Sutoh, K. (1982) *Biochemistry 21*, 3654-3661.
Sutoh, K. (1983) *Biochemistry 22*, 1579-1585.
Sutoh, K., Yamamoto, K., & Wakabayashi, T. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 212-216.
Sutoh, K., Tokunaga, M., & Wakabayashi, T. (1987) *J. Mol. Biol. 195*, 953-956.
Szent-Györgyi, A. G., Szentkiralyi, E. M., & Kendrick-Jones, J. (1973) *J. Mol. Biol. 74*, 179-203.
Takashi, R., Duke, J., Ue, K., & Morales, M. F. (1976) *Arch. Biochem. Biophys. 175*, 279-283.
Taylor, K. A., & Amos, L. A. (1981) *J. Mol. Biol. 147*, 297-324.
Toyoshima, C., & Wakabayashi, T. (1985) *J. Biochem. (Tokyo) 97*, 219-243.
Tyler, J. M., & Branton, D. (1980) *J. Ultrastruct. Res. 71*, 95-102.
Weeds, A. G., & Taylor, R. S. (1975) *Nature (London) 257*, 54-56.
Winkelmann, D. A., & Lowey, S. (1986) *J. Mol. Biol. 188*, 595-612.
Winkelmann, D. A., Lowey, S., & Press, J. L. (1983) *Cell 34*, 295-306.
Youngburg, G. E., & Youngburg, M. V. (1930) *J. Lab. Clin. Med. 16*, 158-166.

The DNA Binding Domain of GAL4 Forms a Binuclear Metal Ion Complex†

Tao Pan‡ and Joseph E. Coleman*
Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, Connecticut 06510
Received August 4, 1989; Revised Manuscript Received November 27, 1989

ABSTRACT: The transcription factor GAL4 from *Saccharomyces cerevisiae* requires Zn(II) or Cd(II) for specific recognition of the UAS_G sequence (Pan & Coleman, 1989). An N-terminal fragment consisting of the first 63 amino acid residues of GAL4 [GAL4(63)] has been obtained by partial tryptic proteolysis of a cloned and overproduced N-terminal domain of 149 residues, GAL(149*). We show that GAL4(63) contains the minimal GAL4 DNA binding domain. GAL4(63) binds tightly 1-2 mol of Zn(II) or 2 mol of Cd(II). ^{113}Cd NMR of ^{113}Cd(II)-substituted GAL4(63) reveals structural identity between the metal binding domains of GAL4(63) and that of the larger precursor GAL4(149*). ^{113}Cd(II) can be substituted for the Zn(II) in GAL4(63), and two ^{113}Cd NMR signals are observed at 706 and 669 ppm, both suggesting coordination of ^{113}Cd(II) to three or four $-S^-$ ligands. With the exception of the N-terminal methionine, the only sulfur-containing residues are the six highly conserved cysteines. High-resolution ^{1}H NMR of Zn(II)-GAL4(63) and Cd(II)-GAL4(63) show the two proteins to have almost identical conformations and to be present as monomers in solutions up to millimolar concentration. This leads us to postulate that GAL4 does not possess a TFIIIA-like "Zn-finger" but forms a binuclear metal cluster involving all six cysteines in a "cloverleaf"-like array. GAL4(63) contains about 60% α-helix, estimated from circular dichroism. Removal of the native Zn(II) causes substantial unfolding of the secondary structure. Unlike GAL4(149*), the resultant apoprotein is not induced to refold by readdition of Zn(II) at low concentrations.

The transcription factor GAL4 from *Saccharomyces cerevisiae* contains a sequence of Cys residues, $\underline{Cys}^{11}X_2\underline{Cys}^{14}X_6Cys^{21}X_6\underline{Cys}^{28}X_2\underline{Cys}^{31}X_6Cys^{38}$, within its N-terminal DNA binding domain (Keegan et al., 1986). Although a Zn(II) binding site involving four Cys (as underlined above), similar to the tetrahedral Zn(II) complex in the "Zn-fingers" of TFIIIA (Miller et al., 1985), has been proposed (Johnston, 1987), all six Cys residues are highly conserved among several other transcription factors from yeast. Other transcription factors from eukaryotic organisms, e.g., the steroid receptors, also contain similar clusters of Cys residues [Evans and Hollenberg (1988) and references cited therein]. It has been shown not only that the mutations of the four putative Cys ligands affect DNA binding but also that a $Cys^{38} \rightarrow Gly^{38}$ mutant is deficient in specific DNA binding (Johnston & Dover, 1987, 1988). Although a Cys^{21} mutation has not been reported to date, it is possible that this particular cysteine could also function as a ligand.

We have shown previously (Pan & Coleman, 1989) that a polypeptide consisting of the N-terminal 147 amino acids of GAL4 plus two additional amino acids at the C-terminus from the cloning vector [denoted GAL4(149*)],[1] overexpressed and purified from *Escherichia coli*, incorporates 1.0-1.5 mol of Zn(II). Zn(II) can be reversibly removed from GAL4(149*) accompanied by a total loss of specific DNA binding affinity. Zn(II) can be substituted by Cd(II) with complete restoration of DNA binding properties. ^{113}Cd NMR of the latter protein shows the protein to contain two ^{113}Cd binding sites. The chemical shifts of the two bound ^{113}Cd(II), 707 and 669 ppm, suggest that each is ligated to at least three sulfurs (Pan &

† This work was supported by NIH Grants DK09070 and GM21919. The 500-MHz NMR was supported by NIH Grant RR03475, NSF Grant DMB-8610557, and ACS Grant RD259.
* Corresponding author.
‡ This work is in partial fulfillment of the requirements for the Ph.D.

[1] Abbreviations: GAL4(149*), cloned polypeptide fragment of the GAL4 transcription factor from yeast containing the N-terminal 147 amino acid residues plus 2 from the cloning vector; GAL4(63), tryptic fragment of GAL4(149*) containing the N-terminal 63 amino acid residues; GAL4(62*), cloned polypeptide fragment of GAL4 containing the N-terminal 61 amino acid residues plus 1 from the cloning vector; NMR, nuclear magnetic resonance; COSY, two-dimensional correlated spectroscopy; CD, circular dichroism; UAS_G, DNA sequence known as the upstream activation sequence specific for genes of the galactose operon; DIPF, diisopropyl fluorophosphate; TSP, sodium (trimethylsilyl)tetradeuteriopropionate.

