Structure
Article

Structural Basis for Dimerization in DNA Recognition by Gal4

Manqing Hong,[1,2] Mary X. Fitzgerald,[1,2,3] Sandy Harper,[1] Cheng Luo,[1] David W. Speicher,[1] and Ronen Marmorstein[1,2,*]
[1]The Wistar Institute
[2]Department of Chemistry
University of Pennsylvania, Philadelphia, PA 19104, USA
[3]Present Address: Center for Advanced Biotechnology and Medicine, 679 Hoes Lane, Room 020, Piscataway, NJ 08854, USA
*Correspondence: marmor@wistar.org
DOI 10.1016/j.str.2008.03.015

SUMMARY

Gal4 is a Zn_2Cys_6 binuclear cluster containing transcription factor that binds DNA as a homodimer and can activate transcription by interacting with the mutant Gal11P protein. Although structures have been reported of the Gal4 dimerization domain and the binuclear cluster domain bound to DNA as a dimer, the structure of the "complete" Gal4 dimer bound to DNA has not previously been described. Here we report the structure of a complete Gal4 dimer bound to DNA and additional biochemical studies to address the molecular basis for Gal4 dimerization in DNA binding. We find that Gal4 dimerization on DNA is mediated by an intertwined helical bundle that deviates significantly from the solution NMR structure of the free dimerization domain. Associated biochemical studies show that the dimerization domain of Gal4 is important for DNA binding and protein thermostability. We also map the interaction surface of the Gal4 dimerization domain with Gal11P.

INTRODUCTION

The Gal4 DNA binding transcriptional activator is a yeast protein required for the metabolism of galactose and melibiose (Hopper et al., 1978; Kew and Douglas, 1976; Platt, 1984). Gal4 recognizes a 17 base pair consensus sequence, and there are multiple consensus sites with very similar sequence located within the upstream activating sequence of Gal4-regulated genes, such as Gal7 and Gal10. The Gal4 protein is 881 amino acids long, and distinct functional domains have been assigned (Johnston, 1987; Keegan et al., 1986; Ma and Ptashne, 1987a), including a Zn_2Cys_6 binuclear cluster DNA recognition element (residues 7–40) and a dimerization domain (residues 50–94) (Carey et al., 1989; Himmelfarb et al., 1990; Keegan et al., 1986; Liang et al., 1996; Ma and Ptashne, 1987b) that together with a nine residue linker that connects the binuclear cluster and dimerization region comprise the DNA binding domain. The Gal4 protein also contains two transcriptional activation domains (from residues 148–196 and 768–881). Interestingly, the Gal4 dimerization domain manifests another novel transcriptional activation capability in yeast cells carrying a single point mutation in the Gal11 protein (called Gal11P for Gal11 transcriptional potentiator), a component of the RNA-polymerase II holoenzyme (Barberis et al., 1995; Farrell et al., 1996; Gaudreau et al., 1998; Himmelfarb et al., 1990).

Several aspects of Gal4 binding to DNA, dimerization, and Gal11P interaction have been elucidated through structural and biochemical analysis. An NMR solution structure of a DNA binding region of Gal4 (residues 1–65) and a cocrystal structure of the same DNA binding region in complex with a 17bp consensus DNA sequence have been determined (Baleja et al., 1992; Marmorstein et al., 1992). The NMR structure shows that residues 7–40 of free Gal4(1–65) forms a globular Zn_2Cys_6 binuclear cluster domain, while residues 41–65 are unstructured. The crystal structure of a Gal4(1–65)/DNA complex shows a superimposable binuclear cluster domain, but the rest of the protein sequence is ordered with residues 41–50 and 50–65 forming an ordered, extended loop region and a coiled-coil dimerization element, respectively. Gal4(1–65) binds DNA as a homodimer with the binuclear cluster domains contacting CGG DNA half-sites that are separated by 11 base pairs and the coiled-coil region sitting over the center of the DNA, with its N-terminal dipoles sitting over the DNA minor groove. Together with the ordered loop region residues 41–50 that track the DNA minor groove, the coiled-coil specifies the 11 base pair separation between the CGG DNA half-sites. A comparison of the free Gal4(1–65) with the DNA-bound complex clearly shows that the dimerization element and extended loop region adopt structure upon DNA binding. The solution structure of a complete dimerization domain of Gal4(residues 50–106) was also determined by solution NMR, revealing an extended coiled-coil consisting of three α helices with flexible loops connecting them. In related biochemical studies, Gal11P binding was mapped to the C-terminal helix and a loop between helices 1 and 2 (Hidalgo et al., 2001).

To address the role of the complete Gal4 dimerization domain in DNA binding and Gal11P interaction in the context of a Gal4/DNA complex, we report here the structure of a complete Gal4 dimer bound to DNA and associated biochemical studies. We find that Gal4 dimerization on DNA is mediated by an intertwined helical bundle that deviates significantly from the structure of the dimerization domain off DNA. Together with associated biochemical studies, we show that the dimerization domain of Gal4 is important for DNA binding and protein stability, and map the interaction surface for Gal11P interaction on DNA. We also discuss possible dimerization domain reorganization for DNA binding.

RESULTS

Overall Structure of the Gal4-DNA Complex

The Gal4 protein construct crystallized here contains residues 1–100, which includes the Zn_2Cys_6 binuclear cluster through the complete dimerization domain (residues 50–94). The crystals form in the space group C2, and the asymmetric unit cell contains a complete Gal4 dimer bound to a 20 base pair DNA duplex (Table 1). The two protein subunits of the Gal4(1–100) dimer are nearly, but not perfectly, symmetrical with a root-mean-square deviation (rmsd) of 1.570 Å over all protein atoms, excluding the second helix of the dimerization domain. The analogous regions of the complex also superimpose well with the Gal4-(1–65)/DNA complex with an rmsd between all protein and DNA atoms of 1.4 Å and 1.2 Å, respectively (Figure 1A). The protein-DNA contacts between these protein/DNA complexes are also analogous, with the only exception being the absence of a hydrogen bond between Gln9 and the phosphate backbone in the Gal4(1–100)/DNA complex. The complete Gal4 dimerization domain seen with Gal4(1–100) forms a considerably more extensive dimer interface than Gal4(1–65), with a total solvent-excluded surface of 3052 $Å^2$, as opposed to 1452 Å, respectively.

