Solution Structure of the DNA-binding Domain of GAL4 from *Saccharomyces cerevisiae*

James D. Baleja, V. Thanabal, Ted Mau, and Gerhard Wagner

Department of Biological Chemistry and Molecular Pharmacology
Harvard Medical School
Boston, Massachusetts 02115

1 Introduction

The GAL4 transcriptional activator protein has long been a favorite for the study of transcription in eukaryotic biology. Genetic studies reveal a modular architecture for the protein with different functions associated with each module [1]. A DNA-binding domain of the protein recognizes and binds to a sequence of DNA termed the Upstream Activating Sequence (UAS_G). Other parts of the protein are relevant for activation. They interact with the transcriptional machinery including RNA polymerase to activate transcription. The UAS_G is near the genes that encode the proteins required for galactose utilization. Upon presentation of galactose to the yeast cell, this DNA site is specifically bound by GAL4, the transcription function of RNA polymerase is activated, and enzymes required for galactose utilization are produced [2].

As a dimer of 881 amino acids, GAL4 is too large for the determination of a high-resolution NMR structure and we have instead studied a fragment containing the N-terminal 65 amino acid residues including its DNA-binding domain [3]. In the absence of DNA, GAL4(65) is monomeric in solution, presumably because it does not have the amino acid residues necessary for dimerization [4]. GAL4(65) is dimeric when bound to any of four DNA sites present in the UAS [5]. Each of these binding sites is approximately two-fold palindromic, and like many other dimeric DNA-binding proteins, each monomer of the GAL4 dimer interacts with a half-site of DNA.

The structure and dynamics of the monomeric DNA-binding domain of GAL4 (residues 1-65; Figure 1), an investigation of the binding to a DNA half-site (Figure 2), and the structure of the resultant protein-DNA complex are presented in this paper.

```
1           11          21
MKLLSSIEQA CDICRLKKLK CSKEKPKCAK
31          41          51
CLKNNWECRY SPKTKRSPLT RAHLTEVESR
61
LERLE
```

Figure 1. Primary amino acid sequence for the DNA-binding domain of GAL4.

1	2	3	4	5	6	7	8	9	10
C	C	G	G	A	G	G	A	C	T
•	•	•	•	•	•	•	•	•	•
G	G	C	C	T	C	C	T	G	A
20	19	18	17	16	15	14	13	12	11

Figure 2. Nucleotide sequence for the one-half DNA-binding site of GAL4.

2 ^{1}H, ^{15}N, and ^{113}Cd NMR resonance assignments

Assignment of specific nuclei to the observed NMR resonance frequencies is the first step in determining the structure of a protein using NMR techniques [6]. Assignments for the ^{1}H, ^{15}N, and ^{113}Cd NMR resonances of GAL4(65) were made using homonuclear and hetero-nuclear NMR experiments and by following standard protocol ([6], Baleja and Wagner, unpublished results).

Zinc is required for the DNA-binding activity of GAL4 [4]. It can be replaced by NMR-active $^{113}Cadmium$ without loss of DNA-binding. The ^{113}Cd NMR spectrum (Figure 3) shows that two metal ions are coordinated by a Cys - $(X)_2$ - Cys - $(X)_6$ - Cys - $(X)_6$ - Cys - $(X)_2$ - Cys - $(X)_6$ - Cys motif using six cysteines and forming a bimetal-thiolate cluster [7]. Hetero-nuclear correlation experiments between the ^{113}Cd and ^{1}H define the liganding of the two central metal ions (Figure 3).

Having assigned resonance frequencies to specific nuclei, the correspondence of cross peaks to the protons can be made and local structural information was can then be determined. A section of the two-dimensional NOESY spectrum is shown in Figure 4. The region shows the cross peaks among the amide protons of the protein backbone and protons of the aromatic side-chains.