0006-2960/90/0429-3023$02.50/0

Coleman, 1989). Since GAL4(149*) contains only six Cys, it is possible that these cysteines are arranged such that a binuclear site could be formed with at least two shared $-S^-$ between Cd(II) ions. GAL4(149*) also contains three Met residues that could possibly function as ligands, although none of the Met are located within the DNA binding domain indicated by mutagenesis (residues 1–74) (Keegan et al., 1986).

Limited tryptic proteolysis of Zn(II)- or Cd(II)-GAL4(149*) generates a 13–14-kDa DNA binding core that retains DNA binding affinity for the UAS_G sequence (Pan & Coleman, 1989). As we show here, in the presence of excess Zn(II) a second stable fragment can be isolated from a tryptic digest. This polypeptide consists of the N-terminal 63 residues of GAL4 and in addition a small amount of the N-terminal fragment of 60 residues due to cleavage after Arg 60. These fragments are capable of specific DNA binding. Moreover, ^{113}Cd NMR of these fragments reveals that the Cd(II) binding motif of GAL4 involves only the six highly conserved Cys residues.

Materials and Methods

Isolation and Purification of GAL4(63). GAL4(149*) was overexpressed and purified from *E. coli* as described (Pan & Coleman, 1989). Tryptic proteolysis of GAL4(149*) was carried out in 10 mM Tris-HCl/300 mM NaCl/1 mM EDTA/7 mM β-mercaptoethanol/0.2 mM $ZnCl_2$/10% (v/v) glycerol, pH 8.0, at a ratio of 1:200 (w/w) trypsin:GAL4(149*) for 30–60 min at 22 °C. The proteolysis was stopped by the addition of 1 μL of DIPF/20 μg of trypsin and dialyzed against 10 mM Tris-HCl/1 mM EDTA/1 mM β-mercaptoethanol/10% (v/v) glycerol, pH 8.0 (STD buffer), plus 50 mM NaCl. The dialyzate was then applied to a Bluegel agarose column (Bio-Rad), washed with STD buffer plus 50 mM NaCl and with STD buffer plus 150 mM NaCl, and then eluted with STD buffer plus 250 mM NaCl. The eluted GAL4(63) was homogeneous as judged by SDS–PAGE.

Gel retardation assays were carried out as described (Pan & Coleman, 1989). The biotinylated GAL4(63) was obtained by addition of 1 μL of NHS-biotin (Sigma) in dehydrated acetonitrile at various concentrations (0.23, 0.78, and 1.56 mM final) to 14 μL of 840 μM GAL4(63) in 50 mM phosphate, pH 8.8/250 mM NaCl and incubated with vigorous vortexing for 1 h at room temperature. Avidin (Sigma) was dissolved in 1× buffer G and added in an appropriate amount to achieve a biotin:avidin molar ratio of 2:1.

^{113}Cd NMR spectra were recorded at 298 K on a Bruker AM-500 spectrometer (11.75 T, 110.9 MHz for ^{113}Cd) with a broad-band tunable probe using 10-mm NMR tubes. Spectral width was 15.2 kHz (136 ppm). The pulse angle was 45°, and a delay time of 2 s was used. Chemical shifts are reported relative to the signal of 0.1 M $Cd(ClO_4)_2$. The ^{113}Cd(II)-substituted GAL4(63) was obtained by addition of 2–3 ^{113}Cd(II) ions per molecule of Zn(II)-GAL4(63) followed by incubation for at least 12 h at room temperature. The excess ^{113}Cd(II) and exchanged free Zn(II) was then removed by dialysis. With this method, a ^{113}Cd(II)-GAL4(63) can be generated that contains 2.0 mol of Cd(II) and no Zn(II). The apo-GAL4(63) can be prepared by dialysis of Zn(II)-GAL4(63) at pH 5.0 in the presence of EDTA as described for GAL4(149*) (Pan & Coleman, 1989).

Results

Specific DNA Binding by GAL4(63). Upon limited tryptic proteolysis of GAL4(149*), a stable 13–14-kDa fragment (fragment I) is formed in the presence of Zn(II) or Cd(II). Fragment I possesses equal DNA binding affinity for specific GAL4 DNA as shown by GAL4(149*) (Pan & Coleman, 1989). However, when the proteolysis is carried out in the presence of a large excess of Zn(II) and β-mercaptoethanol (see Materials and Methods), this 13–14-kDa fragment is cleaved into two additional fragments (fragments II and III, Figure 1A) which are resistant to further proteolysis. Using carboxypeptidases A and B and amino acid analysis, we have identified fragment I as being cleaved after Arg^{123}, fragment II as being cleaved after Arg^{63} with a minor cleavage after Arg^{60}, and fragment III as being cleaved after Arg^{47}. Fragment II can be purified by applying the tryptic digest to a Bluegel agarose column, washing with 50 mM NaCl in standard buffer, and then eluting with 250 mM NaCl in standard buffer. After this treatment, fragment II contains approximately 1 mol of tightly bound Zn(II). In contrast, fragment III losses zinc rapidly when exogenous zinc is removed from the dialysis buffer.

In contrast to GAL4(149*), fragment II [denoted as GAL4(63)] does not retard the specific 17-mer derived from the UAS_G sequence (Keegan et al., 1986) on a 6% nondenaturing gel (Pan & Coleman, 1989). The likely explanation is that, in contrast to GAL(149*) and GAL4(123), folded GAL4(63) does not possess the interface required for dimer formation which is crucial for binding to its specific sequence