The dimerization domain of each Gal4(1–100) monomer contains three α helices. The first helix (residues 51–71) forms a typical coiled-coil, and the second and third helices (residues 74–82 and 86–95 of subunit 1, respectively) fold back in an antiparallel fashion to form a helical bundle interaction near the C-terminal end of the first helix of the opposing subunit (Figures 1B and 1C). Interestingly, the dimerization domain of Gal4(1–100) bound to DNA shows considerable divergence from the more elongated DNA-free dimerization domain as determined by solution NMR (Hidalgo et al., 2001; Figure 1D). This will be discussed in more detail below.

Several apolar residues participate in Gal4(1–100) dimer interactions on DNA. This includes L67 and F68 of helix α1; I71 and F72 of loop1; L77, I80, and L81 of helix α2; and I89 and L93 of helix α3. Among these residues, L67, I71, I80, L81, and L93 show high sequence conservation among fungal Gal4 homologs. Residue L67 makes van der Waals contacts with the side chains of I80, I89, L92, and L93 of the symmetry-related Gal4 subunit of the dimer. In addition, residues I71 and F72 make van der Waals interactions with the symmetry-related Gal4 side chains of I71, F72, and L93. Residues I80 and L81 of helix α2 make Van der Waals contacts with the side chains of R60, R63, L64, and L67 on the opposing subunit, and I89 is in contact with L70. Together, these interactions create a hydrophobic core in the middle of the helical bundle. This hydrophobic core greatly stabilizes the Gal4(1–100) dimer interface (Figures 2A–2D). Beyond the hydrophobic core, there are two hydrogen bonds made by the side chain of R63 with the backbone oxygen atoms of L81 and M83. These hydrogen bonds may help position the α2 helix for more optimal van der Waals interactions.

Role of Gal4 Dimerization on DNA Binding

To directly investigate the role of Gal4 dimerization in DNA binding, we evaluated the relative DNA binding properties of wild-type Gal4(1–100) and selected Gal4 mutants with an electrophoretic mobility shift assay (EMSA). As a control, we first confirmed earlier studies showing that Gal4(1–100) harboring a complete dimerization domain binds DNA much more avidly than Gal4(1–65) harboring only a segment of helix 1 of the dimerization element (Carey et al., 1989; Keegan et al., 1986; Ma and Ptashne, 1987a). As shown in Figures 3A and 3B, Gal4(1–100) binds DNA with an apparent dissociation constant of 24 nM, while Gal4(1–65) binds DNA with an apparent dissociation constant of greater than 400 nM. This result confirms that the Gal4 dimerization domain plays an important role in DNA binding by Gal4.

We next analyzed the DNA binding properties of several alanine substitution mutants within the dimerization interface. In particular, we targeted residues 67, 71, 80, 81, 89, and 93, which appear to play important roles in dimerization. We initially prepared single site alanine mutations, and found that these mutations either had no effect or very small effects on Gal4(1–100) binding to DNA (data not shown). We then prepared several double and triple mutations, including L67A/I71A, L67A/I80A/L81A, L67A/I89A, and L67A/L93A for EMSA. As shown in Figures 3B

Table 1. Crystallographic Data Statistics for the Gal4(1–100)/DNA Complex

	Gal4/DNA (Zn MAD)			
	Peak	Edge	Remote	Native
Data collection				
Space group	C2	C2	C2	C2
Cell Parameters (Å)				
a	126.2	126.2	126.2	126.5
b	40.7	40.7	40.7	40.8
c	89.7	89.7	89.7	90.4
Wavelength	1.2823	1.2830	1.2448	
Resolution range	50–3.3	50–3.3	50–3.3	50–2.6
Total reflections	44,842	45,415	46,761	
Unique reflections	6798	6839	6892	
Completeness	95.6 (76.4)	96.1 (80.4)	97.4 (89.7)	
I/σ	16.2 (5.7)	20.3 (6.1)	18.1 (6.8)	
R_{merge} (%)[a]	8.4 (19.1)	7.2 (18.1)	7.6 (20.0)	
Refinement parameters				
No. of atoms				
Non-hydrogen atoms				2311
Waters				33
Zinc atoms				4
R_{factor} (%)[b]				23.0
R_{free} (%)[c]				27.5
Rmsds				
Bond length (Å)				0.014
Bond angles (°)				1.567

Values for the outer resolution shell are given in parentheses. MAD = multiwavelength anomalous dispersion; rmsd = root-mean-square deviation.

[a] $R_{merge} = \Sigma\Sigma_i|I_i - < I >|\Sigma/< I >$, where $< I >$ is the mean intensity of the N reflections with intensities I_i and common indices *h*, *k*, and *l* for the native and derivative crystals, respectively.

[b] $R_{factor} = \sum_{hkl}||F_{obs}| - |F_{cal}||/\sum_{hkl}|F_{obs}|$, where F_{obs} and F_{cal} are observed and calculated structure factors, respectively.

[c] For R_{free}, the sum is extended over a subset of reflections (10%) that were excluded from all stages of refinement.

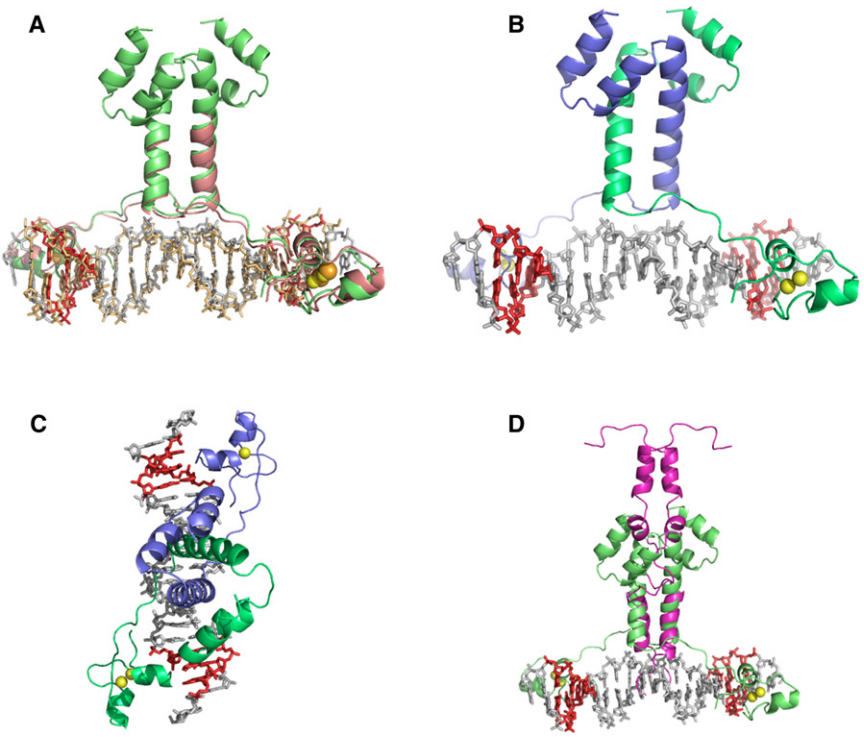

Figure 1. Crystal Structure of the Gal4-(1–100)/DNA Complex and Comparison to Related Structures

(A) The structure of the Gal4(1–100)/DNA (green/gray) and Gal4(1–65)/DNA (pink/orange) complexes are superimposed.