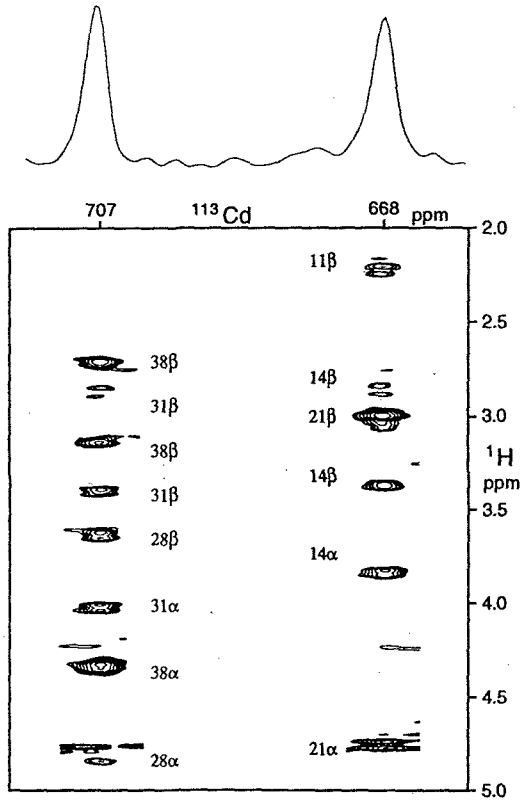

Figure 3. Heteronuclear ^{113}Cd-^{1}H correlation experiments. The standard reverse INEPT pulse sequence [8] was followed by a short MLEV-17 TOCSY transfer [9] and proton detection.

Indicating the presence of two α-helices, there are a series of close approaches between amide proton resonance frequencies for several sequential residues. Other NOEs arise between amino acid residues more distant in sequence and define the unique topological features for the protein. NOE intensities at 56, 150, and 250, and 500 millisecond mixing times were converted to the corresponding interproton distances using the usual inverse r^6 relationship [3]. In case of overlap in the homonuclear NOESY, spectrum analysis was also made using a three-dimensional ^{15}N-^{1}H-^{1}H NOESY-HMQC spectrum recorded on a uniformly ^{15}N-labeled protein (Figure 5, [10]). From the NOESY data,

residues 9-40 are observed to form a well-defined, compact globular cluster. The terminal residues 1-8 and 41-66 show little persistent structure in solution. There are many inter-residue NOE crosspeaks for the central core (30 per residue, on average), but only a few weak NOE cross peaks for protons in the disordered region. In addition, the amide protons of the flexible trails are in rapid exchange with solvent [10] and have narrow resonance lines, indicating that these residues have considerable conformational mobility in the absence of DNA. This unstructured character for parts of the GAL4 protein may be vital to its biological function. On the other hand, at least some aspects of this mobility may merely be a consequence of the truncation used for GAL4 in this study. Nonetheless, our picture of GAL4 is one in which two N-terminal recognition modules are connected by flexible linkers to a dimeric core [4].

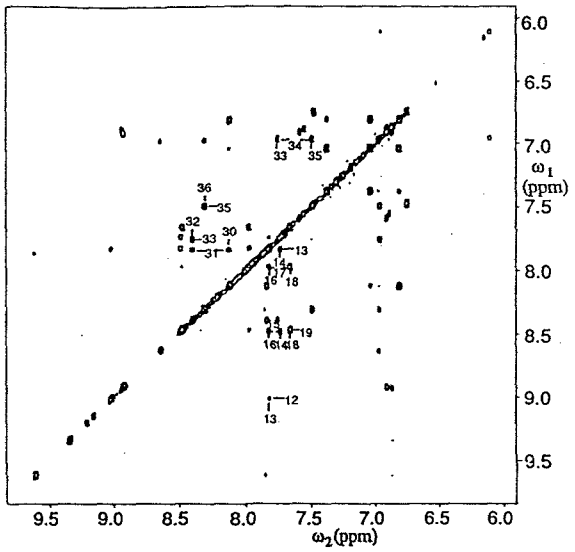

Figure 4. NOESY spectrum of amide and aromatic protons of GAL4(65). Sample concentration was 1.5 mM in 0.2 M NaCl, 20 mM sodium phosphate, pH 7.0 at 25°C. Numbers indicate the residues for which sequential strong amide to amide cross peaks are observed which are indicative of α-helices.

