Racker, E. (1987) *Chem. Scr.* 27B, 131-135.
Rance, M., & Byrd, R. A. (1983) *J. Magn. Reson. 52*, 221-240.
Rance, M., Jeffrey, K. R., Tulloch, A. P., Butler, K. W., & Smith, I. C. P. (1980) *Biochim. Biophys. Acta 600*, 245-262.
Salmon, A., Dodd, S. W., Williams, G. D., Beach, J. M., & Brown, M. F. (1987) *J. Am. Chem. Soc. 109*, 2600-2609.
Seelig, A., & Seelig, J. (1974) *Biochemistry 13*, 4839-4845.
Seelig, A., & Seelig, J. (1977) *Biochemistry 16*, 45-50.
Seelig, J. (1978) *Biochim. Biophys. Acta 575*, 105-140.
Seelig, J., & Browning, J. L. (1978) *FEBS Lett. 92*, 41-44.
Seelig, J., & Waespe-Šarčevič, N. (1978) *Biochemistry 17*, 3310-3315.
Seelig, J., Tamm, L., Hymel, L., & Fleischer, S. (1981) *Biochemistry 20*, 3922-3932.
Smith, I. C. P. (1985) in *Nuclear Magnetic Resonance of Liquid Crystals* (Emsley, J. W., Ed.) pp 533-566, D. Reidel Publishing Co., Dordrecht, Holland.
Stinson, S. C. (1989) *Chem. Eng. News 67* (42), 37-70.
Straume, M., & Litman, B. J. (1987a) *Biochemistry 26*, 5113-5120.
Straume, M., & Litman, B. J. (1987b) *Biochemistry 26*, 5121-5126.
Taber, D. F., Phillips, M. A., & Hubbard, W. C. (1982) *Methods Enzymol. 86*, 366-369.
Thurmond, R. L., Dodd, S. W., & Brown, M. F. (1991) *Biophys. J. 59*, 108-113.
Wiedmann, T. S., Pates, R. D., Beach, J. M., Salmon, A., & Brown, M. F. (1988) *Biochemistry 27*, 6469-6474.
Williams, G. D., Beach, J. M., Dodd, S. W., & Brown, M. F. (1985) *J. Am. Chem. Soc. 107*, 6868-6873.

Sequential Assignments of the ^1H NMR Resonances of $Zn(II)_2$ and $^{113}Cd(II)_2$ Derivatives of the DNA-Binding Domain of the GAL4 Transcription Factor Reveal a Novel Structural Motif for Specific DNA Recognition†

Tao Pan‡ and Joseph E. Coleman*
Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510
Received August 8, 1990; Revised Manuscript Received December 10, 1990

ABSTRACT: The DNA-binding domain of the GAL4 transcription factor, consisting of the 62 N-terminal amino acid residues and denoted GAL4(62*), contains a novel $Zn(II)_2Cys_6$ or $Cd(II)_2Cys_6$ binuclear cluster [Pan, T., & Coleman, J. E. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2077]. Specific DNA recognition requires residues located within as well as C terminal to this binuclear cluster. ^1H NMR sequential assignments have been carried out on $Zn(II)_2$- and $^{113}Cd(II)_2$GAL4(62*) by using DQF-COSY, relayed COSY, double-relayed COSY, and NOESY. The ligands of the two tetrahedral metal-binding sites have been identified as Cys^{11}, Cys^{14}, Cys^{21}, and Cys^{31} to one metal ion and Cys^{28}, Cys^{38}, Cys^{21}, and Cys^{31} to the other metal ion with Cys^{21} and Cys^{31} as ligands shared between the two metal ions. No α-helices can be found within the GAL4(62*) structure, which consists of a series of turns to accommodate the metal cluster, followed by irregular loops and turns from residues 42 to 60, the "specificity region", whose sequence contributes importantly to specific DNA recognition. Long-distance NOE's are observed between residues forming the binuclear cluster and several residues within the specificity region, indicating that the latter is folded compactly onto the metal cluster. The requirement of the $Zn(II)_2Cys_6$ binuclear cluster and the specificity region for binding to DNA reveals GAL4 as a member of a class of specific DNA-binding proteins using a new structural motif for the recognition of specific DNA sequences. Specific DNA binding by this class of proteins is achieved by use of turns and loops that enclose a $Zn(II)_2Cys_6$ binuclear cluster, instead of α-helices or β-strands as observed in specific DNA-binding proteins described previously.

The GAL4 protein is a transcription factor from *Saccharomyces cerevisiae* required for the transcriptional activation of the genes encoding the galactose-metabolizing enzymes [for a review, see Johnston (1987a)]. Although the intact GAL4 protein consists of 881 amino acid residues, its DNA-binding domain has been located within the 62 N-terminal residues by limited proteolysis (Pan & Coleman, 1990a,b). The DNA-binding domain of GAL4 contains a Cys^{11}-X_2-Cys^{14}-X_6-Cys^{21}-X_6-Cys^{28}-X_2-Cys^{31}-X_6-Cys^{38} motif, conserved among a group of eleven fungal transcription factors (Andre, 1990). The cysteine cluster in GAL4 has been proposed (Johnston, 1987b) and was shown to bind Zn(II) (Pan & Coleman, 1989). The binding of Zn(II) is required for specific DNA recognition (Pan & Coleman, 1989). Cd(II) can also efficiently restore specific DNA binding to the apo GAL4 protein (Pan & Coleman, 1989). Contrary to prediction, two rather than one Zn(II) have been found to bind tightly to the DNA-binding domain of GAL4 (Pan & Coleman, 1989, 1990a). Both ^{113}Cd NMR and phase-sensitive ^1H COSY show unambiguously that only the six highly conserved cysteine residues act as ligands to Zn(II) or Cd(II) (Pan & Coleman, 1990a,b). ^1H-^{113}Cd coupling patterns have identified two cysteine residues in which the $-S^-$ is shared between the two bound metal ions (Pan & Coleman, 1990b).

On the basis of these observations, we have proposed that GAL4 forms a $Zn(II)_2Cys_6$ or $Cd(II)_2Cys_6$ binuclear cluster

† This work was supported by NIH Grants DK09070 and GM21919. The 500-MHz NMR facility was supported by NIH Grant RR03475, NSF Grant DMB-8610557, and ACS Grant RD259.
* To whom correspondence should be addressed.
‡ Present Address: Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309.

0006-2960/91/0430-4212$02.50/0

(Pan & Coleman, 1990a,b). In the previous work the two bridging ligands had tentatively been identified as Cys^{21} and Cys^{38} on the basis of a preliminary sequential assignment. Such a model of the binuclear cluster placed the preponderance of positive charges on one side of the cluster formed by two loops, Cys^{14}–Cys^{21} and Cys^{21}–Cys^{28}. The positive charge distribution between Cys^{14} and Cys^{21} is maintained in all eleven fungal transcription factors of this class. This suggests that the residues within the cluster (Cys^{11}–Cys^{38}) may function primarily in nonspecific DNA binding and that amino acid residues outside the cluster may be more important in specific DNA sequence recognition. In fact, both point mutations (Johnston & Dover, 1987, 1988) and segment-swap experiments (Corton & Johnston, 1989) indicate that amino acid residues in the sequence immediately C terminal to the binuclear cluster are essential for specific DNA recognition. On the other hand, mutations within the cluster as well as C terminal to the cluster can affect the affinity of GAL4 for DNA.

We report here the sequential assignments of the ^1H NMR resonances of $Zn(II)_2GAL4(62^*)$ and its Cd(II) derivative, $^{113}Cd(II)_2GAL4(62^*)$, as an essential step toward the determination of the solution structure of the DNA-binding domain of GAL4 by 2D NMR methods. Zn(II) and Cd(II) differ in their ionic radii as reflected in the different lengths of their metal–sulfur bonds. Although the chemical shifts of GAL4(62*) are sensitive to the metal ion species bound, both $Zn(II)_2$- and $^{113}Cd(II)_2GAL4(62^*)$ recognize the specific UAS_G DNA sequence. In contrast to the prediction made from circular dichroic spectra, GAL4(62*) does not contain any α-helices. The majority of the residues are found to form various kinds of loops and turns. Thus GAL4 appears to represent a new class of specific DNA-binding proteins.