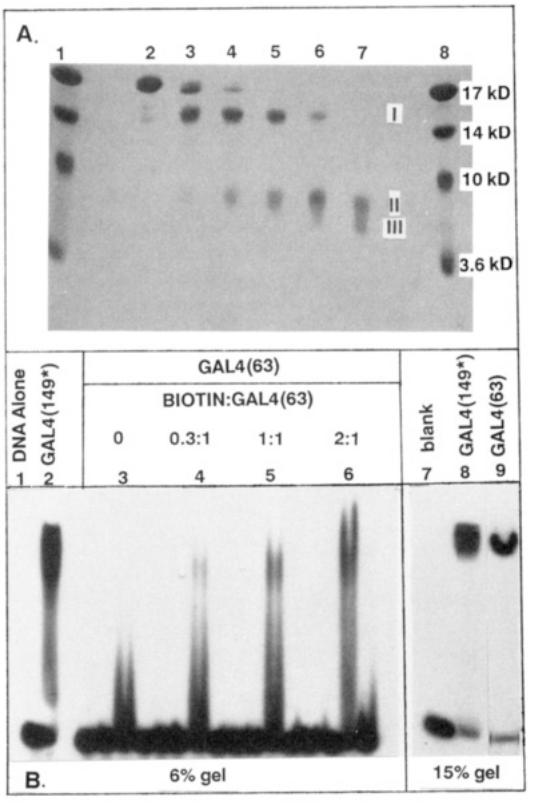

FIGURE 1: (A) Time course of limited tryptic proteolysis of GAL4(149*). Samples were withdrawn at 0, 10, 20, 30, 60, and 120 min, columns 2–7, respectively. Trypsin cleavage was stopped by boiling in SDS buffer for 5 min. Columns 1 and 8 are standard MW markers as indicated. (B) Gel retardation of biotinylated GAL4(63). All samples contained a 1000-fold excess of unlabeled calf thymus DNA relative to labeled specific UAS_G DNA. Protein and DNA were mixed and incubated for 20 min at room temperature prior to loading onto a 6% or 15% nondenaturing polyacrylamide gel as indicated on the figure. Lanes are as follows: (1) UAS_G DNA alone; (2) GAL4(149*), 1 μM; (3) GAL4(63), 100 μM, no biotin; (4) plus 0.3:1 biotin:GAL4(63); (5) plus 1:1 biotin:GAL4(63); (6) plus 2:1 biotin:GAL4(63); (7) UAS_G DNA plus biotin and avidin; (8) GAL4(149*); (9) GAL4(63) plus 1:1 biotin:GAL4(63). Solutions used for lanes 3–6 and 9 contained 1:2 avidin:biotin. The general conditions for both (A) and (B) are described under Materials and Methods.

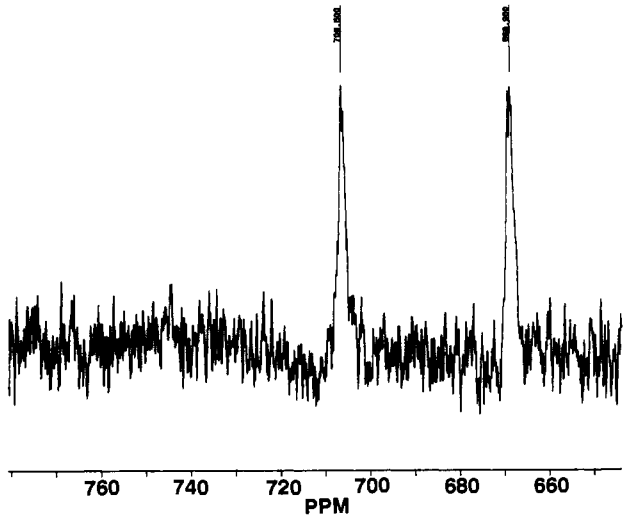

FIGURE 2: ^{113}Cd NMR spectra of N-terminal fragment of GAL4 protein, GAL4(63) (700 μM). Spectral width is 15.2 kHz (136 ppm), and the number of transients is 20000. A 20-Hz line broadening was applied for spectral enhancement.

as suggested by the nearly palindromic UAS_G consensus sequence. To test this possibility, we cross-linked on average one NHS-biotin per monomer of GAL4(63). Upon addition of an appropriate amount of avidin, a fraction of the biotinylated GAL4(63) will form a dimer of appropriate geometry for specific DNA binding (Figure 1B). The requirement of a higher concentration of biotinylated GAL4(63) relative to GAL4(149*) for equal retention of DNA (about 100-fold higher) is not surprising, since the formation of biotinylated GAL4(63) with the correct orientation for the dimer to align specific DNA binding surfaces is likely to be reasonably small, since there are 11 Lys residues in GAL4(63). All Lys are unlikely to be cross-linked in the proper conformation to allow binding of the dimer.

^{113}Cd NMR of GAL4(63) Protein. We have used ^{113}Cd NMR previously as a probe to determine the number of sulfurs involved in Cd(II) ligation [Pan & Coleman, 1989; Giedroc et al., 1989; for a review, see Armitage and Otvos (1982)]. ^{113}Cd(II)-GAL4(149*) shows two signals at 707 and 669 ppm relative to 0.1 M $^{113}Cd(ClO_4)_2$. The chemical shifts of both signals suggest ligation of Cd(II) to three if not four sulfurs (Pan & Coleman, 1989). ^{113}Cd(II)-GAL4(63) gives rise to the same two ^{113}Cd signals at 706 and 669 ppm, respectively (Figure 2). Since ^{113}Cd chemical shifts are highly sensitive to the ligand type, the almost identical chemical shift of the ^{113}Cd(II) in GAL4(63) is suggestive that GAL4(63) has the same conformation and ligands as does the same metal binding subdomain in GAL4(149*). Since the only residues containing sulfurs in addition to the N-terminal Met in GAL4(63) are the six conserved Cys, the Cd(II) ligation almost certainly involves all six cysteines. The most likely ligand arrangement to explain the ^{113}Cd chemical shifts is a two-metal cluster with two of the $-S^-$ ligands shared between ^{113}Cd(II) ions.