(B) A view of the Gal4(1–100)/DNA complex looking perpendicular to the 2-fold axis. The two subunits of the dimer are colored in blue and green, respectively, with zinc ions in yellow. The DNA is colored gray with the CGG half-sites colored red.

(C) A view looking down the 2-fold axis.

(D) Superimposition of the crystal structure of the Gal4(1–100)/DNA complex with the NMR solution structure of the Gal4(50–106) dimerization domain (magenta).

and 3C, each of these mutants showed reduced DNA binding, although the effects were relatively modest, with the most debilitating L67A/I80A/L81A triple mutant showing only about a 2-fold reduction in binding. We also prepared more drastic mutations to glutamate residues of residues 67, 80, 81, 89, and 93; however, each of these mutant proteins were unstable and could not be purified, arguing that more drastic mutations are not tolerated (data not shown). Taking this result together with the more dramatic reduction in DNA binding for Gal4(1–65) relative to Gal4(1–100), we conclude that dimer interactions are not additive, but function cooperatively to facilitate Gal4 binding to DNA.

Role of Gal4 Dimerization on Protein Stability

To investigate the role of Gal4 dimerization on protein stability, we studied the wild-type and mutant Gal4(1–100) proteins with differential scanning calorimetry (DSC). The wild-type protein shows a complex thermal denaturation curve that is irreversible and yields three melting transitions (T*m1*, T*m2*, and T*m3*) (Figure 4A and Table 2). We hypothesize that T*m1* corresponds to melting of the Gal4 dimer, with T*m2* corresponding to melting of the dimerization domain within a protein subunit, and T*m3* corresponding to the melting of the binuclear cluster domain.

The wild-type protein shows melting values of 69.4°C, 83.4°C, and 88.0°C for T*m1*, T*m2*, and T*m3*, respectively. Analysis of the L67A/I71A, L67A/I80A/L81A, L67A/I89A, and L67A/L93A mutants reveals decreased T*m* values for each of the three transitions, suggesting an overall reduced thermal stability of each of the mutants. Interestingly, T*m1*, which we hypothesize corresponds to the dimer-monomer transition, shows the greatest decrease in thermostability, with a decrease of between 8°C and 14°C, depending on the specific mutant. The other thermal transitions show more modest decreases in thermal stability of between 1°C and 6°C. This result demonstrates that the dimerization domain of Gal4 plays an important role in Gal4 thermostability.

We further evaluated the effect of dimerization mutants (L67A/I71A, L67A/I80A/L81A, L67A/I89A, and L67A/L93A) on the mobility of the Gal4(1–100)/DNA complex using molecular dynamic (MD) simulation to account for protein flexibility within a simulated solution environment. The starting structures for wild-type and mutants were directly extracted or mutated from the wild-type crystal structure reported here. All systems were subjected to 600 ps MD simulation separately and the averaged root-mean-square fluctuation (rmsf) values of each residue were used to measure the protein stability and to gain insight into possible structural fluctuation. The results of this analysis are presented in Figure 4B and reveal that each of the protein mutants produces larger protein fluctuations than the wild-type protein. Taking together the experimental and theoretical results, our data are consistent with an important role for the Gal4 dimerization interface that is observed in the structure of the Gal4(1–100)/DNA complex for protein thermostability.

Gal4 Dimerization as a Function of DNA Binding

A surprising outcome of our studies is that the dimerization domain of Gal4(1–100) bound to DNA shows considerable divergence from the more elongated DNA-free dimerization domain, as determined by solution NMR (Hidalgo et al., 2001; Figure 1D). This comparison suggests that either the Gal4 dimerization domain might undergo a DNA-induced conformational change from the DNA-free elongated NMR conformation to the more compact DNA-bound conformation. Alternatively, the observed solution NMR structure might be an artifact of the conditions used in the NMR study, and, therefore, does not represent a populated solution conformation in vivo.

Given the unusual elongated shape of the free Gal4 dimerization domain, as measured by solution NMR, we first tested whether the NMR solution structure represents the only dimer conformation that could form in solution. To do this, we prepared

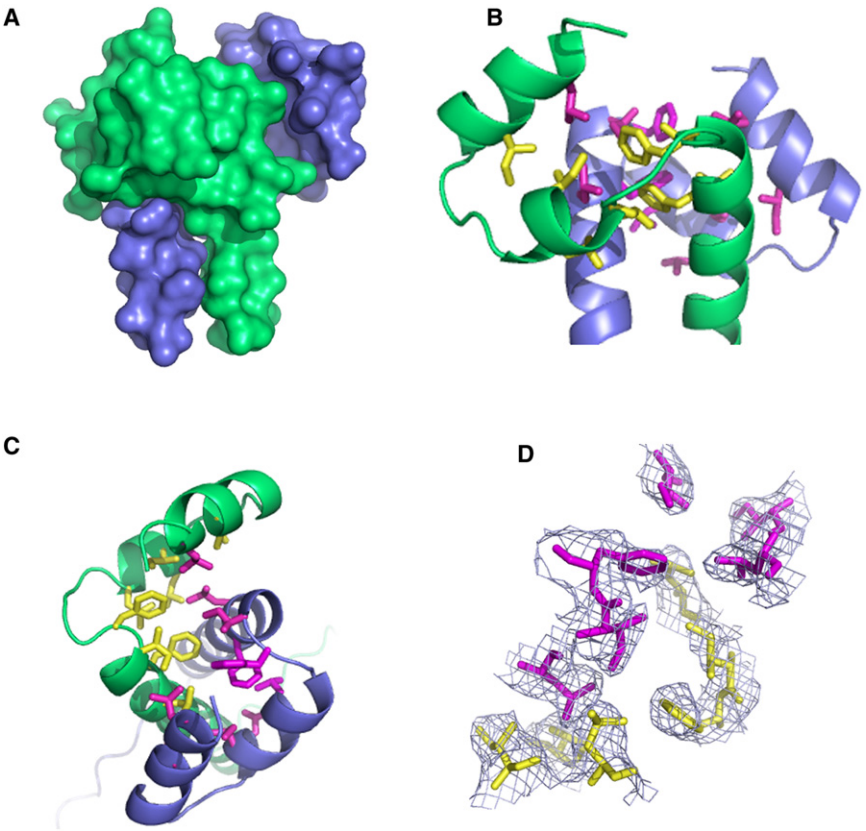

Figure 2. The Gal4 Dimerization Domain
(A) A surface representation of the Gal4 dimerization interface. The two Gal4 subunits of the dimer are colored green and blue.
(B) Details of the dimerization interface highlighting the protein side chains (yellow and purple for the two subunits, respectively). The view is perpendicular to the 2-fold axis.
(C) As in (B), but with a view along the 2-fold axis.
(D) Simulated annealing omit map of a region of the Gal4 dimerization interface.