Materials and Methods

Preparation of $Zn(II)_2$ and $^{113}Cd(II)_2GAL4(62^*)$. Cloning, overexpression, and purification of GAL4(62*) from *E. coli* have been described previously (Pan & Coleman, 1990b). GAL4(62*) consists of the natural 61 N-terminal residues of GAL4, while residue 62 has been changed from Glu to Asp in this construct. $Zn(II)_2GAL4(62^*)$ was obtained by dialysis of 2 mM Zn(II) protein against 4 mM $ZnCl_2$, pH 5.4. $^{113}Cd(II)_2GAL4(62^*)$ was prepared by exchange of Zn(II) for $^{113}Cd(II)$ as described earlier (Pan & Coleman, 1990b). The pH was adjusted to 5.0 by the addition of small aliquots of HCl. Both $Zn(II)_2$- and $^{113}Cd(II)_2GAL4(62^*)$ were then lyophilized and redissolved in 90% H_2O/10% D_2O. The 1D ^1H NMR spectra of the samples prepared before and after lyophilization are essentially identical (data not shown). We have collected 2D ^1H NMR spectra and ^{113}Cd NMR spectra on GAL4 derivatives at pH values that range from 8 to 5. GAL4(62*) is stable over this range as judged by the similarity of both the 2D ^1H COSY spectra as well as the identical chemical shifts of the ^{113}Cd resonances; the latter are very sensitive to conformational changes even outside the immediate vicinity of the coordinating residues. We have chosen to use samples at pH 5.0 for the $^{113}Cd(II)$ and pH 5.4 for the Zn(II) derivatives. As in much previous 2D NMR work with proteins, the slower amide proton exchange is an essential requirement for a protein like GALA4(62*) and avoids low-temperature experiments. We have previously published a number of the 2D ^1H COSY spectra taken at pH 7–8 (Pan & Coleman, 1990b), and they are very comparable to those presented in this work; e.g., the ^1H–^{113}Cd heteronuclear coupling patterns and coupling constants shown in this work at the low pH (see Figure 1) are entirely comparable to those shown in Pan and

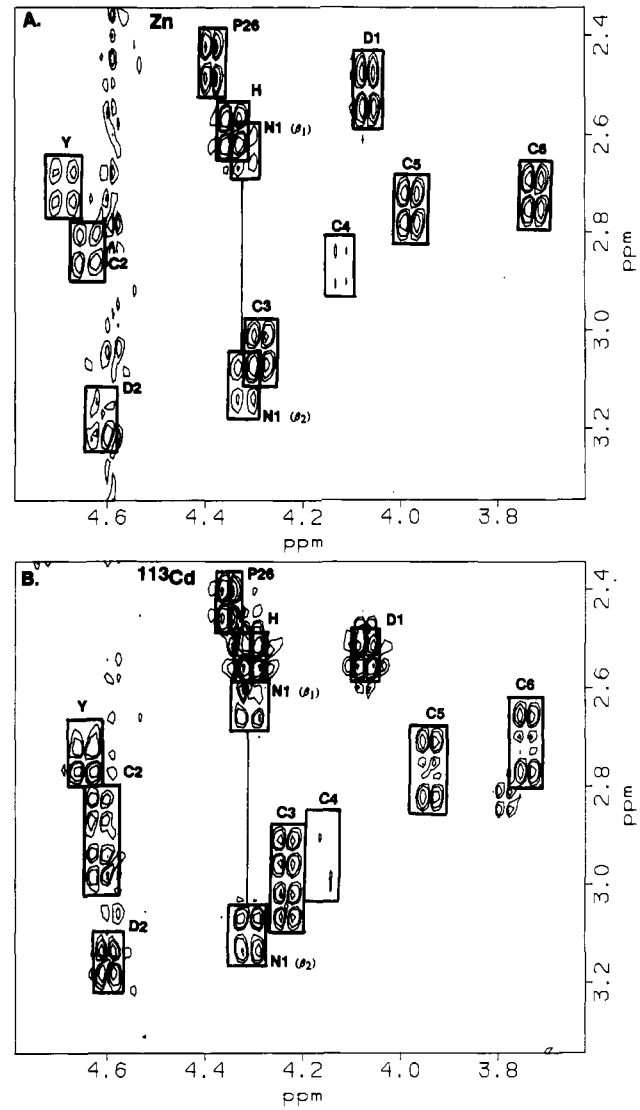

FIGURE 1: Phase-sensitive DQF ^1H COSY of $Zn(II)_2GAL4(62^*)$ (A) and $^{113}Cd(II)GAL4(62^*)$ (B) showing the $\alpha\beta$ cross-peaks of several AMX spin systems. The additional splittings of the cysteine $\alpha\beta$ cross-peaks in B are caused by ^1H–^{113}Cd heteronuclear J coupling.

Coleman (1990b). The concentrations of the $Zn(II)_2$- and $^{113}Cd(II)_2GAL4(62^*)$ were 3 and 15 mM, respectively. No obvious line broadening of $Zn(II)_2$- or $Cd(II)_2GAL4(62^*)$ was observed at all concentrations up to 15 mM, indicating that GAL4(62*) remains a monomer under the conditions used for NMR. All samples were stored under nitrogen to prevent oxidation of the sulfhydryl groups.

^1H NMR. All ^1H NMR spectra were obtained on a Bruker AM-500 spectrometer at 35 °C. The data were transferred by magnetic tape to a microVAX computer for processing with the FTNMR software by Dennis Hare. Chemical shifts are referred to H_2O and HDO, which at 35 °C resonates at 4.60 ppm relative to sodium (trimethylsilyl)tetradeuteriopropionate (TSP).

All COSY and NOESY experiments were recorded in the phase-sensitive mode with standard automation programs provided by Bruker. The carrier was placed in the center of the spectrum on the water signals, and a recycle delay of 1.5–2.0 s was used for all experiments. The relayed and double-relayed COSY spectra were recorded in the absolute-value mode. During the recycle delay and the mixing time, the water signal was supressed by continuous low-power irradiation from the proton decoupler channel. The spectral

width in F_2 was 5681.8 Hz, and 512 increments with 64 scans each were collected with 2048 data points in a typical experiment. The F_1 domain was subsequently zero-filled to 1024 points to provide 1024 × 1024 final matrices. Fourier transformation in both dimensions was performed after multiplication by a skewed sine-bell squared-window function with a 0° phase shift.

Results

Zn(II) and Cd(II)GAL4 bind specific DNA with equal affinity (Pan & Coleman, 1989). For GAL4(62*), $Zn(II)_2$ and $Cd(II)_2$ species can be prepared with ease, and their 1D spectra show very similar characteristics (Pan & Coleman, 1990b). Sequential assignments were completed for both species in order to detect any significant difference in structure between the derivatives. The two data sets also provide independent checks on the assignment. Because of the different ionic radii [0.74 and 0.97 Å for Zn(II) and Cd(II), respectively] and metal–sulfur bond lengths [av 2.33 Å for Zn(II) and 2.50 Å for Cd(II)], chemical shifts of the amino acid residues located within the binuclear cluster are likely to differ. In addition, residues C or N terminal to the cluster that interact with residues within the cluster may also show diverse chemical shifts between the two metal-ion species. These variations in chemical shift with metal-ion species also point to amino acid residues outside the cluster that interact with the binuclear metal cluster.

The assignment strategy described by Wuthrich (1986) was applied for $Zn(II)_2$- and $^{113}Cd(II)_2$GAL4(62*). First, the resonances were categorized as to residue type from phase-sensitive COSY, relayed, and double-relayed COSY. Sequential assignments were then made on the basis of the sequential interresidual αH–NH, NH–NH, or βH–NH NOE's. GAL4(62*) has an abundance of Lys and Arg residues, 16 out of the 62 residues. At the initial stage, significant overlap of some of the αH–NH cross-peaks imposed difficulties for the unambiguous assignment of all observable cross-peaks in the αH–NH "fingerprint" region (see below). Nevertheless, using the spectra from both $Zn(II)_2$GAL4(62*) and $^{113}Cd(II)_2$GAL4(62*), it was possible to make sequential assignments of all NH, αH, and in most cases βH resonances of all 62 residues.

Spin System Identification. Resonances of cysteine residues were identified from comparison of phase-sensitive COSY spectra of $^{112}Cd_2$- and $^{113}Cd_2$GAL4(62*) as described previously (Pan & Coleman, 1990b). For $Zn(II)_2$GAL4(62*), it was assumed that the cross-peak patterns and positions for cysteine residues do not differ significantly from $Cd(II)_2$ species, and assignment of the αH's or βH's of the Cys residues from phase-sensitive COSY was then relatively straightforward as shown for the αβ cross-peaks of five of the six Cys residues in Figure 1A.

The assignment of the Cys residues is greatly helped by the pattern of ^1H–^{113}Cd heteronuclear coupling observed for the ^{113}Cd derivative (Pan & Coleman, 1990b). While this has been discussed in detail previously, an example is given in Figure 1B to illustrate this additional feature possible in the ^1H NMR assignment of a Zn(II) metalloprotein. The assignment of these resonances to cysteines was also confirmed upon completion of sequential assignments (see below). The unique Trp^{36} and Tyr^{40} can be assigned through their AMX spin systems and intraresidual strong NOE's between the protons of the aromatic ring and the β protons. Similarly, both Asn β protons were assigned upon finding the NOE's between the β- and γ-NH protons. β protons of one of the asparagines have unusual upfield chemical shifts caused by ring current

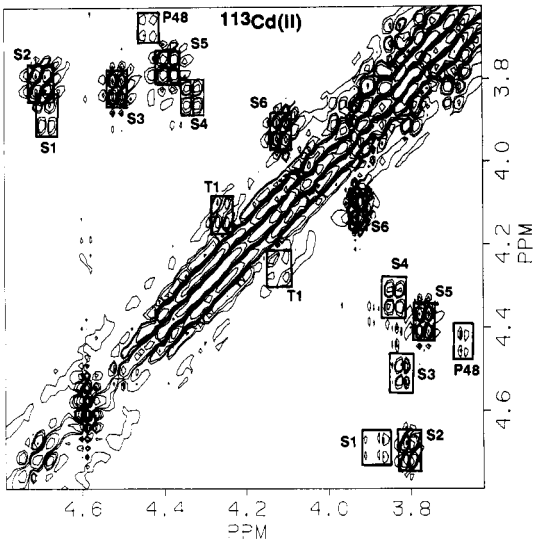

FIGURE 2: Phase-sensitive DQF ^1H COSY of $^{113}Cd(II)_2$GAL4(62*) showing the αβ cross-peaks of six Ser and one Thr residues.

shifts induced by Trp^{36}. The remaining αβ cross-peaks in the region shown in Figure 1 are from 2 Asp, 1 His, and 1 Pro residue.

The β protons of Ser were assigned on the basis of their chemical shifts, which are expected to be around 3.80 ppm (Figure 2). Since GAL4(62*) does not contain any Gly residues, all of the αβ cross-peaks within this region are either Ser or Thr αβ cross-peaks (Figure 2). One cross-peak is unaccounted for, and this appears to be the αβ cross-peak of Pro^{48} based on the complete sequential assignment (see below).