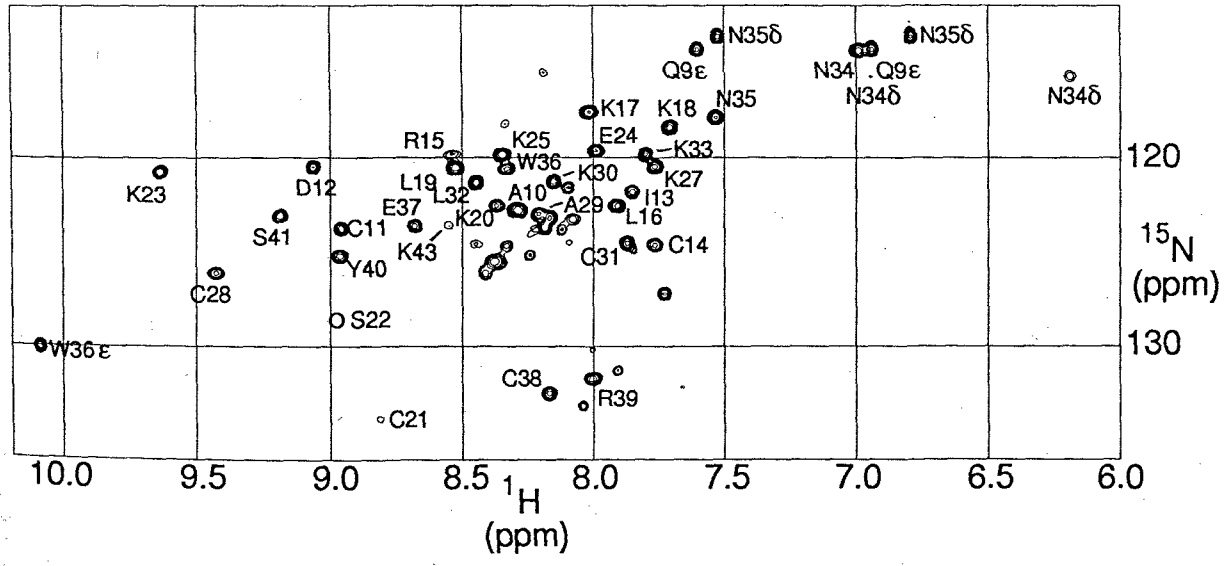

Figure 5. Heteronuclear single quantum coherence ^{15}N-^{1}H spectra [11] of uniformly ^{15}N-labeled GAL4(65). The sample concentration was 1.5 mM in a buffer of 0.1 M NaCl, 20 mM phosphate, pH 7.18.

3 Structure of the GAL4 DNA-binding domain

Sets of interproton distances and torsion angles formed the basis for structure determination. Distances were determined from the NOESY data [3]. ϕ torsion angles were interpreted from measured NH-α coupling constants and χ_1 angles were derived from ^{15}N-β and α-β coupling constants [3]. Sulfur-Cd liganding distances were imposed to be 2.35 to 2.45 Å, Cys Cβ-Cd distances to be less than 3.4 Å, and sulphurs liganding the same Cd to be between 3.3 and 4.2 Å. 614 distance and 41 torsion angle measurements were used with the distance geometry package DG-II to generate a set of structures (0.6 Å backbone atom rmsd) in agreement with the NMR data [3]. A schematic of the structure (Figure 6) shows the two central metal ions coordinated by the six cysteines. The DNA-binding domain consists of an α-helix and an extended structure, then a sharp turn that contains a *cis*-proline bond, and then a 2nd α helix followed by an extended structure. If the central metal binding subdomain of GAL4 is split, and the two halves superimposed, a striking correspondence between each part is revealed. The conformation of the 13 residue segment from residues 10 to 22 is almost identical to that of residues 27 to 39, with an rmsd of 0.8 Å for the backbone atoms. Although the structural integrity of the protein would be provided mainly by the way in which the two metal ions are liganded, some hydrophobic packing is observed with the side chains of W36 and Y40 [3]. Slowly exchanging amides of GAL4, in both the free and DNA-bound forms [10], can be attributed, in part, to hydrogen bonding to carbonyl oxygens within the α-helices, and to sulfurs of the cysteinyl side-chains [12].

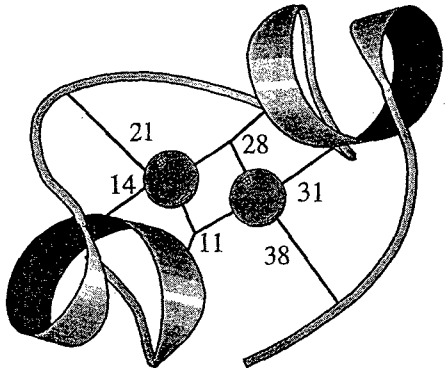

Figure 6. Structure of the central core for the DNA-binding domain of GAL4. Cysteines that ligand the two central metal ions are numbered.