residues 50–100 of the Gal4 dimerization domain and subjected it to sedimentation velocity experiments that are sensitive to molecular size and shape. This analysis reveals a single peak with a sedimentation coefficient of 1.28 s with an rmsd of less than 0.006 absorbance units (Figure 5A). For comparison, we calculated theoretical sedimentation coefficients for residues 50–100 of the Gal4 dimers, as observed in the DNA-bound Gal4(1–100) structure and the Gal4(50–106) solution NMR structure. To obtain these theoretical sedimentation coefficients, we used the program HYDROPRO for each sample with the respective atomic coordinates (Garcia de la Torre et al., 2000; Garcia de la Torre 2001). This analysis produced sedimentation coefficients of 1.18 and 1.08 s for the DNA-bound Gal4(1–100) structure and the Gal4(50–106) solution NMR structure, respectively (Table 3). These data demonstrate that the free Gal4 dimerization domain can adopt a relatively compact structure in solution in the absence of DNA that is more similar to that observed in the crystal structure in the presence of DNA than the DNA-free solution NMR structure. In addition, the difference in sedimentation coefficients between the experimental and theoretical sedimentation values of the Gal4 dimerization domain based on the crystal structure of the Gal4(1–100)/DNA complex suggests that the dimerization domain might undergo some sort of

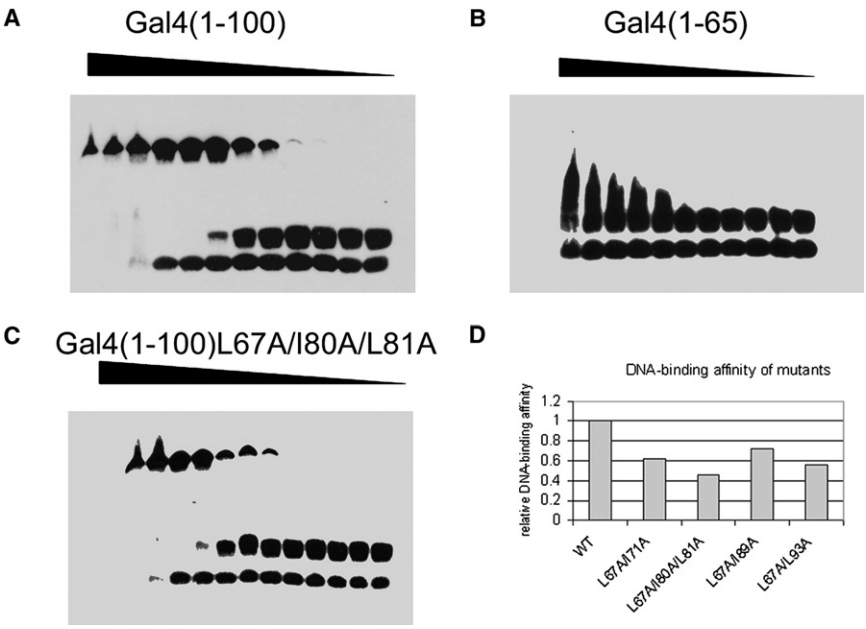

Figure 3. EMSA of Gal4(1–100) and Mutants
(A) EMSA of wild-type Gal4(1–100) with USA_{Gal4}. The dimer concentration starts at 400 nM and decrease in the direction of the half arrow to 1.56 nM.
(B) EMSA of wild-type Gal4(1–65) with USA_{Gal4}.
(C) EMSA of the Gal4(1–100)L67A/I80A/L81A mutant with USA_{Gal4}.
(D) Summary of relative binding affinity of Gal4-(1–100) and mutants for USA_{Gal4}. The apparent K_d for wild-type Gal4(1–100) WT and mutants L67A/I71A, L67A/I80A/L81A, L67A/I89A, and L67A/L93A are 24.58 ± 4.60 nM, 40.00 ± 3.54 nM, 54.17 ± 5.89 nM, 34.17 ± 7.20 nM, and 44.38 ± 3.20 nM respectively. The apparent K_d for Gal4(1–65) is out of the measurement range used here, and can be approximated to be >400 nM.

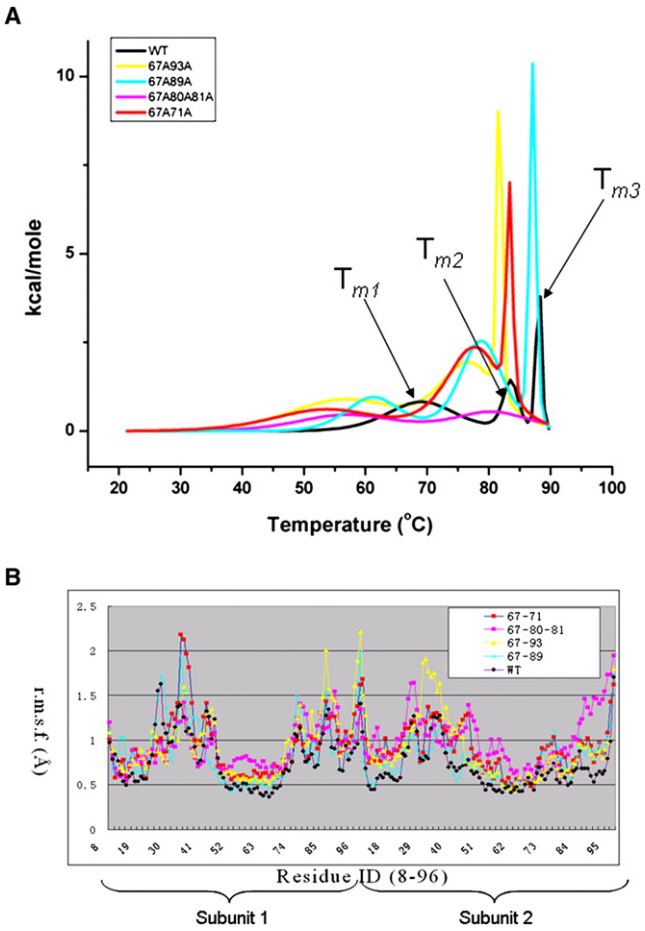

Figure 4. DSC and MD Simulations of Gal4(1–100) and Mutants

(A) DSC profiles of wild-type and mutant Gal4(1–100) proteins are color coded, as indicated.

