Recognition of DNA by GAL4 in Solution: Use of a Monomeric Protein–DNA Complex for Study by NMR†

James D. Baleja,*,‡ Ted Mau,§ and Gerhard Wagner*

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

Received November 12, 1993; Revised Manuscript Received January 4, 1994®

ABSTRACT: The complex of a monomer of GAL4 with DNA has been investigated by two-dimensional ^{1}H nuclear magnetic resonance (NMR) spectroscopy. Previous X-ray analysis has revealed a structure in which a dimer of the N-terminal 65-residue fragment of GAL4 forms a complex, 27 kDa in molecular mass, with a 19 base pair full-binding-site DNA [Marmorstein, R., Carey, M., Ptashne, M., & Harrison, S. C. (1992) *Nature 356*, 408–414]. We have developed a smaller system, half in molecular mass, which is amenable for detailed analysis using NMR. Titration of a 10 base pair half-binding-site DNA with GAL4-(65) shows 1:1 binding, illustrating that one monomer of the protein binds in a specific manner to half-site DNA. The components of the protein–DNA complex are mainly in fast exchange on the NMR chemical shift time scale, with an equilibrium dissociation constant of 161 ± 12 μM. With a basis of chemical shift data for free GAL4 protein and for the free half-site DNA, the fast exchange facilitates ^{1}H resonance assignments in the complex since cross-peak positions can be examined at different protein:DNA ratios. Chemical shift changes in the DNA reveal the base pairs that are important for recognition by GAL4. Intermolecular NOE cross-peaks are also observed in spectra of the protein–DNA complex. Their identification places the C-terminal end of the first α-helix (residues 12–17) in a position such that the amino acids are able to read the DNA sequence in a manner entirely consistent with the X-ray structure of the related complex. Dimerization of GAL4 therefore is not required for specific recognition of the base pairs present in half of the binding-site DNA.

The GAL4 protein activates transcription of the genes required for galactose utilization in *Saccharomyces cerevisiae* (Johnston, 1987). The protein of 881 amino acids binds as a dimer to approximately 2-fold symmetric DNA sites present in the galactose upstream activating sequence (UAS_G)[1] (Bram & Kornberg, 1985; Carey et al., 1989; Chasman & Kornberg, 1990; Giniger et al., 1985). The N-terminal portion (residues 1–65) contains the amino acid residues responsible for DNA recognition and binding. The core of the DNA-binding domain contains a Cys-$(X)_2$-Cys-$(X)_6$-Cys-$(X)_6$-Cys-$(X)_2$-Cys-$(X)_6$-Cys motif in which the six cysteines, located on two helix/extended strand units connected by a long loop, coordinate two central zinc ions, forming a bimetal–thiolate cluster (Baleja et al., 1992; Gardner et al., 1991; Kraulis et al., 1992; Marmorstein et al., 1992; Pan & Coleman, 1989). The structure of the GAL4 DNA-binding domain represents a novel DNA-binding motif and has been termed a class III zinc-containing DNA-binding protein (Harrison, 1991). The core of the domain changes little upon binding DNA. Essentially the same conformation is observed for the compact domain in solution using NMR techniques as for the central core of GAL4(65) bound to DNA using crystallographic techniques (Marmorstein et al., 1992).

GAL4 appears to have a modular architecture consisting of domains associated with DNA binding, catabolite repression, transcriptional activation, and GAL80 regulation (Johnston, 1987; Stone & Sadowski, 1993). Use of an N-terminal portion of GAL4, which contains the DNA-binding domain, has been a favorite tool in eukaryotic molecular biology for the creation of fusion proteins that are directed to bind DNA carrying the GAL4-binding site (Ma & Ptashne, 1987; Sadowski et al., 1988). The fragment of GAL4 studied here is monomeric in solution (Baleja et al., 1992). Crystallographic analysis of the same fragment in complex with full-site DNA shows that GAL4(65) becomes dimeric using helices that are formed by residues 50–64 and through contacts to DNA in the two half-sites (Carey et al., 1989; Marmorstein et al., 1992). Our model for the GAL4 DNA-binding domain is one in which two N-terminal DNA recognition modules are connected by flexible linkers to a dimerization subdomain (Baleja et al., 1992; Marmorstein et al., 1992). The 2-fold symmetric nature of the DNA sequence and the dimeric form of the protein would suggest that each half of the DNA-binding site is mainly recognized by one monomer of the protein dimer (Kim & Little, 1992). Indeed, the structure of the GAL4–DNA complex shows preservation of 2-fold symmetry in the complex and mainly monomer ↔ half-site DNA recognition. Therefore, we have studied a system in solution that highlights the recognition of DNA by this eukaryotic transcriptional factor while avoiding complications of protein dimerization and at the same time maintaining a molecular size amenable for NMR analysis of the protein–DNA complex. Here, we describe the formation and structural analysis of a protein–DNA complex in which one monomer of GAL4(65) is bound to a 10 base pair DNA duplex containing half of the DNA recognition element. Our results are consistent with