4 Formation of a GAL4-DNA complex

Imino protons of one half of the consensus dimeric GAL4 DNA-binding were monitored in forming the complex between the protein and DNA (Figure 7). One imino proton is present for each base-pair of a 10 base-pair DNA duplex.

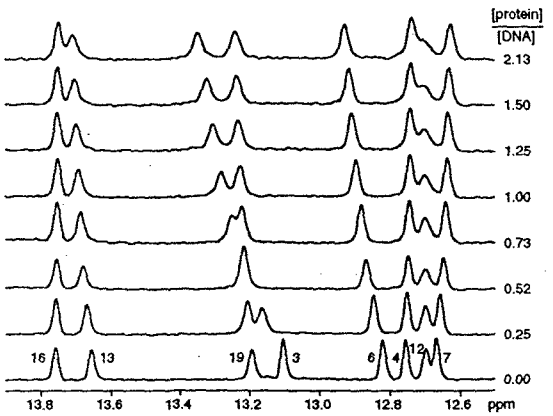

Figure 7. Titration of DNA with GAL4.

Experimentally, eight imino proton resonances are observed, since the imino protons on the base-pairs at each end of the duplex exchange rapidly with the bulk solvent at room temperature. Each imino resonance has been assigned to a specific base-

pair. Resonances shift as the environment around each proton changes upon addition of GAL4. All resonances broaden as the more slowly protein-DNA complex is formed. The equilibrium dissociation constant is 169±13 μM. Resonances are in fast exchange between the protein-DNA complex and the free components indicating rapid equilibrium between free and bound forms.

NOESY and TOCSY two-dimensional data for the protein-DNA complex have been collected (Figure 8). These spectra are promising for full analysis since the resonance lines are not too broad for resonance assignment, chemical shift dispersion is good, and solubility of the protein-DNA complex is adequate. The DNA resonances have been completely re-assigned and the protein resonances have been partially re-assigned within the protein-DNA complex.

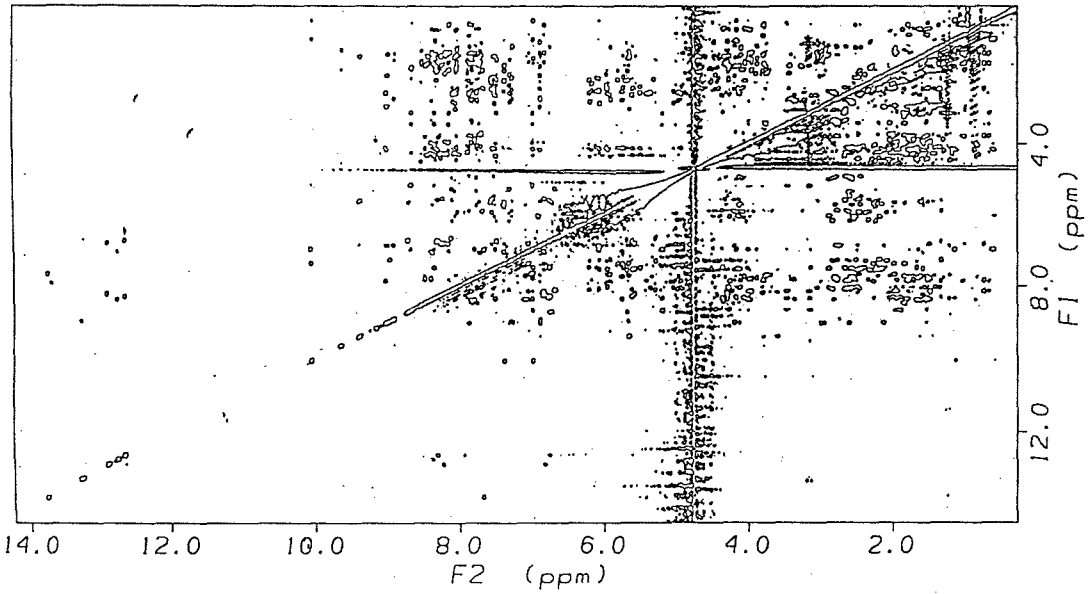

Figure 8. NOESY spectrum of the GAL4:DNA complex. The concentration of the complex was approximately 0.5 mM. Buffer conditions were 0.15 M NaCl, 20 mM PO_4, pH 7, 25°C.