The αH–NH fingerprint regions of $Zn(II)_2$- and $^{113}Cd(II)_2$GAL4(62*) are shown in Figure 3. The spin-system assignments of the αH–NH cross-peaks are based on relayed and double-relayed COSY of the same samples in H_2O. One example of a relayed COSY spectrum is shown in Figure 4 where nearly all the backbone NH chemical shifts of the AMX systems can be identified. Similarly, assignments of the other spin systems were carried out by combining the information obtained from the COSY, the relayed, or the double-relayed COSY spectra.

Sequential Assignments. The NOESY spectra shown in Figures 5–8 were taken with a mixing time of 250 ms. We have also taken NOESY spectra with a 100-ms mixing time, and the data obtained were consistent with the assignments shown in Figures 5–8. The strategy for the sequential assignment of GAL4(62*) can be divided into three phases. Within the $Zn(II)_2Cys_6$ binuclear cluster (residues Gln^9–Ser^{41}), most of the sequential NOE's can be easily located. The unbroken sequential NOE connectivities for the region C terminal to the binuclear cluster (Pro^{42}–Asp^{62}) are much shorter, and assignment has to be based on combinations of αH–NH and βH–NH connectivities. The region N terminal to the binuclear cluster (Met^1–Glu^8) appears to be very flexible in solution, and assignments are based mostly on exclusion of the spin systems, since sequential NOE's are missing. Comparison of the chemical shifts between $Zn(II)_2$- and $^{113}Cd(II)_2$GAL4(62*) is extremely useful in assignments of all three parts of the molecule. Significant shifts of specific resonances between the two derivatives, while the connectivity pattern is maintained, help clarify a number of ambiguities (see Figure 5).

The Binuclear Cluster (Gln^9–Ser^{41}). When the spin systems of the unique Trp^{36} and the adjacent Asn^{34} and Asn^{35} were assigned, the sequential NOE's between these three adjacent

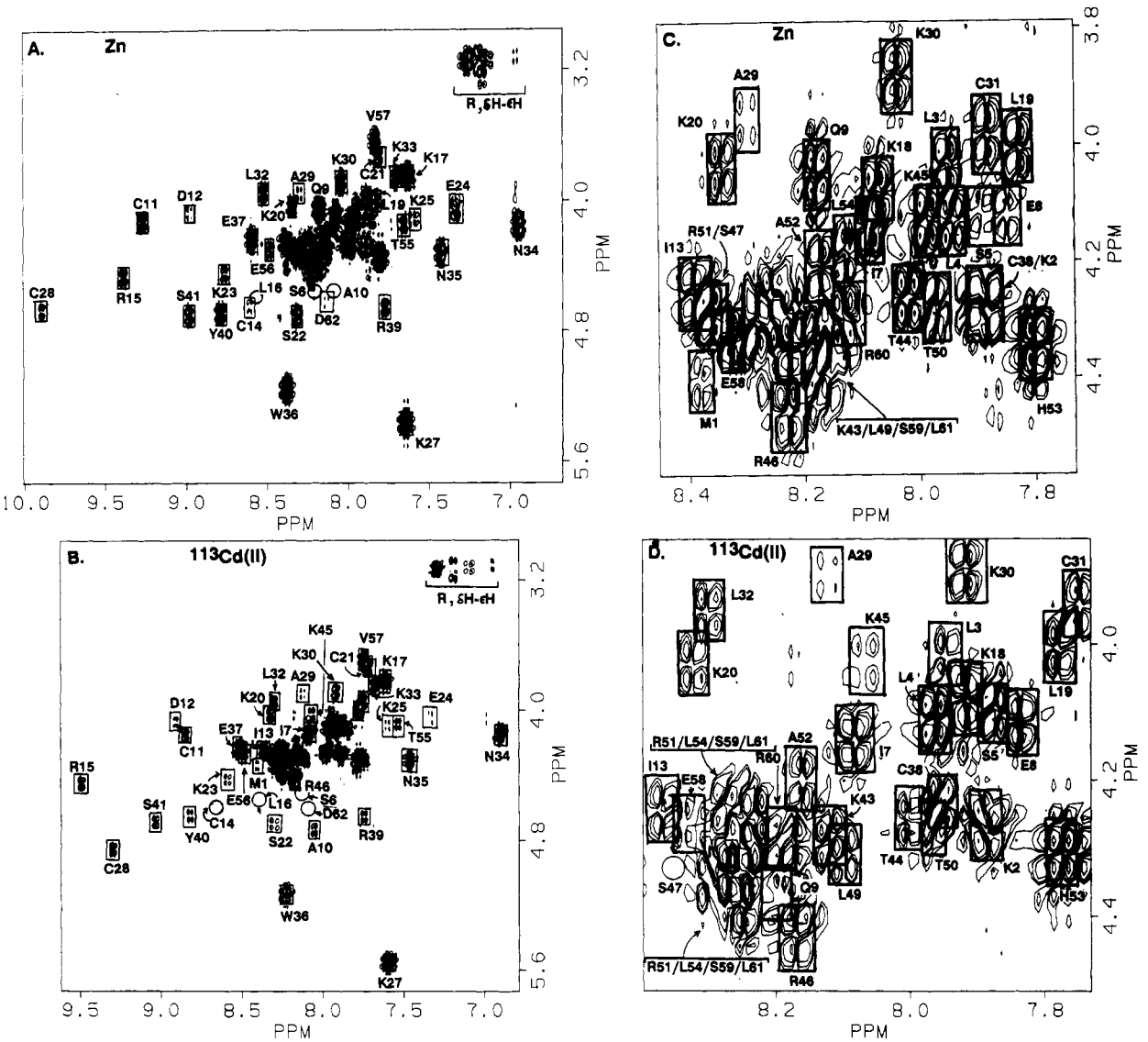

FIGURE 3: Phase-sensitive DQF ^{1}H COSY of $Zn(II)_2$- (A) and $Cd(II)_2$GAL4(62*) (B) of the αN "fingerprint" regions. The heavily overlapped regions are shown in the enlarged spectra for $Zn(II)_2$- (C) and $Cd(II)_2$GAL4(62*) (D).

residues were easily identified [Figures 5 and 6 showing αN-(i,i+1) and NN(i,i+1) NOE's, respectively]. Sequential NOE connectivities, αN(i,i+1), can be extended to Pro^{26} and Arg^{39} (Figure 5) and to Cys^{31} by the use of the NN(i,i+1) NOE's (Figure 6). At the same time, βN(i,i+1) NOE's connect all the residues from Pro^{26} to Ser^{41} (Figure 7). These sequential NOE links complete sequence-specific assignments for three cysteine ligands, Cys 1, Cys 5, and Cys 3 [designation used in Pan and Coleman (1990b)], as Cys^{28}, Cys^{31}, and Cys^{38}, respectively. Since Cys 5 is one of the two shared cysteine residues (Pan & Coleman, 1990a), one of the shared cysteine ligands is Cys^{31}. A second stretch of NN(i,i+1) NOE connectivities assigns six sequential residues (Figure 6). One of them is assigned to Cys 6 on the basis of the corresponding chemical shift changes in the $Zn(II)_2$- and $^{113}Cd(II)_2$GAL4-(62*) species. This cysteine can also be connected through an αN(i,i+1) NOE to a serine residue. Since the only Cys–Ser segment within GAL4(62*) is Cys^{21}–Ser^{22}, this allows assignment of the Leu^{16}–Cys^{21} segment through sequential NN(i,i+1) NOE's. This also establishes the second shared cysteine ligand as Cys^{21} [Cys 6 in Pan and Coleman (1990b)].

The assignment of Cys^{11} (Cys 4) is based on its αN(i,i+1) NOE connections to its neighbors, an Asp, identified in Zn-$(II)_2$- and $Cd(II)_2$GAL4(62*), and an Ala residue readily identified in $^{113}Cd(II)_2$GAL4(62*) (Figure 5). Cys 2 must be Cys^{14}, since it has αN(i,i+1) and βN(i,i+1) NOE connectivities to an Ile residue that can therefore be assigned as Ile^{13}. Thus, all six cysteine ligands can be uniquely identified. Interresidual NOE's are observed between Cys^{28} and the Cys^{38}/Cys^{31} pair. Likewise Cys^{11} has NOE's with the Cys^{14}/Cys^{21} pair (Figure 8). This pattern of NOE's establishes that the metal-binding ligands occur as two groups, Cys^{11}, Cys^{14}, Cys^{21}, and Cys^{31} to one site and Cys^{28}, Cys^{38}, Cys^{21}, and Cys^{31} to the second site with Cys^{21} and Cys^{31} as shared ligands. Assignment of the remaining residues within the second loop of the cluster, Lys^{23}-Glu^{24}-Lys^{25}, was carried out at later stages. There are αN(i,i+1) NOE's between these three residues in $Zn(II)_2$GAL4(62*). The only other residue left within the cluster, Arg^{15}, was identified by its βN(i,i+1) NOE connectivity to Cys^{14} (Figure 7).