(B) Residue specific root-mean-square fluctuation (rmsf) values for wild-type and mutant Gal4(1–100) bound to DNA as calculated with MD simulations.

conformational change upon DNA binding, although one that is less dramatic than implied from the solution NMR structure.

We also carried out a sedimentation velocity experiment on the Gal4(1–100)/DNA complex and obtained a single peak with a sedimentation coefficient of 3.21 s that is very similar to a theoretical value of 3.12 s (Figure 5B and Table 3). This result is consistent with the conclusion that the observed crystal structure of Gla4(1–100)/DNA is representative of the complex that forms in solution.

Mapping of Gal4 Interaction with Gal11P

Gal11 is a component of the RNA polymerase holoenzyme, and, under normal conditions, does not interact with the dimerization domain of Gal4. However, a single mutation of N342 to one of several hydrophobic amino acids in Gal11 converts it into a transcriptional potentiator (Gal11P) that interacts directly with the dimerization domain of Gal4. Based on EMSA analysis and NMR titration studies, residues F68, L70, I71, F72, D78, S85, K90, L92, T94, G95, L96, and D100 of the Gal4 dimerization domain were hypothesized to participate in Gal11P interaction (Hidalgo et al., 2001). A mapping of these residues onto the dimerization

Table 2. DSC Analysis of Gal4(1–100) and Mutants

Melting Transition (°C)	WT	L67A/I71A	L67A/I80A/L81	L67A/I89A	L67A/L93
Tm1	69.4	54.8	56.8	61.4	55.6
Tm2	83.4	76.8	80.2	78.8	76.8
Tm3	88.0	82.9	N/A	87.1	81.7

domain of Gal4 within the Gal4(1–100)/DNA complex shows that Gal11P interacts with a clustered patch on the top surface of helical bundle region (Figures 6A and 6B). This proposed, more localized Gal4 contact surface with Gal11P differs from the more extended contact surface that was proposed based on the solution NMR structure of the more extended Gal4 dimer (Hidalgo et al., 2001; Figure 6C). We propose that the Gal11P contact surface mapped here onto the Gal4(1–100)/DNA complex might be more biologically relevant.

DISCUSSION

It has long been known that Gal4 dimerization is essential for Gal4 function in vivo (Keegan et al., 1986; Ma and Ptashne, 1987a, 1987b). In this study, we reveal the molecular basis for Gal4 dimerization and the sterochemical basis for how this dimerization facilitates DNA-specific binding. We also show that Gal4 dimerization contributes to protein thermostability, and map a Gal4 dimerization surface for interaction with the Gal11P transcriptional coactivator.

Gal4 was one of the earliest eukaryotic DNA binding transcription factors that were characterized and analyzed at the cellular and molecular level, and has thus been used by many as a paradigm for understanding DNA binding by eukaryotic transcription factors (Ptashne et al., 1982). Gal4 is now known to be a member of over 80 fungal-specific proteins that contain a Zn_2Cys_6 binuclear cluster domain and that bind predominantly as homodimers to DNA sites containing CG-rich DNA half-sites and, most often, CGG that are contacted by the Zn_2Cys_6 domain through major groove and phosphate backbone interactions (Marmorstein and Fitzgerald, 2003; Todd and Andrianopoulos, 1997). While proteins that contain a Zn_2Cys_6 domain bind the same or highly related CGG half-sites, specificity for the half-site separation and polarity is dictated by the distinct configurations of the linker and dimerization regions of these proteins. Previous studies on the Gal4 (Marmorstein et al., 1992), Ppr1 (Marmorstein and Harrison, 1994), Put3 (Swaminathan et al., 1997), Hap1 (King et al., 1999a, 1999b; Lukens et al., 2000), and Leu3 (Fitzgerald et al., 2006) Zn_2Cys_6 domain proteins reveal that dimerization often involves a conserved coiled-coil domain, but more divergent features directly C terminal and N terminal (linker region to the Zn_2Cys_6 domain) to the coiled-coil domain. While these N-terminal regions clearly play a role in DNA binding specificity by these proteins as they modulate the inter-half-site spacing preference, the C-terminal regions have less clear functions. In the case of Gal4, this region forms a helical bundle with the coiled-coil region to effect not only DNA binding, but also protein stability and interaction with other transcription factors, such as Gal11P. It is not yet clear what role the corresponding regions play in other Zn_2Cys_6 domain proteins, but the sequence divergence in this region suggests that they might form different