^{113}Cd NMR of GAL4(62).* Upon finding that the proteolytic product of GAL4(149*), GAL4(63), was soluble in marked contrast to the cloned N-terminal 74 amino acid residues (Pan & Coleman, 1989), we cloned and overproduced the N-terminal 62 amino acids, designated GAL4(62*), a construct which also turned out to be highly soluble. We now generally use GAL4(62*) for most experiments rather than the proteolytic product, GAL4(63). The ^{113}Cd NMR of GAL4(62*) is identical with that of GAL4(63) (Figure 3). A ^{113}Cd COSY spectrum of GAL4(62*) is also shown in Figure 3, an attempt to detect ^{113}Cd–^{113}Cd J coupling between

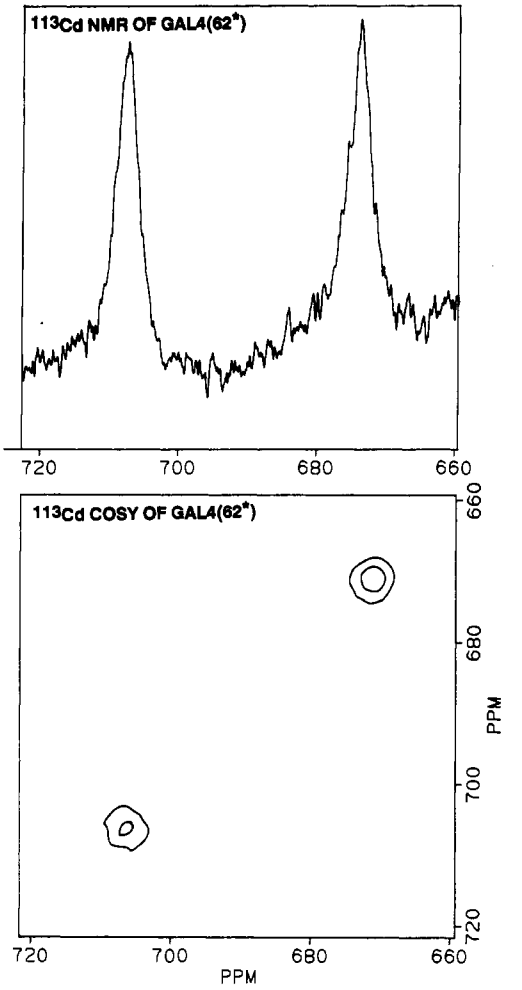

FIGURE 3: ^{113}Cd COSY of 4 mM ^{113}Cd(II)-GAL4(62*) in 40 mM Tris-HCl, pH 8.0. Spectral width is 7000 Hz; 1024 data points and 220 experiments with 256 scans each in the t_1 domain were collected. The normal ^{113}Cd NMR spectrum of GAL4(62*) is shown at the top of the 2D contour plot.

the two ^{113}Cd(II) ions which might be expected if bridging Cys sulfur ligands are present. No matter how many contours were plotted, no evidence could be found for ^{113}Cd–^{113}Cd cross peaks.

We now have obtained ^{1}H COSY spectra of the ^{112}Cd(II) ($I = 0$) and ^{113}Cd(II) ($I = {}^1/_2$) derivatives of GAL4(62*) in which the pattern of ^{113}Cd–$\beta\beta^1$H coupling demonstrates the presence of two bridging Cys ligands (Pan and Coleman, unpublished results). The detailed ^{1}H COSY study will be published separately but raises the question of why ^{113}Cd–^{113}Cd coupling is not observed in the ^{113}Cd COSY of Figure 3. There are two possible reasons: the coupling is too small, or there is an exchange process of the right frequency to broaden the ^{113}Cd resonances and obscure the J coupling. That there is an exchange process modulating the ^{113}Cd(II) sites in GAL4(62*) is shown by the broadening of the ^{113}Cd NMR signals as the frequency is lowered from 110 to 44 MHz (Figure 4). Ordinarily, the line width would be expected to be narrower at 44 MHz as observed for most ^{113}Cd(II)-substituted proteins. In the case of GAL4(63), however, both signals broaden, the 669 ppm signal more than the 706 ppm signal. Integration shows no significant loss of total area under the ^{113}Cd resonances. One of the few explanations for this phenomenon would be exchange broadening. The precise frequency of such a process (more effective at 44 MHz) is determined by a combination of the chemical shift difference between the exchanging species and the frequency of the ex-

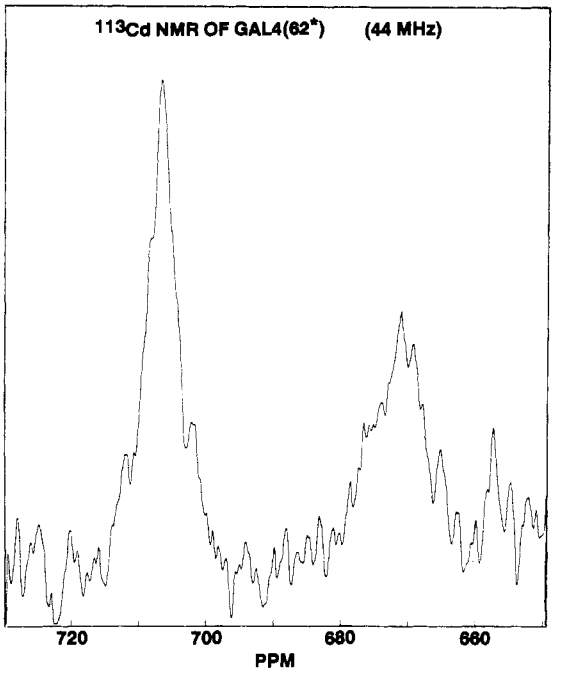

FIGURE 4: ^{113}Cd NMR spectrum of the same sample as in Figure 3 but acquired at 44.4 MHz. The number of transients is 54 000, and the relaxation delay is 1.5 s.

change. Previous modeling of such processes using several possible distributions of species and reasonable chemical shift differences suggests that the frequencies of effective intermediate exchange processes for $^{113}Cd(II)$-substituted proteins are on the order of 10^3 or 10^4 s^{-1}. These rates would not be expected to affect ^{113}Cd–^1H coupling as we have observed (Pan and Coleman, unpublished results).

$^{113}Cd(II)$–Zn(II) Exchange. In order to determine if both Cd(II) bound to GAL4(149*) could be replaced by Zn(II), a 10-fold excess of Zn(II) was added to the $^{113}Cd(II)$-GAL4(149*) sample shown in Figure 5A. Both ^{113}Cd signals disappear simultaneously, and after 4-h incubation at 22 °C the signals are no longer detectable (Figure 5B).

Cd(II)–Zn(II) Hybrid Form of GAL4(63). While a 10-fold excess of Zn(II) readily displaces the two $^{113}Cd(II)$ ions from GAL4(149*), as shown in Figure 5B, if the two $^{113}Cd(II)$ in GAL4(63) are titrated with increasing concentrations of Zn(II) from a 1:1 to 5:1 Zn(II):Cd(II) ratio, then the $^{113}Cd(II)$ ion with the NMR signal at 669 ppm can be differentially replaced. The ^{113}Cd NMR spectrum after incubation of $^{113}Cd(II)$-GAL4(62*) at a 3:1 molar ratio of Zn(II):protein is shown in Figure 5C. The $^{113}Cd(II)$ giving rise to the 669 ppm ^{113}Cd signal is largely displaced with Zn(II), while the $^{113}Cd(II)$ giving rise to the 706 ppm ^{113}Cd signal is still present.