The Region C Terminal to the Binuclear Cluster (Pro^{42}–Asp^{62}). The αH–NH NOE cross-peaks for most of these residues are crowded within a much narrower range than those arising from the binuclear cluster. No sequential NN(i,i+1) NOE links can be found within this region (Figure 6). Except for the short stretch, Val^{57}–Asp^{62}, sequential NOE's do not extend beyond three residues. Assignment relies heavily on the chemical shift similarities or differences between $Zn(II)_2$-

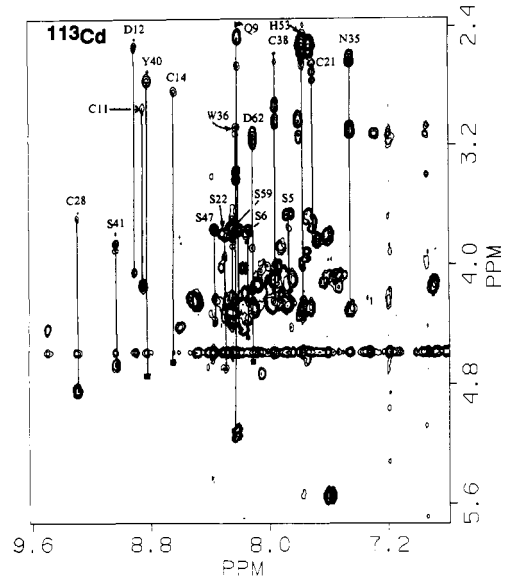

FIGURE 4: Relayed-COSY spectrum of $Cd(II)_2$ GAL4(62*) in H_2O. The mixing time was 32 ms for this spectrum. NH–βH connections to the most AMX spin systems are shown.

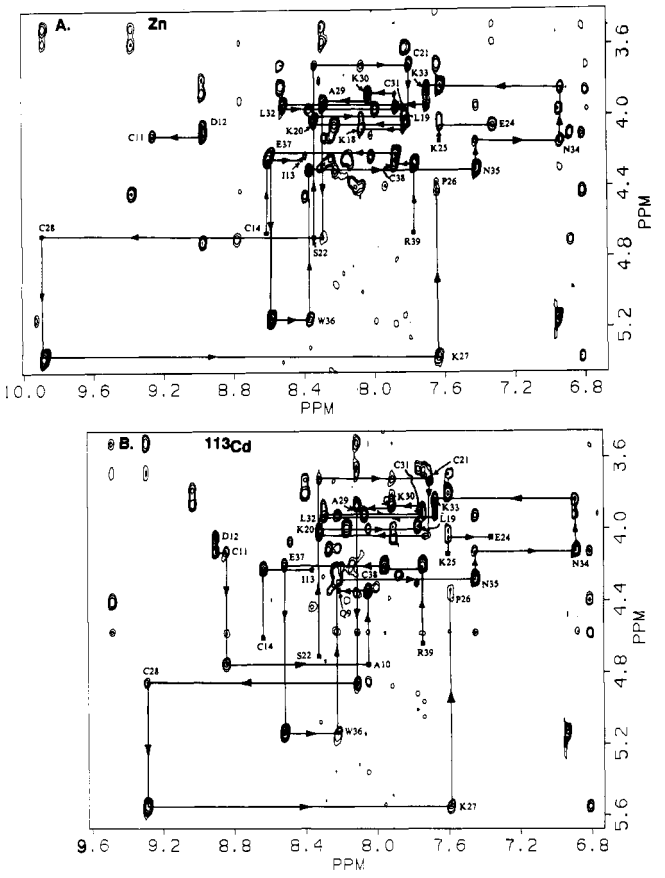

FIGURE 5: Assignment of the sequential $\alpha N(i,i+1)$ NOE contacts between residues within the binuclear cluster (Gln^9–Ser^{41}). (A) $Zn(II)_2$GAL4(62*). (B) $^{113}Cd(II)_2$GAL4(62*).

and $^{113}Cd(II)_2$GAL4(62*). Some of the NOE connectivities are shown in Figure 9. NOE connectivities, $\alpha N(i,i+1)$, are evident for Val^{57}–Asp^{62} and Lys^{45}–Ser^{47}. Assignment of the Lys^{45}–Ser^{47} sequence is also confirmed by their $\beta N(i,i+1)$ sequential NOE's (not shown). The remaining assignment is carried out from combinations of $\alpha N(i,i+1)$ and $\beta N(i,i+1)$ NOE's and the uniqueness of some spin systems, e.g., Ala^{52} and His^{53} (see Figure 10).

The N-Terminal Region of GAL4(62) (Met^1–Glu^8).* Due

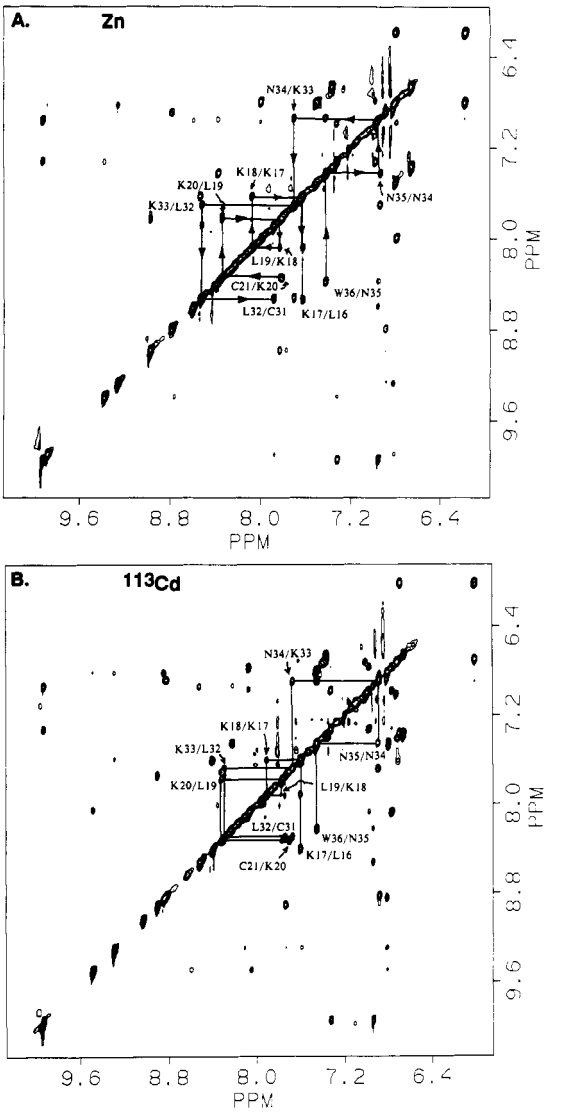

FIGURE 6: NOESY spectra of the backbone amide and aromatic region of $Zn(II)_2$- (A) and $Cd(II)_2$GAL4(62*) (B).

to the flexibility of the N-terminal eight residues in solution, very few NOE's are observed under our experimental conditions. Glu^8 gives rise to an $\alpha N(i,i+1)$ NOE to Gln^9. The rest of the residues are assigned mainly by excluding other spin systems. Ser^5 and Ser^6 were differentiated by finding a weak $\alpha N(i,i+1)$ NOE cross-peak between Ser^6 and Ile^7 in $^{113}Cd(II)_2$GAL4(62*). Separation of Leu^3 and Leu^4 is uncertain and is based on a weak $\beta N(i,i+1)$ NOE cross-peak. The findings that much less chemical shift dispersion is observed for the residues within this region and that few differences exist between the $Zn(II)_2$- and $^{113}Cd(II)_2$ species suggest that this region does not interact with the binuclear cluster (see Table I).

The Secondary Structure of GAL4(62).* A summary of the sequential $\alpha N(i,i+1)$, $NN(i,i+1)$, and $\beta N(i,i+1)$ NOE connectivities is shown in Figure 10. Chemical shift assignments of $Zn(II)_2$- and $^{113}Cd(II)_2$GAL4(62*) are listed in Table I. The complete numbers give the chemical shifts for the protons of the $Zn(II)_2$GAL4(62*), while the numbers in parentheses give the $\Delta\delta$s occurring in the chemical shifts of the same protons upon the substitution of Cd(II) for Zn(II). The $\Delta\delta$s are expressed as $\delta_{Zn} - \delta_{Cd}$, hence negative is a downfield shift, while positive is an upfield shift. Residues for which at least one NH, αH, βH, or γH proton shifts by 0.10 ppm or more consequent to the metal-ion substitution are

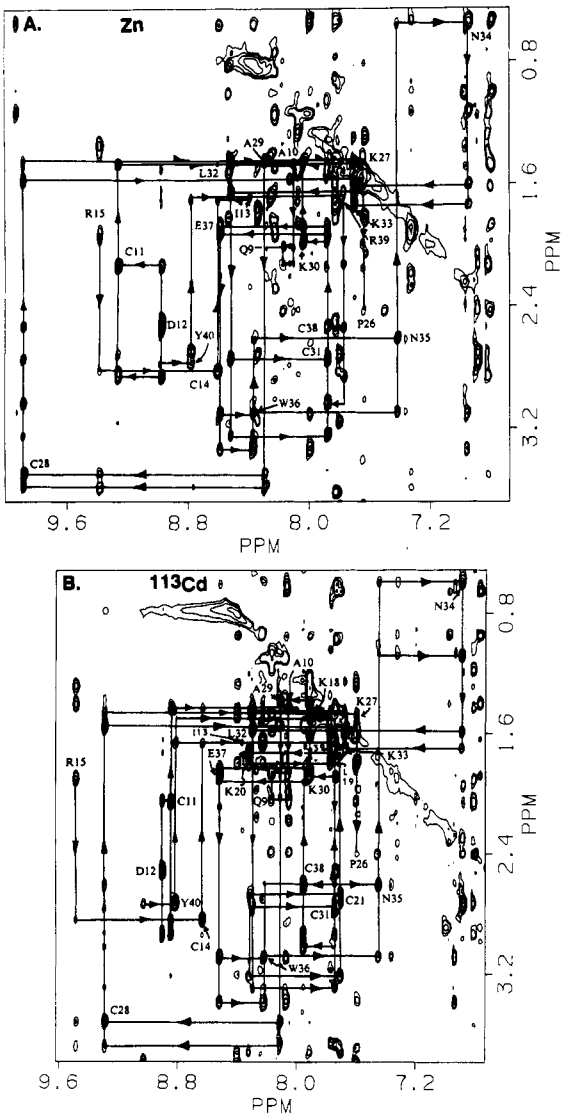

FIGURE 7: Assignment of the sequential βN(i,i+1) NOE links between residues within the binuclear cluster. $Zn(II)_2$- (A) and Cd-$(II)_2$GAL4(62*) (B). The sequential βN(i,i+1) NOE connectivities for Lys^{17}–Ser^{22} are not shown in A for clarity.