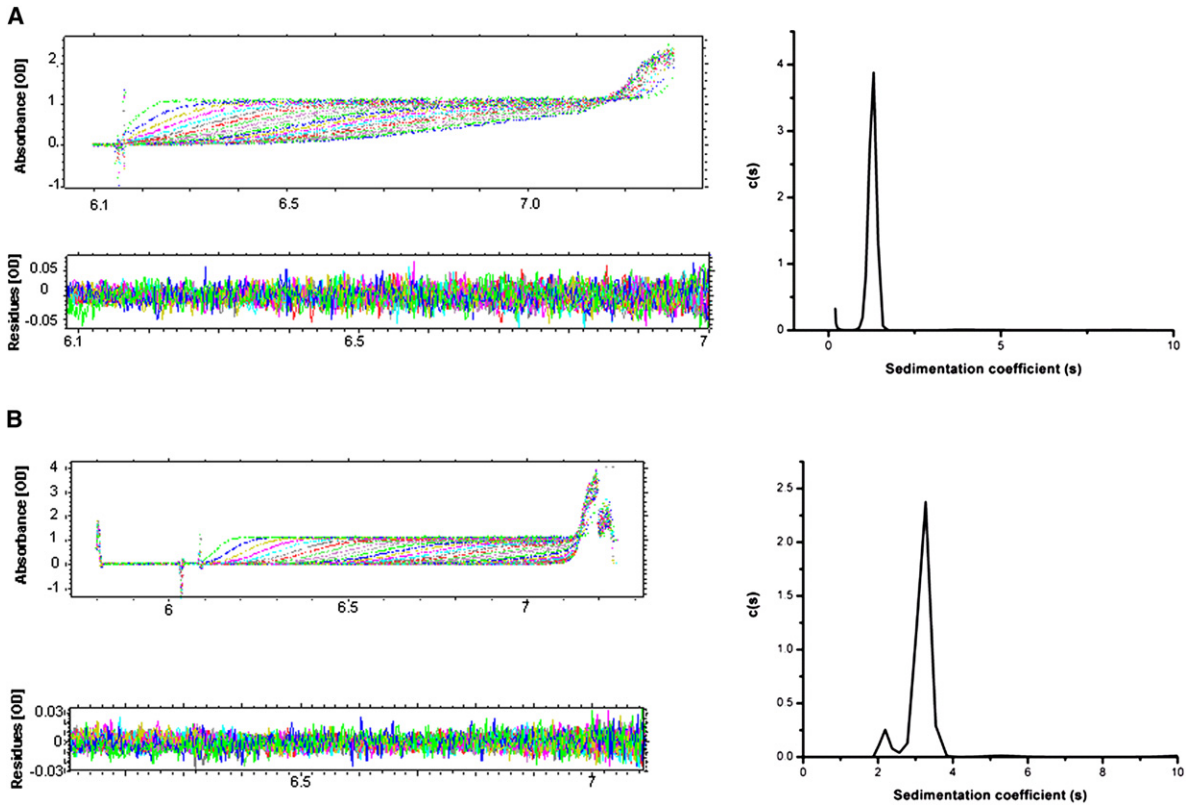

Figure 5. Sedimentation Coefficient Distribution of Gal4 Proteins

A sedimentation velocity profile is shown for (A) Gal4(50–100) and (B) Gal4(1–100)/DNA complex. The upper panels show the raw sedimentation signals acquired at different time points, the middle panels show the residual maps of the fittings, and the lower panels are the calculated sedimentation coefficient distributions for the corresponding samples.

structures with different functions that contribute to protein-specific activities among the family of proteins that contain a conserved Zn_2Cys_6 domain.

EXPERIMENTAL PROCEDURES

Protein and DNA Preparation

Gal4(1–100) was expressed in XA90 *Escherichia coli* cells, grown at 37°C in LB media to an OD_{595} of ~0.6, induced with 0.5 mM IPTG and 100 μM $Zn(OAc)_2$ and grown for an additional 3.5 hr. The cell pellet was recovered by centrifugation and sonicated in buffer A containing 20 mM HEPES (pH 7.5), 150 mM NaCl, 10 μM $Zn(OAc)_2$, 1 mM DTT, and 100 μg/ml PMSF, and centrifuged to isolate the supernatant. The supernatant was loaded on an SP Sepharose fast flow resin (GE Healthcare), washed with buffer A and eluted with a 0.15 mM to ~1 M NaCl gradient in buffer A. The peak fractions containing Gal4(1–100) were pooled and precipitated with 50% w/v $(NH4)_2SO_4$. The centrifuged pellet was resuspended in buffer A without DTT and PMSF and applied to a Superdex 75 (Amersham) gel filtration column. The resulting peak fractions were pooled and centrifuged with a Millipore Amicon Ultra 5K concentrator to a concentration of ~50 mg/ml and stored at −80°C before crystallization. Protein purity was judged to be over 95% by SDS-PAGE electrophoresis using Simple Blue Safestain (Invitrogen). Purified oligonucleotides were obtained from Integrated DNA Technologies (IDT).

Table 3. Sedimentation Coefficients of Gal4 Constructs and DNA Complex

	Gal4(50–100)	Gal4(1–100)/DNA
$S_{Theoretical}$		
NMR structure	1.08	2.85
Crystal structure	1.18	3.12
$S_{Experimental}$	1.28	3.21

S = Svedberg.

Crystallization, Data Collection, and Structure Determination

Crystals of Gal4(1–100) bound to a consensus 20 base pairs oligonuleotide (5′-TCC GGA GGA CTG TCC TCC GG-3′ and 5′-ACC GGA GGA CAG TCC TCC GG-3′) were obtained by vapor diffusion at room temperature with a 2 μl hanging drop containing 0.6 mM protein, 0.8 mM DNA duplex, 40 mM $Mg(OAc)_2$, 25 mM sodium phosphate (pH 5.5), 5% PEG400, and 5% MPD equilibrated over a reservoir solution containing twice the concentrations of salts, buffer, and precipitant. The crystals were cryoprotected by a direct dunk into a reservoir solution supplemented to 25% MPD, and then flash frozen in propane. A Zn multiwavelength anomalous dispersion (MAD) dataset was collected at BNL, beamline X-25, and data were processed and scaled with HKL2000 (Otwinowski and Minor, 1997). Two Zn atoms were found and refined to give model phases with the program SOLVE (Terwilliger and Berendzen, 1999). A solvent-flattened electron density map was obtained with the program CNS (Brunger, 2007; Brunger et al., 1998), and this map revealed unambiguous density for most of the DNA duplex, the Zn binding domain, linker, and about two thirds of the coiled-coil region of the dimerization domain. The program O (Jones et al., 1991) was used to build the complex, with the Gal4-(1–65)/DNA complex as a starting point. Following initial model building, the structure was refined in CNS to a resolution of 3.2 Å resolution. This partially built and refined model was then used as a search model with molecular replacement in CNS with a 2.6 Å native dataset. Additional electron density was observed for the rest of the dimerization domain, which was modeled

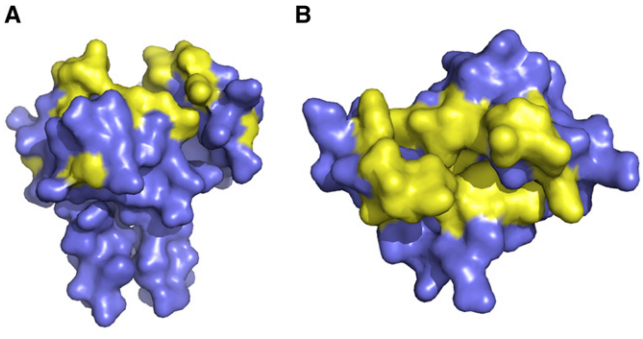

Figure 6. Mapping of Gal4 Contact Surface for Gal11P

(A and B) Orthogonal views of the mapping of the Gal11P contact surface onto the dimerization domain of the Gal4(1–100)/DNA structure. Gal4 residues that have been implicated to play a role in Gal11P contact are colored yellow.

