by which protein dimerization and DNA recognition are interrelated. Moreover, the mechanism for HAP1 binding UASCYC7 is distinct from other members of the ZnCys family, such as GAL4, PPR1 and PUT3, which mediate base-specific contacts via symmetrical ZnCys domain interactions with the two conserved DNA half-sites (1,4,5). This analysis highlights the versatile nature by which the ZnCys domain can be used for DNA recognition. It will be interesting to determine the structure of the ZnCys domain proteins LEU3 and PDR3 bound to DNA, since they are believed to bind DNA by yet another mechanism (29).

Our comparison of the UASCYC7 complexes with HAP1, HAP1-18 and HAP1-PC7 reveal a dramatically rearranged protein–DNA interface outside a central core DNA sequence leading us to conclude that protein–DNA interactions have direct allosteric effects on transcriptional activation. The unusual transcriptional activation properties of the HAP1 DNA-binding domain may share functional and structural properties with other transcriptional activator proteins. For example, the DNA-binding domains of USF2 (30) and the glucocorticoid (31,32) and retinoid receptors (33,34) have been shown to play an active role in transcriptional activation. Moreover, a model whereby DNA can act as an allosteric effector of transcriptional activation has been proposed to rationalize the phenotype of selected mutations within the DNA-binding domains of the glucocorticoid steroid receptor (31). Indeed, the DNA-binding and transcriptional activation functions of transcriptional regulators may be more interrelated then previously appreciated.

ACKNOWLEDGEMENTS

The authors wish to thank Adrienne Clements and Yi Mo for useful discussions. This work was supported by awards from the NIH (GM52880, S10 RR 12823) and the Fannie E. Rippel Foundation to R.M., and a Monica Shander Fellowship Award (Wistar Institute) to D.K.

REFERENCES