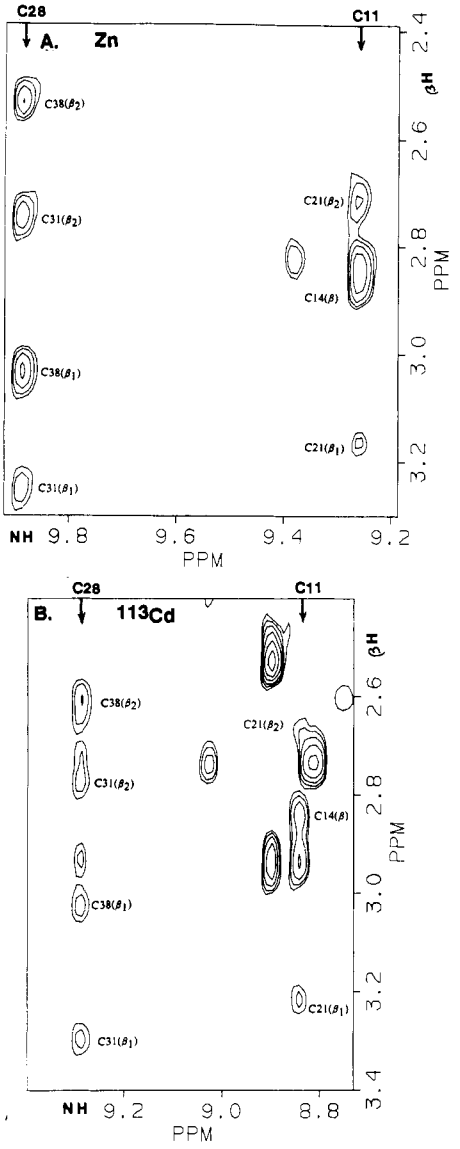

FIGURE 8: Long-range NOE's between the βH and NH of the cysteine residues establish the cysteine ligands in two groups.

in boldface in Table I. From the data shown in Figure 10, one can conclude that there are no α-helices within GAL4-(62*). α-Helices are characterized by medium NN(i,i+1) and αN(i,i+3) and very weak αN(i,i+1) NOE's. In GAL4(62*), two stretches showing medium NN(i,i+1) NOE's are observed, each connecting six residues. Among these residues, however, only one αN(i,i+3) (Cys^{31}–Asn^{34}) can be found between Cys^{31} and Trp^{36} and none in the Lys^{16}–Cys^{21} segment. In addition, medium-range αN(i,i+1) NOE's are observed between Cys^{31} and Trp^{36} and through most of the Lys^{16}–Cys^{21} segment. α-Helix formation may be possible in the much less defined region N terminal to the binuclear cluster. This region is, however, not required for specific DNA binding (see Discussion). Some β-strand structure may be present in the regions where there are strong αN(i,i+1) and no NN(i,i+1) NOE sequential connectivities. In GAL4(62*), these include Glu^8–Cys^{14} and Val^{57}–Asp^{62}.

The overall secondary structure of GAL4(62*) strongly resembles that of Cd_7-metallothionein (Wagner et al., 1986; Schultze et al., 1988). Cd_7-metallothionein does not contain regular secondary elements such as α-helices or β-strands. Rather it forms a series of turns and loops to accommodate formation of the Cd_3 and Cd_4 clusters. In order to accommodate formation of the binuclear metal cluster, the peptide loops within the "core" of GAL4(62*) must form a series of turns. The conformations of these turns, partly dictated by the formation of the metal–sulfur cluster, are probably responsible for the unusual pattern of sequential NOE's observed in GAL4(62*) from residues 8 to 42. Some of the backbone NH groups may satisfy their hydrogen-bonding potential by formation of NH–S bonds as found in the 1.8-Å crystal structure of metallothionein (Robbins et al., 1991). This will affect the pattern of sequential d_{NN} NOE's. Until a better picture of the conformation of these turns is achieved from the best possible energy-minimized structure derived from all the NOE's (sequential and long distance), one cannot define a unique secondary structure for GAL4(62*).

The eight N-terminal residues of GAL4(62*) must be disordered in the free protein. The paucity of strong sequential NOE's in the C-terminal region (residues 43–62) may be due in part to flexibility of this portion of the molecule in the absence of DNA. However, the long-range NOE's in GAL4(62*) establish a number of contacts between residues of the C-terminal region and the cluster. These include Lys^{45}–Arg^{46} and Leu^{54}–Val^{57}, residues that are folded back onto Trp^{36}, Asn^{34}, Cys^{21}, and Leu^{19} (see Discussion). Hence the C-terminal region must fold back on the metal cluster to

Table I: ^{1}H NMR Resonance Assignments for $Zn(II)_2$GAL4(62*) (35 °C, pH 5.4) and $^{113}Cd(II)_2$GAL4(62*) (35 ‰C, pH 5.0)[a]

residue[b]	NH	αH	βH	other
1 Met	8.38	4.40 (0.06)		
2 Lys	7.89	4.27	1.79, 1.63	
3 Leu	7.96	4.03	2.02 (0.08)	γ 2.01; δ 0.88, 0.82
4 Leu	7.93	4.13	1.81 (0.29)	γ 1.47; δ 0.86, 0.80
5 Ser	7.88	4.12	3.92, 3.77 (−0.10)	
6 Ser	8.16	4.52	3.84, 3.79	
7 Ile	8.09	4.14	1.79	γ 1.35, 1.08, 0.81 (0.06); δ 0.73 (0.08)
8 Glu	7.84	4.12	1.74 (0.10), 1.53	γ 2.17, 1.87 (0.06)
9 Gln	8.18	4.05 (−0.21)	2.12 (−0.12), 2.02 (0.14)	γ 2.43; δ 7.50 (0.08), 6.80
10 Ala	8.11 (0.07)	4.60 (−0.14)	1.47	
11 Cys	9.26 (0.41)	4.13	2.88 (−0.08), 2.15 (0.07)	
12 Asp	8.98 (0.08)	4.07	2.56, 2.51	
13 Ile	8.39	4.26	1.70	γ 1.44, 0.92 (−0.09), 0.78; δ 0.58
14 Cys	8.60	4.64	2.84, 2.84	
15 Arg	9.38 (−0.11)	4.46	1.98	
16 Leu	8.54 (0.14)	4.58	1.85, 1.80	γ 2.01; δ 0.88
17 Lys	7.63	3.85	2.08	
18 Lys	8.07 (0.16)	4.08	1.91 (0.40)	
19 Leu	7.82	4.01	1.75, 1.64	γ 1.50; δ 0.79
20 Lys	8.34	4.04	1.85, 1.78	
21 Cys	7.79 (0.08)	3.73	3.17 (−0.06), 2.73	
22 Ser	8.31	4.71	3.82, 3.82	
23 Lys	8.76 (0.19)	4.45	1.50	
24 Glu	7.33	4.06	1.39, 1.29 (0.09)	
25 Lys	7.60	4.11	1.50	
26 Pro		4.40 (0.06)	2.47	γ 1.83 (0.07); δ 3.72, 3.43
27 Lys	7.64 (0.05)	5.37 (−0.19)	1.58, 1.47	
28 Cys	9.89 (0.60)	4.67 (−0.19)	3.60 (0.10), 3.51	
29 Ala	8.30 (0.18)	3.95 (−0.05)	1.45	
30 Lys	8.03 (0.12)	3.88	1.96, 1.86 (0.06)	
31 Cys	7.87 (0.12)	3.98	3.25 (−0.06), 2.75	
32 Leu	8.52 (0.22)	3.95	1.66 (0.06), 1.46	δ 0.68
33 Lys	7.70	3.85	1.74, 1.59	
34 Asn	6.93	4.16	1.13, 0.56	γ 6.80 (0.10), 6.18 (0.17)
35 Asn	7.42	4.32	3.11, 2.62	γ 7.36, 6.65
36 Trp	8.37 (0.15)	5.16	3.34 (−0.05), 3.13	H1 9.94; H2 6.96; H4 8.00 (−0.08); H5 6.79; H6 6.97; H7 7.34
37 Glu	8.59 (0.08)	4.24	1.90, 1.85	γ 2.19, 2.12
38 Cys	7.89 (−0.07)	4.29 (0.06)	3.04, 2.55 (−0.08)	
39 Arg	7.77	4.66	1.66	
40 Tyr	8.78	4.69	2.77, 2.71	H2,6 6.89; H3,5 6.82
41 Ser	8.98 (−0.05)	4.71	3.89, 3.81	
42 Pro		4.49	2.28	γ 2.00; δ 3.84, 3.70
43 Lys	8.28 (0.17)	4.30	2.03 (0.26), 1.87 (0.23)	
44 Thr	8.02	4.26	4.12	γ 1.13
45 Lys	7.98 (−0.08)	4.13 (0.10)	1.53 (0.32), 1.77	
46 Arg	8.23 (0.06)	4.46	2.26 (0.08), 1.29 (0.06)	γ 1.56; δ 3.14; ε 7.26
47 Ser	8.38	4.30	3.92 (0.08), 3.78	
48 Pro		4.44	3.69	γ 1.97; δ 3.78, 3.70
49 Leu	8.18 (0.08)	4.34	1.53	γ 1.26; δ 0.87
50 Thr	7.95	4.28	4.17	γ 1.13
51 Arg	8.36 (0.08)	4.29 (−0.06)	2.01, 1.85	
52 Ala	8.17	4.21	1.28	
53 His	7.80	4.35	2.59 (0.05)	H2 8.42; H4 7.23
54 Leu	8.13 (−0.10)	4.18 (0.12)	1.66, 1.55	γ 1.25; δ 0.84
55 Thr	7.65 (0.12)	4.15 (0.06)	4.13 (0.05)	γ 1.26
56 Glu	8.48	4.30	1.92 (0.06), 1.87	γ 2.25 (0.05)
57 Val	7.83 (0.09)	3.63 (−0.06)	1.74	γ 0.53 (−0.15)
58 Glu	8.32	4.33 (0.07)	1.56, 1.49	
59 Ser	8.22	4.40	3.83 (0.07), 3.80	
60 Arg	8.13 (−0.05)	4.30	1.81 (0.11), 1.67 (0.13)	
61 Leu	8.28	4.26	1.72, 1.62	γ 1.56; δ 0.84, 0.79
62 Asp	8.13	4.62	3.15	