^1H NMR of GAL4(63) and GAL4(62).* Freshly prepared GAL4(63) shows a highly resolved ^1H NMR spectrum (Figure 6A). Both GAL4(149*) and GAL4(123) form aggregates at NMR concentrations with broadened ^1H NMR signals (spectra not shown). Peak 3 of GAL4(63) is assigned to the 2-H of His^{53}. The protons responsible for peak 2 readily exchange with deuterons, while those responsible for peak 3 slowly exchange as well over several weeks at 4 °C. From the relayed ^1H COSY of GAL4(62*), the 4-H of His^{53} resonates at 6.9 ppm in the midst of the signals from Tyr^{40} (see below). Resonance 2 of GAL4(63) must be a slowly exchanging amide proton. Unexpectedly, this proton exchanges much more rapidly with deuterons in the case of GAL4(62*). The structural reason for this consequent to the removal of Arg^{63} is unclear.

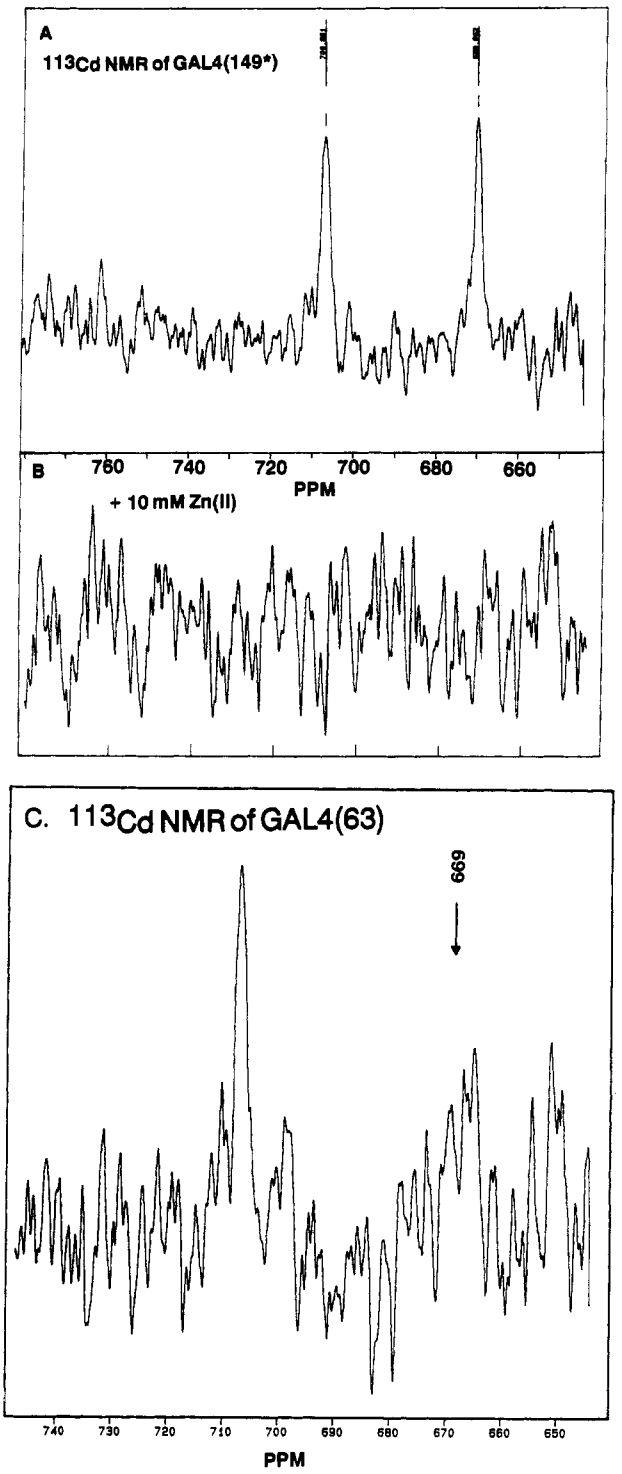

FIGURE 5: (A) ^{113}Cd NMR of $^{113}Cd(II)$-GAL4(149*) protein [450 μM, 1 mM in $^{113}Cd(II)$]. The number of transients is 7000 for this spectrum under the same conditions as in Figure 2. (B) Spectrum of the sample in (A) under the same acquisition conditions but after addition of 10 mM $ZnCl_2$ and 10 mM β-mercaptoethanol. (C) $^{113}Cd(II)$-GAL4(63) plus 3 equiv of Zn(II) per $^{113}Cd(II)$ ion. Zn(II) was added to $^{113}Cd(II)$-GAL4(63) [0.42 or 0.85 mM in $^{113}Cd(II)$] in 40 mM Tris/250 mM NaCl, pH 8.0, to a final concentration of 2.6 mM. The number of transients collected was 2200.

The aromatic proton spectrum of GAL4(62*) consisting of the protons of Trp^{36}, Tyr^{40}, and His^{53} is easily assigned on the basis of the relayed ^1H COSY shown in Figure 6B. The proton NMR spectra of the Cd(II)- and Zn(II)-GAL4(63) are very similar with the same line widths, suggesting that both derivatives are probably monomers. The only changes in the proton NMR spectrum induced by the Cd(II) substitution are