5 NMR structure of the GAL4-DNA complex

Several NOE contacts have been observed between protein and DNA (dashed lines, Figure 9) yielding a preliminary stucture for this GAL4-DNA complex. The recognition helix for the free form of the protein was docked onto a B-DNA. The amino acid residues responsible for recognizing DNA are part of the metal-binding cluster and interact with edges of base-pairs exposed in the major groove of the DNA.

The structure of the DNA-binding domain of GAL4 (residues 1-65) bound to full site DNA containing two GAL4 binding sites has been determined crystallographically. In agreement with our observed intermolecular NOE cross peaks (Figure 8), the C-terminal end of the first α-helix of the metal binding cluster (residues 9-40) provides for sequence-specific re-cognition of DNA by GAL4. When bound to full site DNA, parts of GAL4(65), before unstructured in solution, adopt a regular conformation. Residues 41 to 49 form a linker region which interacts with

the backbone of the DNA. In addition, residues 50-64 form a small dimerization domain using a coiled-coil type packing arrangement [4]. Our preliminary results on an intact dimerization element for GAL4 (residues 50-106) indicate that the α-helical character for the coiled-coil of the protein extends beyond residue 64 (Baleja, Marmorstein, and Wagner, unpublished results). The same conformation (1.1 Å rmsd) is observed for the recognition module of GAL4(65) in solution using NMR techniques as for the central metal-binding cluster of GAL4(65) bound to DNA using crystallographic techniques. Thus the core of the DNA-binding domain changes little upon binding DNA.

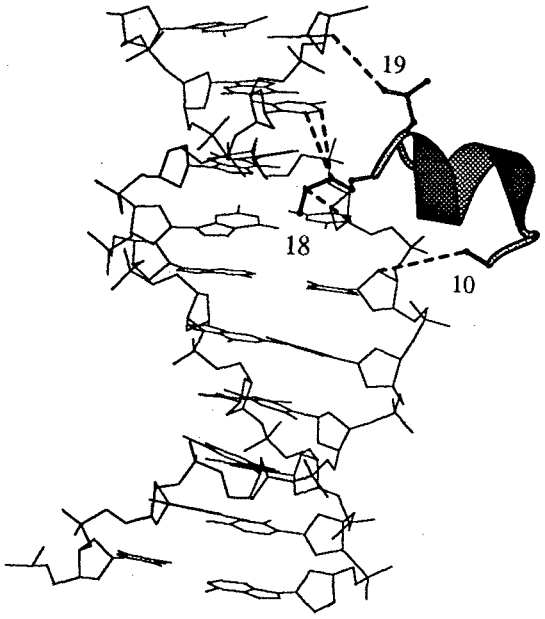

Figure 9. Preliminary structure for a GAL4-DNA complex.

6 Summary

The DNA-binding domain of GAL4 has many interesting features. It is a novel DNA-binding motif with a globular two-metal cluster that has hydrogen-bonds to sulfur and two-fold internal symmetry. GAL4 reads the sequence of DNA through the amino acids present at the C-terminal end of the first α-helix. The two DNA-reading modules of intact GAL4 are tethered by flexible linkers to the central body of the protein. Once bound to DNA through the N-terminal recognition domains, the C-terminal portion of GAL4 is in the correct position to interact with the components of the transcriptional machinery to bring about transcriptional activation.

7 References

1. Keegan, L., Gill, G., and Ptashne, M., Science 231, 699 (1986).

2. Johnston, M., Micro. Rev. 51, 458 (1987).

3. Baleja, J. D., Marmorstein, R., Harrison, S. C., and Wagner G., Nature 356, 448 (1992).

4. Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S. C., Nature 356, 408 (1992).

5. Carey, M., Kakidani, H., Leatherwood, J., Mostashari, F., and Ptashne, M., J. Mol. Biol. 209, 423 (1989).

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

7. Pan, T., and Coleman, J. E., Biochemistry 209, 3023 (1990).

8. Brühwiler, D., and Wagner, G., J. Magn. Reson. 69, 546 (1986).

9. Bax, A., and Davis, D. G., J. Magn. Reson. 65, 355 (1985).

10. Mau, T., Baleja, J. D., and Wagner, G., Protein Science (submitted).

11. Bodenhausen, G., and Ruben, D. J., Chem. Phys. Lett. 69, 185 (1980).

12. Kraulis, P., Raine, A. R. C., Gadhavi, P. L., and Laue, E. D., Nature 356, 448 (1992).