[a] Chemical shifts were measured relative to the water peak at 4.60 ppm; TSP, sodium (trimethylsilyl)tetradeuteriopropionate, is the chemical shift zero. The numbers in parentheses are the $\Delta\delta$s ($\delta_{Zn} - \delta_{Cd}$) when Cd(II) is substituted for Zn(II); (−) = downfield shifts. If $\Delta\delta < 0.05$ ppm, no entry is made. [b] Residues in boldface indicate those with at least one proton that alters its chemical shift by at least 0.1 ppm when Zn(II) is replaced by Cd(II).

give a reasonably compact structure.

Discussion

A stable N-terminal domain of GAL4 made up of the 63 N-terminal amino acid residues was obtained initially from the partial tryptic proteolysis of a larger fragment, GAL4-(149*) (Pan & Coleman, 1990a). This 63-residue subdomain folds independently and binds to the specific DNA sequence, UAS_G, recognized by GAL4 (Pan & Coleman, 1990a). This DNA-binding domain, like the larger constructs, binds two Zn(II) ions (Pan & Coleman, 1989; 1990a). ^{113}Cd NMR of ^{113}Cd(II)-substituted GAL4(63) initially showed two ^{113}Cd signals whose chemical shifts were consistent with ligation of three or four sulfurs to each ^{113}Cd(II) ion (Pan & Coleman, 1989, 1990a). A cloned construct consisting of the 62 N-terminal amino acid residues also produced a protein struc-

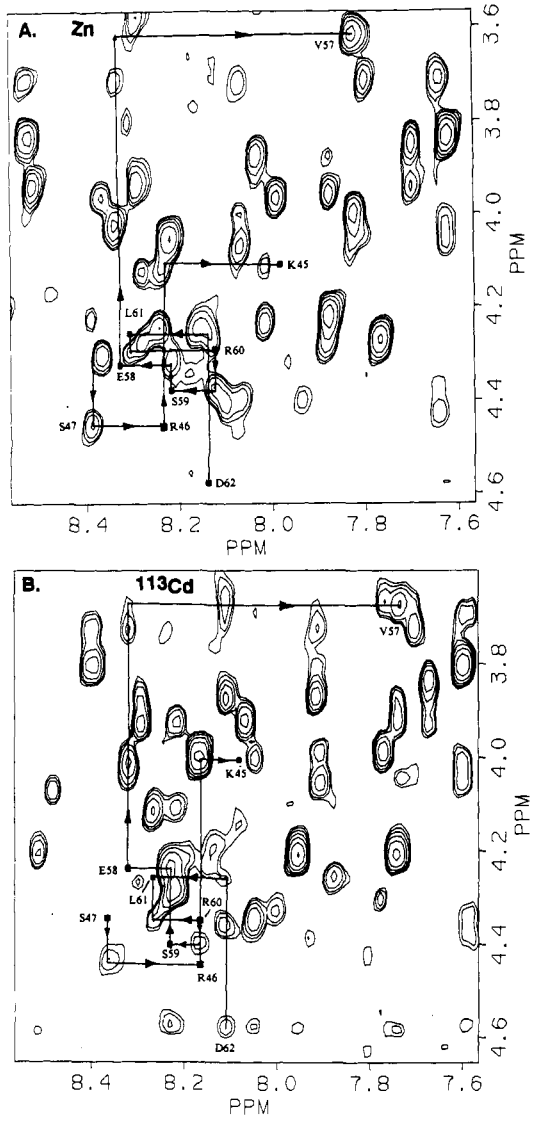

FIGURE 9: Sequential $\alpha N(i,i+1)$ connectivities of some residues from the region C terminal to the binuclear cluster. (A) $Zn(II)_2$- and (B) $Cd(II)_2GAL4(62^*)$.

turally and functionally identical with GAL4(63) whose ^{113}Cd derivative showed the same ^{113}Cd NMR spectrum. ^1H–^{113}Cd heteronuclear multiple-quantum spectroscopy and phase-sensitive ^1H COSY of $^{112}Cd_2$- and $^{113}Cd_2GAL4(62^*)$ identified six cysteine residues as ligands to both metal ions, two of which were bridging ligands between the two Cd(II) ions (Pan & Coleman, 1990b). The necessity of forming such a binuclear cluster for DNA binding would explain why all six cysteine residues are conserved among the 11 fungal transcription factors whose amino acid sequences are related to GAL4.

The $Zn(II)_2Cys_6$ Binuclear Cluster. In the previous work we were able to identify two shared cysteine ligands, designated as Cys 5 and Cys 6. The tentative partial sequential assignment we had at that time suggested that these were Cys^{38} and Cys^{21}, respectively. Our final sequential assignment presented here shows Cys 5 to be Cys^{31} rather than Cys^{38}. Thus in the actual structure the loop from Cys^{28} to Cys^{31} slips under the six-residue loop from Cys^{31} to Cys^{38} rather than the reverse as in the original model (Pan & Coleman, 1990b). The βH–NH interresidual NOE's between the quartets of cysteine ligands strongly support the division of the cysteine ligands into two groups as assigned, Cys^{11}, Cys^{14}, Cys^{21}, and Cys^{31} coordinated to one metal ion and Cys^{28}, Cys^{38}, Cys^{21}, and Cys^{31} coordinated to the second metal ion (Figure 8).

$Zn(II)_2GAL4(62^*)$ Compared to $Cd(II)_2GAL4(62^*)$. Binding of the metal ions to GAL4(149*) and GAL4(62*) induce the polypeptide conformation required for the specific recognition of the UAS_G DNA sequence (Pan & Coleman, 1989, 1990a). Zn(II) and Cd(II) are both effective in inducing this conformation. Nevertheless, the Cd(II) ion has a significantly larger ionic radius than Zn(II), 0.97 vs 0.74 Å, and makes longer metal–sulfur bonds, ~2.5 vs ~2.3 Å. The complete COSY and NOESY spectra of the Zn(II) and Cd(II) derivatives of GAL4(62*) show that Cd(II) substitution causes some significant changes in the chemical shifts of particular amino acid residues (Table I). Within the core binuclear cluster, the changes in chemical shift are centered on the NH protons of Cys^{11}, Cys^{28}, and the immediately adjacent residues. The NH protons of Cys^{11} and Cys^{28} show the largest shifts in the protein, 0.41 and 0.60 ppm upfield, respectively, when Cd(II) is substituted for Zn(II), while the NH protons of the other four Cys ligands are much less affected (Table I). That the proton chemical shift changes are centered predominantly on just two of the six Cys ligands, Cys^{11} and Cys^{28}, suggests that these two ligand residues may be located in the GAL4(62*) core region in positions where they can shift position most easily to accommodate the larger Cd(II) ions without disrupting the structure significantly. Thus the Cd(II) substitution may lead to an asymmetrical distortion of the cluster.

Not surprisingly, the changes in chemical shift of the NH protons of Cys^{11} and Cys^{28} induced by Cd(II) are accompanied by some alterations in the magnitude of the sequential NOE's in the immediate vicinity of these two ligand residues (Figure 10). Strong αH NOE's extend from Glu^{12} to Gln^9 in the Cd(II) protein but only from Glu^{12} to Cys^{11} in the Zn(II) protein. Likewise there is an additional strong NOE from Ala^{29} to Cys^{28} in the Cd(II) protein as well as an additional medium-strength NOE from Cys^{31} to Lys^{30}. These could reflect the fact that Cd(II) forms a more stable complex with GAL4(62*) than Zn(II) as shown by metal ion exchange studies (Pan & Coleman, 1990a). Therefore the peptide chain may become less conformationally labile; however, changes in the geometry of the turns around the metal cluster may be responsible as well.