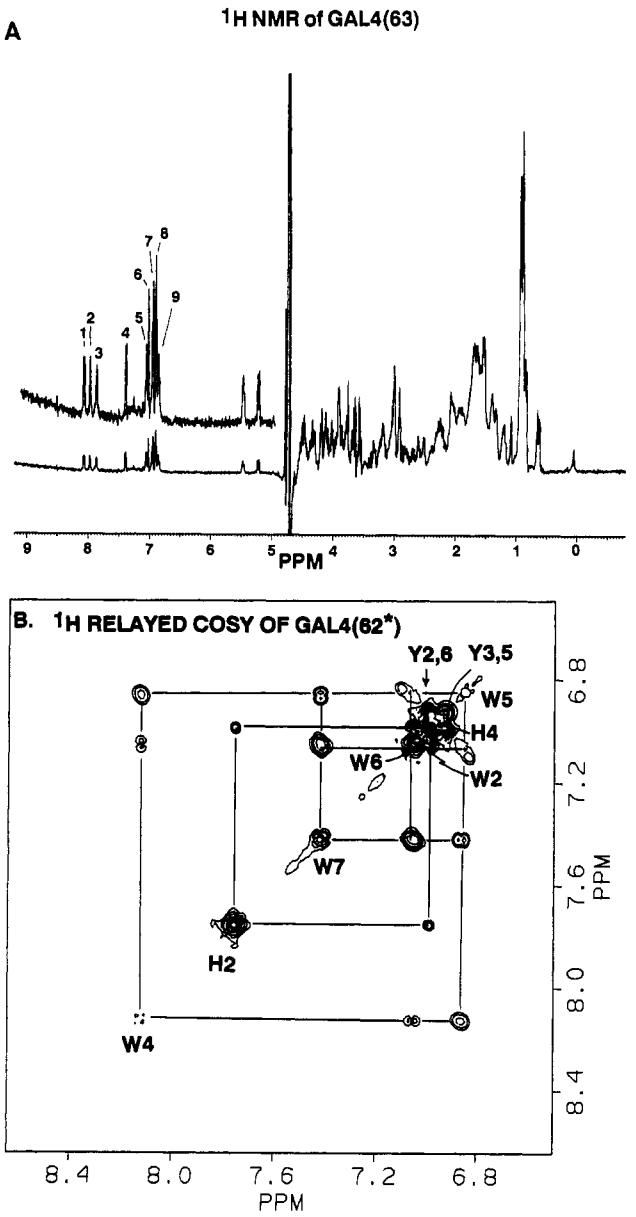

FIGURE 6: (A) ^1H NMR spectra of 600 μM Zn(II)-GAL4(63) in 50 mM phosphate/100 mM NaCl/D_2O, pH 8.0. The number of transients is 16, and the relaxation delay is 1 s. Chemical shift is plotted relative to that of TSP. (B) Aromatic region of the relayed ^1H COSY of $^{113}Cd(II)$-GAL4(62*) (see text).

small changes in the chemical shift of several of the aromatic and aliphatic protons (data not shown). Thus, there cannot be radical changes in conformation induced by the Cd(II) substitution.

Metal Ion Contents of the Zinc and Cadmium Derivatives of the GAL4 DNA Binding Domain. All three derivatives, GAL4(149*), GAL4(63), and GAL4(62*), bind 2 mol of Cd(II)/mol of protein (Table I). None of the Cd(II) is readily removed by dialysis. On the other hand, the cloned GAL4-(149*) is usually isolated containing between 1.2 and 1.5 mol of Zn(II)/mol of protein (Pan & Coleman, 1989) (Table I). If the GAL4(149*) (30 μM) is dialyzed against a buffer containing 6 μM Zn(II), then it maintains a content of 2 mol of Zn(II)/mol of protein (Table I). If GAL4(63) is isolated without maintaining additional Zn(II) in the buffers, then the proteolytic product tends to lose most of the second Zn(II) (Table I). The same is true of the cloned GAL4(62*). Thus, a second Zn(II) ion appears to bind less tightly than a second Cd(II) ion in the metal binding domain of GAL4.

Table I: Metal Content of GAL4(63)

protein	treatment	Zn	Cd
GAL4(149*)	dialysis[a]	2.03	
	dialysis (no exogenouse Zn)[b]	1.14–1.49	
GAL4(63)	preparation 1[c]	1.36	
	preparation 2	1.15	
	dialysis	1.02	
apoprotein	dialysis + 10 mM EDTA, pH 5	0.01	
Cd(II)	exchange	<0.05	2.0
fragment III [GAL4(47)]	dialysis (15 h)	0.38	

[a] Dialysis of a 30 μM protein solution was carried out against standard buffer containing 6 μM Zn(II). [b] Taken from Pan and Coleman (1989). [c] GAL4(63) eluted from Bluegel–agarose column.

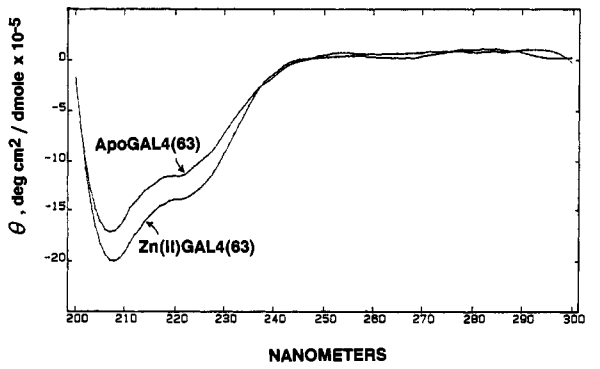

FIGURE 7: Circular dichroism of Zn(II)- and apo-GAL4(63) in 10 mM Tris/150 mM NaCl/1 mM β-mercaptoethanol/50 μM EDTA, pH 8.0, at 27 °C. Protein concentrations were 5 μM for both species.

Circular Dichroism of Zn(II)- and Apo-GAL4(63). The Zn(II)-GAL4(63) has significant molar ellipticity in the wavelength region of the peptide bond chromophores, -20.0×10^5 deg cm^2 dmol^{-1} at 208 nm and -13.9×10^5 deg cm^2 dmol^{-1} at 222 nm (Figure 7). A reasonable graphical fit is obtained by a combination of 60% α-helix, 0% β-sheet, and 40% random coil (Greenfield & Fasman, 1969), in good agreement with a Chou–Fasman prediction of the secondary structure of GAL4(63) (66% α-helix, 0% β-sheet, 34% β-turns + random coil). Removal of Zn(II) causes a relatively large change in the secondary structure, as shown by a decrease in negative molar ellipticity of $+2.3 \times 10^5$ deg cm^2 dmol^{-1} at 222 nm and $+3 \times 10^5$ deg cm^2 dmol^{-1} at 208 nm. In contrast to GAL4(149*), readdition of Zn(II) (20 μM) to apo-GAL4(63) (5 μM) does not reconstitute the original CD spectrum of Zn(II)-GAL4(63) under these conditions.