† This research was supported by NIH Grant GM47467. J.D.B. acknowledges the support of the Medical Research Council of Canada and the Alberta Heritage Foundation for Medical Research.

* Correspondence should be addressed to these authors.

‡ Present address: Department of Biochemistry, Tufts University School of Medicine, Boston, MA 02111.

§ Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143.

® Abstract published in *Advance ACS Abstracts*, February 15, 1994.

[1] Abbreviations: UAS_G, galactose upstream activating sequence; NMR, nuclear magnetic resonance; HPLC, high-performance liquid chromatography; TSP, (trimethylsilyl)propionate; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; TOCSY, total correlation spectroscopy; TPPI, time proportional phase incrementation; E-COSY, edited correlation spectroscopy.

0006-2960/94/0433-3071$04.50/0

the interaction sites of the complex determined by X-ray crystallography (Marmorstein et al., 1992). In addition, the NMR data provide an accurate measurement of the equilibrium dissociation constant between protein and DNA.

EXPERIMENTAL PROCEDURES

NMR Sample Preparation. The DNA-binding domain containing the N-terminal 65 amino acid residues of GAL4 plus a C-terminal Phe derived from the cloning construct was prepared as described previously (Marmorstein et al., 1992; Mau et al., 1992). Protein concentrations were measured using a Coomassie blue dye-binding assay (Bio-Rad Laboratories). Oligonucleotides corresponding to each strand of the DNA duplex were synthesized on a Milligen DNA synthesizer and purified by reverse-phase HPLC. The nucleotide base sequence used was: The duplex was prepared

$$\begin{array}{l} 5'\text{d}(C_1\ \ C_2\ \ G_3\ \ G_4\ \ A_5\ \ G_6\ \ G_7\ \ A_8\ \ C_9\ \ T_{10})3' \\ 3'\text{d}(G_{20}\ G_{19}\ C_{18}\ C_{17}\ T_{16}\ C_{15}\ C_{14}\ T_{13}\ G_{12}\ A_{11})5' \end{array}$$

by mixing equimolar amounts of each strand (as determined by A_{260}) in a small volume (ca. 50 μL), heating to 80 °C, and slow-cooling overnight to room temperature. NMR samples contained a trace of TSP as an internal proton chemical shift reference and 0.05% NaN_3 to retard bacterial growth. H_2O samples contained 10% D_2O for the deuterium lock signal. Buffers consisted of 0.1 M NaCl, 10 mM Na_2HPO_4, and 10 mM NaH_2PO_4, pH 7.0 ± 0.2 (direct meter readings).

NMR Spectroscopy. For the titration of DNA with Cd-GAL4, a 2.5 mM protein stock solution was added in aliquots of 10–40 μL to 0.5 mM duplex DNA in 0.45 mL of buffer. The titration was carried out at 25 °C and monitored with one-dimensional spectra using a 1-1 binomial water suppression sequence (Hore, 1983). Spectra were collected on a Bruker AMX-500 spectrometer with a proton frequency of 500.14 MHz. The carrier frequency in the proton channel was set on the water resonance. Two-dimensional NOESY spectra were taken midway through the titration ([GAL4]:[DNA] ratio = 0.86) and again at the end ([GAL4]:[DNA] ratio = 2.31). A D_2O sample was prepared by lyophilization and resuspension in 450 μL of 99.996% D_2O, for which NOESY and TOCSY spectra were recorded. No changes ^{1}H NMR resonance positions were observed after lyophilization, indicating that the resolubilized protein–DNA complex retains its native conformation. NOESY spectra were acquired with a mixing time of 150 ms, 128 summed scans, 1024 real t_2 points with a spectral width of 6024 Hz, 512 t_1 TPPI increments, and a relaxation delay of 1.2 s between scans. Spectra were apodized with squared sine bells shifted by 45° in t_2 and 60° in t_1 and zero-filled to a 2K by 1K (real) matrix using the FELIX NMR processing program.