The residues in the C-terminal domain that the long-distance NOE's identify as approaching the cluster surface closely, Lys^{45}, Leu^{54}, and Val^{57}, form a third group whose NH protons undergo significant changes in δs upon metal substitution, ~0.1 ppm (Table I). This is another indication that the C-terminal domain, required for specific DNA recognition, interacts closely with the cluster. While the chemical shift changes upon the Zn(II) to Cd(II) substitution do not indicate the precise nature of the underlying structural changes, the pattern of $\Delta\delta$s observed, greatest near the metal ligands, is compatible with the notion that the structure of GAL4(62*) could be described as one in which the binuclear cluster of cysteine residues and Zn(II) ions forms a central scaffold that folds a DNA-binding surface around it. Thus the slight expansion of the central scaffold induced by Cd(II) can be accommodated without altering DNA binding significantly.

General Folding of GAL4(62).* We have used the present intrachain NOE's as well as the long-distance NOE's to calculate a structure for GAL4(62*) by directly entering the NOE distance contraints into the energy-minimization program X-PLOR (Brunger, 1990). This structure is presently being refined and will be the subject of a separate communication. In order to clarify the relationships discussed here, we present a qualitative ribbon diagram of this structure to

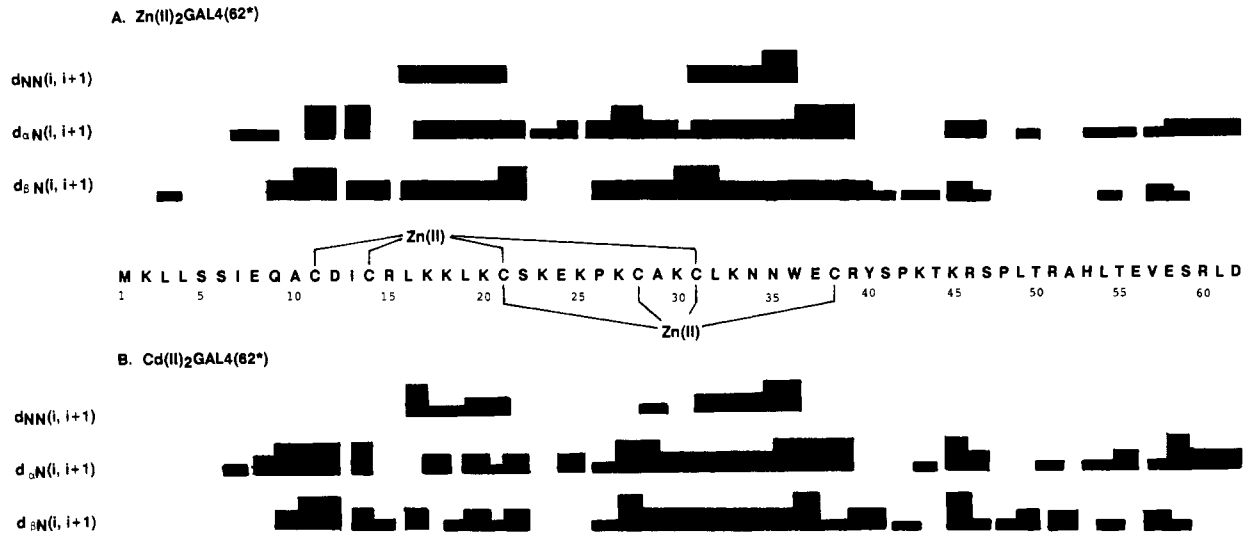

FIGURE 10: Summary of the sequential NOE's observed in $Zn(II)_2$- (A) and $Cd(II)_2GAL4(62^*)$ (B). The intensities of the NOE's at a 250-ms mixing time are indicated by the heights of the bars. For completeness we have included the first eight residues of GAL4(62*) although they do not display significant sequential NOE's. The text should be consulted for the description of the means by which the spin systems of these first eight residues were assigned.

illustrate the relationships within the cluster and the division between the cluster and the C-terminal extension of the DNA-binding domain (Figure 11). The interresidue NOE's that establish the close contacts between the C-terminal subdomain and the cluster are indicated by the dotted lines in Figure 11. The two subdomains are separated in the diagram for clarity. Several residues within loop III (Leu^{32}–Glu^{37}) interact strongly with residues in the C-terminal subdomain. As mentioned in our earlier work, four of the six residues within loop I (Arg^{15}–Lys^{20}) are Lys or Arg residues, and they may function by interacting with the ribose–phosphate backbone of the UAS_G DNA sequence (Pan & Coleman, 1990b). Segment-swapping and mutagenesis studies of GAL4 and PPR1 (Corton & Johnston, 1989) indicate that in GAL4 a protein–DNA charge–charge interaction may also involve Lys^{23}, located in loop II (Ser^{22}–Lys^{27}). Pro^{26} within loop II is also highly conserved among the other fungal transcription factors containing the $Zn(II)_2Cys_6$ motif. Results from mutagenesis of this Pro residue indicate that it is required mainly for proper folding of the protein (Johnston, 1987b).

There are now DNA sequences reported for a total of 11 fungal transcription factors that contain the spacing of 6 Cys residues forming the $Zn(II)_2Cys_6$ binuclear cluster motif as found in GAL4. One other member of this class of proteins, LAC9, has been shown by ^{113}Cd NMR to contain a similar binuclear cluster within its DNA-binding domain (Pan et al., 1990). Although it was not possible to determine by ^{113}Cd NMR which of the cysteines in LAC9 are the two shared ligands, it is reasonable to suggest that they are the cysteines located in the positions in the LAC9 sequence corresponding to the bridging ligands in GAL4, namely, the third and fifth cysteine from the most N-terminal Cys ligand (Figure 11). We expect that the analogous cysteines are the bridging ligands in the other nine GAL4-like transcription factors.

Interactions between the C-terminal Region and the Binuclear Cluster in GAL4(62).* A total of 70 long-range NOE's have been assigned in GAL4(62*), and these form the basis of the three-dimensional structure of GAL4(62*) undergoing refinement. The most interesting quartet among these long-range NOE's involves the backbone and side-chain protons of Asn^{34}, Trp^{36}, Lys^{45}, and Val^{57}. In addition to these long-range NOE's, the chemical shifts of these residues also suggest that these four residues interact closely. Resonances for both

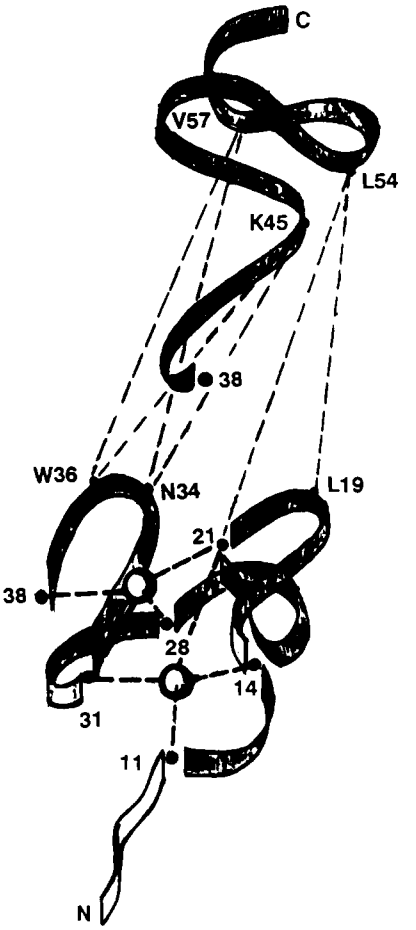

FIGURE 11: Ribbon representation of the $Zn(II)_2Cys_6$ or $Cd(II)_2Cys_6$ binuclear cluster and "specificity region" of the DNA-binding domain of GAL4. The binuclear cluster and the specificity region are disconnected at residue Cys^{38} to illustrate the folding of both parts of the molecule. Dashed lines indicate the long-range NOE's observed between residues Asn^{34}, Trp^{36}, Lys^{45}, and Val^{57} and residues Leu^{54} and Leu^{19}/Cys^{21}.

βH's of Asn^{34} are shifted ~2.0 ppm upfield compared to the shifts expected for the random-coil conformation. Observation of long-range NOE's between a βH of Asn^{34} and the amide proton of Trp^{36} indicates that the unusually large upfield shift

of the former is caused by a ring current shift due to its close approach to Trp[36]. Chemical shifts of the αH and amide protons of Lys[45] differ considerably between the $Zn(II)_2$ and $Cd(II)_2$ species, suggesting its proximity to the cluster (Table I). The γH's of Val[57] are also among the signals that shift upfield upon Zn(II) or Cd(II) binding to the apoprotein.

Strong interresidue NOE's between Val[57] and the two cluster residues, Trp[36] and Asn[34], are among the features that establish close contact between the C-terminal subdomain and the metal cluster structure (Figure 11). The significant upfield chemical shifts of both the β and γ protons of Val[57] (1.74 and 0.68 ppm, ^{113}Cd protein, pH 5.0) are probably due to the close approach of this residue to the indole ring of Trp[36]. The same upfield shifts of the Val[57] protons (1.75 and 0.66 ppm) are observed in the ^1H COSY of ^{113}CdGAL4(62*) at pH 8.0, suggesting that the conformation of the two parts of the DNA-binding domain are preserved over this pH range. Others induced to shift upfield by metal-ion binding include the βH of Asn[34] and the δH of Ile[13] [see Figure 6 in Pan and Coleman (1990b)]. Another triplet of NOE's between the cluster and a residue in the C-terminal subdomain are NOE's between Leu[54] and protons of both Leu[19] and Cys[21]. Both groups of interactions anchor the C-terminal region onto the cluster.