Discussion

The N-Terminal 63 Amino Acids of GAL4 Form the DNA Binding Surface. It has been shown that the N-terminal 74 amino acids of GAL4 contain all the amino acid determinants required for specific DNA binding (Keegan et al., 1986). The number 74, however, is artificial since it is entirely based on an easily accessible restriction site of GAL4 cDNA. Mutational analysis of GAL4 shows that most missense mutations that inactivate the DNA binding of GAL4 cluster within residues 10–57 (Johnston & Dover, 1987, 1988). The present study reveals that tryptic proteolysis defines a GAL4 fragment resistant to proteolysis which is a metal ion binding subdomain consisting of only the N-terminal 63 amino acid residues (Figure 1). This fragment binds Zn(II) or Cd(II) tightly and is capable of specific recognition of the UAS_G sequence, although the overall binding affinity for specific DNA is reduced.

Our earlier GAL4(74+2) construct produced an insoluble protein (Pan & Coleman, 1989). We have been able to construct a soluble fusion product of GAL4(1–74) combined

at the C-terminus with the A-domain of T4 gene 32 protein (residues 264–301). The latter is highly negatively charged. This fusion protein is soluble and when purified from *E. coli* binds specifically in the same manner as does GAL4(149*) to the 17-mer DNA sequence from UAS_G as assayed by a gel retardation assay (data not shown). The binding affinity is rather similar to GAL4(63), i.e., considerably less compared to GAL4(149*) or GAL4(123). It is tempting to suggest that GAL4(63) resembles an isolated DNA binding domain such as the N-terminal 92 residues of λcI repressor which binds DNA specifically but requires higher concentration, since this peptide lacks the ability to form dimers (Pabo et al., 1979; Johnson et al., 1979). The additional residues, 64–123, could easily provide an interface for monomer–monomer interactions, which is crucial for high-affinity DNA binding. On the other hand, the ^1H NMR spectra (Figure 6) suggest that both GAL4(63) and GAL4(62*) are monomeric, while GAL4(123) and GAL4(149*) are both multimeric at ^1H NMR concentrations of protein (200–600 μM), as shown by broad ^1H NMR lines. The enhanced binding of the biotin–avidin cross-linked dimer supports the postulate that a dimer of the GAL4 DNA binding domain is required for tight binding of the protein to the UAS_G sequence (Figure 1B).

GAL4 Zn(II) Binding Site Does Not Resemble the TFIIIA-Type Zn-Finger. On the basis of the arrangement of Cys residues within the DNA binding domain, Johnston (1987) proposed a TFIIIA-like Zn-finger structure for GAL4 utilizing four of the six conserved Cys as ligands for a single metal binding site. Although ^{113}Cd NMR of GAL4(149*) clearly indicated the presence of two potential metal binding sites, each with coordination to at least three sulfur atoms, it remained ambiguous whether the additional sulfur ligands were the thioethers of Met or shared $-S^-$ from Cys (Pan & Coleman, 1989). ^{113}Cd NMR of GAL4(63) and GAL4(62*) reveals that the N-terminal 63 or 62 amino acids are sufficient for formation of both metal binding sites (Figures 2 and 3). Since both signals have an intensity equivalent to one ^{113}Cd ion, each monomer contains two Cd(II), also shown by atomic absorption analysis for Cd (Table I). A metal-linked dimer cannot provide additional $-S^-$ ligands per site. In addition, the ^1H NMR of Cd(II)-GAL4(63) does not suggest the formation of a dimer or significant changes in conformation from that of the Zn(II) protein. The nearly identical ^{113}Cd NMR spectra for GAL4(63), GAL4(62*), and GAL4(149*) strongly support the postulate that the N-terminal 63 amino acid residues contain a metal binding subdomain of the same conformation as in the larger fragments of GAL4.

In previous work we had to clone the N-terminal 147 amino acids before a soluble protein was overproduced in *E. coli*; 74+2 or 92+1 residue constructs were both insoluble (Pan & Coleman, 1989). On the other hand the ability to produce a soluble GAL4(63) by tryptic proteolysis suggested that this may be a natural subdomain of the protein. This appears to be confirmed by the fact that a cloned and overproduced GAL4(62*) in *E. coli* is folded and soluble, even though the GAL4(74+2) with 14 additional C-terminal residues is not (Pan and Coleman, unpublished data).

Metal binding sites containing four sulfur ligands normally have higher affinity for Cd(II) than Zn(II). Cd(II) does readily replace Zn(II) in GAL4(63) when incubated at about equal concentrations with GAL4(63) (Table I). We believe that both sites possess similar binding affinity for Cd(II), while one site binds Zn(II) more strongly than the other. This would explain the fact that isolated GAL4(63) or GAL(149*) contains 1.0–1.5 mol/mol of bound Zn(II), while both Cd(II) ions

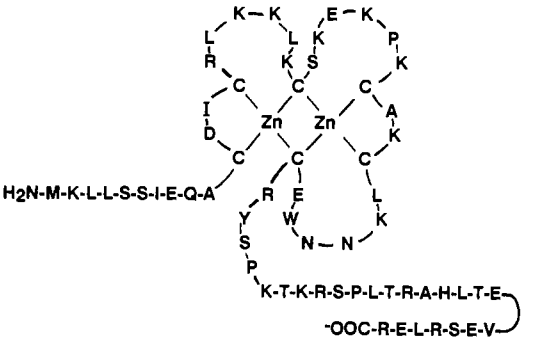

FIGURE 8: Proposed binuclear metal complex formed by the GAL4(63) protein.

seem to be bound simultaneously and cooperatively (Table I). A binuclear site can be constructed by use of all six Cys to form a two-metal cluster as shown in Figure 8. While we have been unable to support this model with the finding of ^{113}Cd–^{113}Cd J coupling (Figure 3), there appear to be structural reasons why such coupling is not observed. The expected pattern of ^{113}Cd–β^1H coupling for the β protons of Cys residues shared between two $^{113}Cd(II)$ ions is observed for two of the six Cys residues in the 2D ^1H COSY of GAL4(62*) (Pan and Coleman, unpublished results). Bridging Cys ligands forming a multinuclear metal ion cluster in a protein have been found previously in the three and four metal ion clusters of metallothionein, a protein that also binds both Zn(II) and Cd(II) (Kagi & Kojima, 1987). A binuclear cluster of this type would be expected to bind Cd(II) more tightly than Zn(II), and this appears to be true for the GAL4 site (Table I). Two Zn(II), however, are bound at greater than micromolar concentrations of free Zn(II) (Table I). There may be a difference in the affinity of the one-Zn(II) and two-Zn(II) derivatives of GAL4 for the specific UAS_G sequence. We are currently working on a binding assay accurate enough to measure such a difference.