(C) A mapping of the Gal11P contact surface onto the dimerization domain of the Gal4(50–106) solution NMR structure.

with O. Additional refinement in CNS with simulated annealing and torsion angle dynamic was carried out. Toward the end of refinement, restrained individual isotropic atomic B factors were adjusted, and solvent molecules were built into regions showing strong $F_o - F_c$ density. In the final step, TLS refinement was carried out, and the final model was checked for errors with a composite omit map (Table 1). Structural figures were created with PyMOL (DeLano, 2002).

EMSA Studies

The following DNA probe (IDT) was used in EMSA: 5′ biotin-TCT TCG GAG GGC TGT CAC CCG AAT ATA. The complementary strand also contained a 5′ biotin group. The DNA was annealed and diluted in renaturation buffer (RnB; 20 mM HEPES (pH 7.5), 50 mM NaCl) to a final reaction DNA duplex concentration of 0.5 nM. For the EMSA assay, protein was diluted in RnB with 100 μg/ml BSA (DB) by 2-fold serial dilution from 4 μM to 15.625 nM (dimer). Protein and DNA were equilibrated at room temperature for 30 min in reaction buffer RnB with an additional 5% w/v Ficol. After incubation, the samples were loaded onto a 6% DNA polyacrylamide retardation gel (Invitrogen) in 0.5× Tris-borate-EDTA (TBE) and run at 150 V for 40 min at 4°C. The gel was blotted on Biodyne B Nylon membrane (Pierce) at 380 mA for 1 hr in 0.5× TBE at 4°C. The transferred DNA was then crosslinked to the membrane with a Stratagene crosslinker. The DNA signal was visualized with the Lightshift kit (Pierce) and exposed to film that was developed and scanned. The band intensity of the protein/DNA complex band was used to calculate the apparent K_d for dimer binding. Specifically, the concentration of the shifted band that corresponded to half the intensity of the DNA-alone band was considered the apparent K_d. Each analysis was carried out in triplicate and used to calculate the overall apparent K_d with its associated standard deviation. In the blotted gel, two bands were seen for the probe. The upper band corresponds to the annealed duplex that is shifted with Gal4 protein, while the lower band likely corresponds to single-stranded DNA that is incompletely annealed to form duplex. We have therefore used the shift of the top probe for K_d calculations. All proteins used in the EMSA analysis were purified to homogeneity through gel filtration chromatography, which demonstrated that all proteins were properly folded.

Analytical Ultracentrifugation and DSC Analysis

Sedimentation velocity experiments were performed with a Beckman Optimal XL-A analytical ultracentrifuge at 20°C in 1× PBS buffer with a 60Ti rotor at 50,000 rpm. Samples were detected with optical absorbance at 280 nm. The experimental A(r, t) data were analyzed with the program SEDFIT, applying the continuous c(s) distribution model (Balbo and Schuck, 2002; Schuck, 2000). The partial specific volumes, the densities, and the viscosities of the buffers were calculated with the program SEDNTERP (Philo, 1997). The program HYDROPRO was used to compute theoretical sedimentation coefficients of the respective macromolecules from their atomic coordinates, with primary data including solvent density, temperature, and molecular weight.

DSC experiments were performed with an MCS differential scanning calorimeter. Each scan was made under nitrogen gas protection with a range of 15°C–90°C at a scanning rate of 90°C/hr. The reference buffer contained 50 mM HEPES (pH 7.5), 150 mM NaCl, and 10 μM $Zn(OAc)_2$, and protein was diluted in the reference buffer to a concentration of 1 mg/ml. All samples and buffers were degassed for 5 min before loading into the cell. Data were analyzed with software provided with the instrument by subtracting the buffer control, subtracting the baseline fitted to the ends of transition, normalizing for the protein concentration, and curve fitting with nonlinear least-squares regression analysis.

MD Simulation

Several MD simulation sets were carried out on wild-type and mutant Gal4/DNA complex structures separately with the AMBER 9.0 program and the Parm99 force field (Pearlman et al., 1995). The complex structures were solvated with a box of TIP3P water molecules extending at least 10 Å away from the boundary of any macromolecule atoms. An appropriate number of counterions were added to neutralize the system. The particle mesh Ewald method was employed to calculate long-range electrostatic interactions. All the MD runs were set up with the same protocol. First, each solvated Gal4/DNA complex was subjected to 200 steps of minimization with the steepest descent method followed by conjugate gradient to remove conflicts possibly existing between solvent molecules and the complexes. During this process, the macromolecules were held fixed. Then, a second minimization of 500 steps was performed on the entire protein/DNA/water complex. The relaxed structures were then subjected to MD simulations. Each system was gradually heated from 0K to 300K in 15 ps with three intervals, and then equilibrated for 25 ps at 300K, followed by a data collection run, giving a total simulation time of 600 ps for all systems. The nonbonded cutoff was set to 8.0 Å, and the nonbonded pairs were updated every 25 steps. The SHAKE method was applied to constrain all covalent bonds involving hydrogen atoms. Each simulation was coupled to a 300K thermal bath at 1.0 atm pressure by applying the algorithm of Berendsen. The temperature and pressure coupling parameters were set as 0.2 ps and 0.05 ps, respectively. An integration time step of the MD calculations was 2 fs. In the energy minimizations and MD simulations, periodic boundary conditions were applied in all directions. The analyses of the simulations focused on the production stages. The rmsf of each residue was calculated similarly. The interactions between the interfaces of Gal4 dimer were analyzed on the completed models with the program LIGPLOT (Wallace et al., 1995).

ACCESSION NUMBERS

Coordinates of the Gal4(1–100)/DNA structure have been submitted to the PDB with accession code 3COQ.

ACKNOWLEDGMENTS

We thank G. Van Duyne, K. Gupta, and G.S. Pesiridis for assistance with the Beckman Optimal XL-A analytical ultracentrifuge and helpful advice; K. Zhao, K. Li, and X. Liu for help on data collection and useful discussions; and the staff at beamline X-25 at the National Synchrotron Light Source (NSLS) and beamline A1 at the Cornell High Energy Synchrotron Source (CHESS) for assistance. CHESS is supported by the National Science Foundation under award DMR 0225180, and the Macromolecular Diffraction at CHESS facility is supported by award RR-01646 from the National Institutes of Health (NIH), through its National Center for Research Resources. The NSLS is supported by the U.S. Department of Energy under contract DE-AC02-98CH10886. This work was supported by an NIH grant to R.M. (GM 052880). The authors declare that they have no competing financial interests.