Specific DNA Binding of GAL4(62).* Studies of prokaryotic repressors have established that many of them recognize their specific DNA sequences by utilizing a helix–turn–helix motif with the second α-helix binding in the major groove of the DNA [reviewed by Pabo and Sauer (1984); for an overview, see Matthews (1988)]. A similar helix–turn–helix motif has also been found for several eukaryotic transcription factors that recognize upstream activation sequences. This variant of the helix–turn–helix motif has been called the homeo domain (Otting et al., 1988; Qian et al., 1989). The "classical" Cys_2His_2 "zinc finger" transcription factors may also utilize in part α-helices for specific DNA recognition as suggested by mutagenesis, NMR studies, and model building (Berg, 1988; Parraga et al., 1988, 1990; Lee et al., 1989). The α-helix in Cys_2His_2 zinc fingers is composed of the C-terminal third of the "finger" including both His ligands (Parraga et al., 1988; Lee et al., 1989). A different recognition motif involves an antiparallel β-strand as observed in the *Arc*, *Mnt*, and *Met* repressors (Knight & Sauer, 1989; Zagorski et al., 1989, Rafferty et al., 1989).

What do we know about specific DNA binding by GAL4? Amino acid sequence comparison between GAL4 and LAC9 provided a useful guide for residues involved in contacting DNA, since both DNA-binding domains recognize the same UAS sequences. The only extensive sequence homology within the binuclear cluster itself exists from Arg[15] to Lys[23]. On the other hand, in the region C terminal to the cluster, there is a stretch of 15 amino acid residues of nearly identical sequence in the two transcription factors, Arg[46]–Arg[60] in GAL4. Genetic studies have indicated that the binuclear cluster of GAL4 or PPR1 alone is not sufficient for specific DNA binding. Binding of PPR1 requires the additional 14 residues C terminal to the last cysteine ligand (Corton & Johnston, 1989). We have therefore termed the region C terminal to the cluster as the "specificity region". In GAL4(62*) it includes residues Pro[42]–Arg[60]. In our earlier paper we proposed that GAL4 could utilize a binuclear cluster–turn–helix motif for specific DNA recognition (Pan & Coleman, 1990b). We postulated that the helical region could be formed by residues Leu[49]–Arg[60] with turns in the region of Pro[42] and Pro[48]. The prediction of α-helix formation was based on the circular dichroism spectrum by fitting the negative band observed near 220 nm (Pan & Coleman, 1990a). While the overall CD of both Zn(II)GAL4(149*) and Zn(II)GAL4(62*) cannot be graphically reproduced by a composite curve constructed from the three conformations of poly(L-lysine), α-helix, and β-sheet, and random coil (Pan & Coleman, 1989), the significant negative ellipticity near 220 nm initially suggested a significant amount of α-helix in the DNA-binding domain (Pan & Coleman 1989, 1990a). This prediction is not supported by the sequential NOE data (Figure 10). Recent CD studies comparing the Zn(II) and Cd(II) forms of GAL4(149*) and GAL4(62*) show that the $-S^- \rightarrow$ metal charge-transfer bands in this protein [220 nm for Zn(II) and 250 nm for Cd(II)] are unusually optically active and make a significant contribution to ellipicity. This optical activity appears to account for a significant part of the near ultraviolet features of the CD spectrum of GAL4(62*) (Basile and Coleman, in preparation).

The sequential assignments of GAL4(62*) show that the idea of any extended helix in the "specificity region" of GAL4(62*) is not possible (Figure 10). Although two stretches of NN(i,i+1) NOE's are observed in GAL4, they are between residues within the binuclear cluster (Figures 6 and 10). Both segments lack the αN(i,i+3) NOE's characteristic of α-helices and show strong αN(i,i+1) NOE's inconsistent with α-helix formation (Wuthrich, 1986). Although we cannot exclude the possibility that the flexible N-terminal region (Met[1]–Glu[8]) is helical, this region is not required for specific DNA binding as indicated by segment-swap experiments (Corton & Johnston, 1989). The C-terminal specificity region does not contain extensive classical β-strands either, e.g., strong patterns of cross-chain αCH–αCH NOE's are missing. The precise nature of hydrogen bonding in this region awaits detailed analysis of the three-dimensional structure. From sequential NOE's, the only residues within the specificity region that could possibly be β-stranded are Val[57]–Asp[62]. This stretch may not be directly involved in specific DNA interactions if results from segment-swap experiments from PPR1 are applied to GAL4. Only residues up to Arg[51] appear to be required for specific contacts with DNA (Corton & Johnston, 1989).

Sequence comparison between GAL4 and LAC9 shows that the sequences between Arg[46] and Arg[51] are identical with one exception, a conservative change of Ser[49] to Thr in LAC9. As mentioned earlier, notable interactions between the binuclear cluster and the specificity region involve Lys[45], Val[57], and Leu[54] (Figure 11). These observations strongly suggest that the segment Arg[46]–Arg[51] (Arg[46]-Ser[47]-Pro[48]-Leu[49]-Thr[50]-Arg[51]) is exposed on the surface of the protein molecule and capable of specific interactions with DNA. As indicated by the absence of NOE's, the Arg[46]–Arg[51] segment probably forms a turn or loop. Site-specific mutants of the specificity region that abolish specific DNA binding are clustered within this sequence (Johnston & Dover, 1987, 1988).

GAL4 is the first specific DNA-binding protein characterized that does not contain regular secondary structural elements such as α-helix or β-strands involved in specific DNA recognition. Because of the unique nature of the $Zn(II)_2Cys_6$ binuclear cluster, the same turn–loop motif may be used by all 11 transcription factors containing this binuclear cluster. Three-dimensional structural determination utilizing the long-range NOE's is now in progress, and this will allow us to gain detailed knowledge of how GAL4 and related transcription factors recognize DNA.

Added in Proof

Heteronuclear ^1H–^{113}Cd COSY spectra show that the β_a

and β_b protons of Cys^{11} and Cys^{28} are coupled to both ^{113}Cd nuclei. Thus, Cys^{11} and Cys^{28} are clearly bridging ligands, contrary to the conclusion based on the coupling patterns observed for the $\beta_a\beta_b$ cross-peaks of these residues in standard ^1H–^1H COSY experiments. The β_a and β_b protons of Cys^{14} and Cys^{38} are each coupled to a single ^{113}Cd nucleus, thus these residues are eliminated as bridging ligands, as was concluded also from the ^1H–^{113}Cd coupling observed in ^1H–^1H COSY spectra. The 2D heteronuclear COSY spectra show the β_b protons of Cys^{21} and Cys^{31} to be coupled to only one ^{113}Cd nucleus, while the β_a protons of these two residues may have small couplings to the second ^{113}Cd nucleus. Thus, comparisons of the ^1H–^{113}Cd coupling patterns for the $\beta_a\beta_b$ cross-peaks (^1H–^1H COSY) for the Cys^{11}, Cys^{21}, Cys^{28}, and Cys^{31}, the candidates for the bridging ligands, do not distinguish which are coupled to both ^{113}Cd ions as clearly as we first thought. Although the $\beta_a\beta_b$ cross-peaks of Cys^{21} and Cys^{31} show identical ^1H–^{113}Cd coupling patterns reflecting nearly identical environments in the complex, symmetrical positions as the bridging ligands appear not to be the explanation for this identity. While the current refinements of the 3D structure of GAL4(62*) can accommodate within the RMS deviation either Cys^{11} and Cys^{28} or Cys^{21} and Cys^{31} as the pairs of bridging ligands, the final refinement may provide confirmation of the new assignment of Cys^{11} and Cys^{28}, as strongly indicated by the 2D heteronuclear COSY.

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

References

Andre, B. (1990) *MGG, Mol. Gen. Genet. 220*, 269.

Berg, J. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 99.

Brunger, A. (1990) *X-PLOR Manual*, Yale University, New Haven, CT.

Corton, J. C., & Johnston, S. A. (1989) *Nature (London) 340*, 724.

Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., & Wright, P. E. (1989) *Science 245*, 635.

Johnston, M. (1987a) *Microbiol. Rev. 51*, 458.

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

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

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

Knight, K. L., & Sauer, R. T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 797.

Matthews, B. W. (1988) *Nature (London) 335*, 294.

Otting, G., Qian, Y.-Q., Muller, M., Affotter, M., Gehring, W., & Wuthrich, K. (1988) *EMBO J. 7*, 4305.

Pabo, C., & Sauer, R. (1984) *Annu. Rev. Biochem. 53*, 293.

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

Pan, T., & Coleman, J. E. (1990a) *Biochemistry 29*, 3023.

Pan, T., & Coleman, J. E. (1990b) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2077.

Pan, T., Halvorsen, Y.-D., Dickson, R. C., & Coleman, J. E. (1990) *J. Biol. Chem. 265*, 21427.

Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., & Klevit, R. E. (1988) *Science 241*, 1489.

Parraga, G., Horvath, S. J., Hood, L., Young, E. T., & Klevit, R. E. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 137.

Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W., & Wuthrich, K. (1989) *Cell 59*, 573.

Rafferty, J. B., Somers, W. S., Saint-Girons I., & Phillips, S. E. (1989) *Nature (London) 341*, 705.

Robbins, A. H., McRee, D. E., Williamson, M., Collett, S. A., Xuong, N. H., Furey, W. F., Wang, B. C., & Stout, C. D. (1991) *J. Mol. Biol.* (in press).

Schultze, P., Worgotter, E., Braun, W., Wagner, G., Vasak, M., Kagi, J. H. R., & Wuthrich, K. (1988) *J. Mol. Biol. 203*, 251.

Wagner, G., Neuhaus, D., Worgotter, E., Vasak, M., Kagi, J. H. R., & Wuthrich, K. (1986) *Eur. J. Biochem. 157*, 275.

Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Zagorski, M. G., Bowie, J. U., Vershon, A. K., Sauer, R. T., & Patel, D. J. (1989) *Biochemistry 28*, 9813.