We have built a molecular model of the proposed structure and find that the six Cys can be brought together to form the binuclear metal ion cluster without obvious steric problems. The Chou–Fasman prediction shows that the N-terminal 10 residues and the C-terminal 15 residues are predicted to form an α-helix, which would account for a large part of the α-helical content indicated by the CD spectrum (Figure 7). The amino acid sequence between the Cys ligands contains most of the turns predicted by the Chou–Fasman procedure.

Two other transcription factors from yeast, LAC9 and PPR1, contain similar clusters of six Cys residues as found in GAL4, although the sequences are located in somewhat different relative positions within the N-terminal region of these proteins (Salmeron & Johnston, 1986). While there are a number of conservative amino acid changes in the amino acid residues between the Cys, the spacing between the Cys residues is conserved, and each sequence can form the cloverleaf arrangement shown in Figure 8. It is not obvious whether mutation of a single Cys would destroy the structure of the metal complex or simply alter it. Thus, the effects of such mutations on the metal ion complex and in turn on DNA binding are not predictable. A correlation between changes in the metal complex caused by mutation and the resultant changes in DNA binding is currently under investigation.

A number of mammalian transcription factors, most notably the steroid receptor proteins, contain within their DNA binding domains clusters of Cys residues rather similar to those found in GAL4 with the exception that two copies of the cluster sequence are present (Evans & Hollenberg, 1988). Of the 10

Cys residues present in the human glucocorticoid receptor proteins, however, one in the first cluster is not always conserved among the larger family of transcription factors containing two putative Cys_2Cys_2 zinc-fingers [for review, see Beato (1989)]. Whether the type of complex that can be formed by the yeast transcription factors is a general motif among this group or is unique to a certain subset requires further investigation.

Secondary Structure of GAL4(63). The relatively large molar ellipticity of Zn(II)-GAL4(63) suggests substantial secondary structure (Figure 7). In contrast to GAL4(149*) which shows relatively minor changes in CD spectra upon removal of Zn(II) (Pan & Coleman, 1989), the effect of Zn(II) removal from GAL4(63) is more significant (Figure 6). Apo-GAL4(149*) can be reconstituted with Zn(II), most of its specific DNA binding affinity being regained at concentrations as low as 10 μM (Pan & Coleman, 1989). This does not hold for GAL4(63). Apparently, the unfolding upon removal of Zn(II) is irreversible at least at concentrations comparable to those of GAL4(149*). Possibly the refolding of apo-GAL4(63) is accompanied by an unfavorably large activation energy.

Although GAL4(63) unfolds substantially upon Zn(II) removal, the tightly bound Zn(II) can nevertheless be replaced by Cd(II) (Figure 2, Table I). From our ^{113}Cd NMR data and the fact that once denatured the apoprotein is not easily reconstitutable by Zn(II) addition, we suggest that the apoprotein is not an intermediate in the Cd–Zn exchange process. Since Cd(II) can fill the second site with high affinity as opposed to Zn(II), a Cd(II)–Zn(II) hybrid may be a likely intermediate in this type of exchange. The exchange experiments utilizing ^{113}Cd NMR as the probe show that it is possible to prepare such a hybrid (Figure 5C). Since the ^{113}Cd resonance at 706 ppm remains narrow in the "hybrid" of Figure 5C, it is unlikely that the disappearance of the 669 ppm signal is due to exchange modulation rather than metal ion replacement. It is true, however, that these small DNA binding fragments of GAL4 can undergo a conformational flux that does result in exchange broadening of both ^{113}Cd resonances (Figure 4).

There are now DNA sequences reported for a total of nine fungal transcription factors which contain a single copy of the GAL4 DNA binding motif. All nine conserve the six Cys residues found in GAL4 if adjustments for one apparent deletion of one residue and two insertions of two and three residues, respectively, are made in three of the factors. This conservation of all six Cys residues is not incompatible with all six sulfhydryls being involved in the formation of the metal ion complex as in GAL4. GAL4(62*) is soluble to millimolar concentrations, and the proton NMR spectrum is of sufficiently high resolution to undertake a 3D structural determination by 2D NMR techniques. A structure of this metal-containing, DNA binding subdomain should provide some clues as to how GAL4 and related transcription factors recognize DNA.

Acknowledgments

We thank Kasim Mookhtiar for many helpful suggestions.

Registry No. Cys, 52-90-4; Zn, 7440-66-6; Cd, 7440-43-9.

References

Armitage, I. M., & Otvos, J. D. (1982) *Biol. Magn. Reson. 4*, 79.

Beato, M. (1989) *Cell 56*, 335.

Evans, R. M., & Hollenberg, S. M. (1988) *Cell 52*, 1.

Giedroc, D. P., Johnson, B. A., Armitage, I. M., & Coleman, J. E. (1989) *Biochemistry 28*, 2410.

Greenfield, N., & Fasman, G. D. (1969) *Biochemistry 8*, 4108.

Johnson, A. D., Meyer, B. J., & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5061.

Johnston, M. (1987) *Nature (London) 328*, 353.

Johnston, M., & Dover, J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2401.

Johnson, M., & Dover, J. (1988) *Genetics 120*, 63.

Kagi, J. H. R., & Kojima, Y. (1987) *Metallothionein II*, Proceedings of the 2nd International Meeting on Metallothionein, Birkhauser Verlag, Basel.

Keegan, L., Gill, G., & Ptashne, M. (1986) *Science 231*, 699.

Miller, J., Mclachlan, A. P., & Klug, A. (1985) *EMBO J. 4*, 1609.

Pabo, C. O., Sauer, R. T., Sturtevant, J. M., & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 1608.

Pan, T., & Coleman, J. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3145.

Salmeron, J. M., Jr., & Johnston, S. A. (1986) *Nucleic Acids Res. 14*, 7767.