Received: February 6, 2008
Revised: March 28, 2008
Accepted: March 28, 2008
Published: July 8, 2008

REFERENCES

Balbo, A., and Schuck, P. (2002). Analytical ultracentrifugation in the study of protein self-association and heterogeneous protein-protein interactions. In Protein-Protein Interactions: A Molecular Cloning Manual, E. Golemis and P.D. Adams, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Baleja, J.D., Marmorstein, R., Harrison, S.C., and Wagner, G. (1992). Solution structure of the DNA-binding domain of Cd2–GAL4 from *S. cerevisiae*. Nature *356*, 450–453.

Barberis, A., Pearlberg, J., Simkovich, N., Farrell, S., Reinagel, P., Bamdad, C., Sigal, G., and Ptashne, M. (1995). Contact with a component of the polymerase II holoenzyme suffices for gene activation. Cell *81*, 359–368.

Brunger, A.T. (2007). Version 1.2 of the crystallography and NMR system. Nat. Protoc. *2*, 2728–2733.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Carey, M., Kakidani, H., Leatherwood, J., Mostashari, F., and Ptashne, M. (1989). An amino-terminal fragment of GAL4 binds DNA as a dimer. J. Mol. Biol. *209*, 423–432.

DeLano, W.L. (2002). The Pymol Molecular Graphics System (Palo Alto, CA: Delano Scientific).

Farrell, S., Simkovich, N., Wu, Y., Barberis, A., and Ptashne, M. (1996). Gene activation by recruitment of the RNA polymerase II holoenzyme. Genes Dev. *10*, 2359–2367.

Fitzgerald, M.X., Rojas, J.R., Kim, J.M., Kohlhaw, G.B., and Marmorstein, R. (2006). Structure of a Leu3-DNA complex: recognition of everted CGG half-sites by a Zn2Cys6 binuclear cluster protein. Structure *14*, 725–735.

Garcia de la Torre, J. (2001). Hydration from hydrodynamics. General considerations and applications of bead modelling to globular proteins. Biophys. Chem. *93*, 159–170.

Garcia de la Torre, J., Huertas, M.L., and Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys. J. *78*, 719–730.

Gaudreau, L., Adam, M., and Ptashne, M. (1998). Activation of transcription in vitro by recruitment of the yeast RNA polymerase II holoenzyme. Mol. Cell *1*, 913–916.

Hidalgo, P., Ansari, A.Z., Schmidt, P., Hare, B., Simkovich, N., Farrell, S., Shin, E.J., Ptashne, M., and Wagner, G. (2001). Recruitment of the transcriptional machinery through GAL11P: structure and interactions of the GAL4 dimerization domain. Genes Dev. *15*, 1007–1020.

Himmelfarb, H.J., Pearlberg, J., Last, D.H., and Ptashne, M. (1990). GAL11P: a yeast mutation that potentiates the effect of weak GAL4-derived activators. Cell *63*, 1299–1309.

Hopper, J.E., Broach, J.R., and Rowe, L.B. (1978). Regulation of the galactose pathway in *Saccharomyces cerevisiae*: induction of uridyl transferase mRNA and dependency on GAL4 gene function. Proc. Natl. Acad. Sci. USA *75*, 2878–2882.

Johnston, M. (1987). A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*. Microbiol. Rev. *51*, 458–476.

Jones, T.A., Zou, J.Y., Cowen, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallog. A. *47*, 110–119.

Keegan, L., Gill, G., and Ptashne, M. (1986). Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science *231*, 699–704.

Kew, O.M., and Douglas, H.C. (1976). Genetic co-regulation of galactose and melibiose utilization in *Saccharomyces*. J. Bacteriol. *125*, 33–41.

King, D.A., Zhang, L., Guarente, L., and Marmorstein, R. (1999a). Structure of a HAP1-DNA complex reveals dramatically asymmetric DNA binding by a homodimeric protein. Nat. Struct. Biol. *6*, 64–71.

King, D.A., Zhang, L., Guarente, L., and Marmorstein, R. (1999b). Structure of HAP1–18-DNA implicates direct allosteric effect of protein-DNA interactions on transcriptional activation. Nat. Struct. Biol. *6*, 22–27.

Liang, S.D., Marmorstein, R., Harrison, S.C., and Ptashne, M. (1996). DNA sequence preferences of GAL4 and PPR1: how a subset of Zn2 Cys6 binuclear cluster proteins recognizes DNA. Mol. Cell. Biol. *16*, 3773–3780.

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.

Ma, J., and Ptashne, M. (1987a). Deletion analysis of GAL4 defines two transcriptional activating segments. Cell *48*, 847–853.

Ma, J., and Ptashne, M. (1987b). A new class of yeast transcriptional activators. Cell *51*, 113–119.

Marmorstein, R., and Harrison, S.C. (1994). Crystal structure of a PPR1-DNA complex: DNA recognition by proteins containing a Zn2Cys6 binuclear cluster. Genes Dev. *8*, 2504–2512.

Marmorstein, R., and Fitzgerald, M.X. (2003). Modulation of DNA-binding domains for sequence-specific DNA recognition. Gene *304*, 1–12.

Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S.C. (1992). DNA recognition by GAL4: structure of a protein-DNA complex. Nature *356*, 408–414.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E., DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995). AMBER: a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput. Phys. Commun. *91*, 1–41.

Philo, J. (1997). An improved function for fitting sedimentation velocity data for low molecular weight solutes. Biophys. J. *72*, 435–444.

Platt, T. (1984). Toxicity of 2-deoxygalactose to *Saccharomyces cerevisiae* cells constitutively synthesizing galactose-metabolizing enzymes. Mol. Cell. Biol. *4*, 994–996.

Ptashne, M., Johnson, A.D., and Pabo, C.O. (1982). A genetic switch in a bacterial virus. Sci. Am. *247*, 128–130, 132, 134–140.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys. J. *78*, 1606–1619.

Swaminathan, K., Flynn, P., Reece, R.J., and Marmorstein, R. (1997). Crystal structure of a PUT3-DNA complex reveals a novel mechanism for DNA recognition by a protein containing a Zn2Cys6 binuclear cluster. Nat. Struct. Biol. *4*, 751–759.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.

Todd, R.B., and Andrianopoulos, A. (1997). Evolution of a fungal regulatory gene family: the Zn(II)2Cys6 binuclear cluster DNA binding motif. Fungal Genet. Biol. *21*, 388–405.

Wallace, A.C., Laskowski, R.A., and Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. *8*, 127–134.