A TOCSY spectrum was collected using identical parameters, except with a mixing time of 43.8 ms using an MLEV-17 mixing sequence (Bax & Davis, 1985), and processed with squared sine bells shifted by 45° in t_2 and 60° in t_1 and zero-filled to a 2K by 1K (real) matrix.

One-to-one binding of protein, P, and DNA, D, to form a protein–DNA complex, PD, can be expressed as

$$\text{P} + \text{D} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \text{PD} \tag{1}$$

The equilibrium dissociation constant, K_d, for this reaction is

$$K_d = \frac{k_d}{k_a} = \frac{[\text{P}][\text{D}]}{[\text{PD}]} \tag{2}$$

With P_0 = [P] + [PD] and D_0 = [D] + [PD], K_d is

$$K_d = \frac{\{P_0 - [\text{PD}]\}\{D_0 - [\text{PD}]\}}{[\text{PD}]} \tag{3}$$

When chemical shift changes are in fast exchange ($\delta_{obs} = \delta_f f_f + \delta_b f_b$, where δ_f is the chemical shift of the unbound species, δ_b is that of the bound species, and f_f and f_b are the fractions free and bound, respectively, with $f_f + f_b = 1$), the observed change in chemical shift, $\Delta\delta$, is

$$\Delta\delta = \frac{[\text{PD}]}{D_0}(\delta_b - \delta_f) \tag{4}$$

or

$$[\text{PD}] = \frac{\Delta\delta}{\delta_b - \delta_f} D_0 = f_b D_0 \tag{5}$$

Substitution of eq 5 into eq 3 yields

$$K_d = \frac{(P_0 - f_b D_0)(D_0 - f_b D_0)}{f_b D_0} \tag{6}$$

The titration data (plotting $\Delta\delta$ versus P_0/D_0) are fit using a nonlinear least-squares method for the quadratic function (Hull, 1975). K_d and $\delta_b - \delta_f$ are solved using the analysis program CRVFIT (R. Boyko, personal communication). Concentrations of protein and DNA are corrected for dilution through the titration by the measured changes in volume. Equation 6 may be rearranged to yield

$$\frac{f_b}{P_0 - f_b D_0} = \frac{f_b}{[\text{P}]} = \frac{-f_b}{K_d} + \frac{1}{K_d} \tag{7}$$

which is the familiar Scatchard plot analysis (plotting $f_b/[\text{P}]$ vs f_b), with the y-intercept equal to K_d^{-1}, the slope of the line equal to $-K_d^{-1}$, and the x-intercept equal to the number of binding sites (Ferguson-Miller & Koppenol, 1981).

RESULTS AND DISCUSSION

The study of the complex of GAL4(65) and the half-site binding-site DNA was begun by examining each component separately. Sequence-specific assignments were obtained for residues 6–44 of the DNA-binding domain for the cadmium-substituted form of GAL4 using standard techniques (J. D. Baleja, T. Mau, V. Thanabal, and G. Wagner, unpublished results; Baleja et al., 1992; Wüthrich, 1986). Our resonance assignments in the metal-binding region (residues 10–40) do not agree well with those previously published for Cd-GAL4 (Gardner et al., 1991; Pan & Coleman, 1991). However, we do observe the same sequential NOE patterns and a great similarity in resonance assignments with the NMR data published for Zn-GAL4, a fact which reflects the homology in structure between the zinc and cadmium forms (Baleja et al., 1992; Gadhavi et al., 1990; Kraulis et al., 1992; Shirakawa et al., 1993). Since the poorly structured regions (residues 1–9 and 41–65) exhibit either weak or no NOEs, complete assignments have not yet been obtained for residues 1–5 or residues 45–65 of Cd-GAL4 (Baleja et al., 1992; Kraulis et al., 1992; Shirakawa et al., 1993). Our partial assignments in this region agree with those completed for Zn-GAL4 (Gadhavi et al., 1990; Shirakawa et al., 1993). Analysis of the GAL4(65)–10mer DNA complex is not compromised since

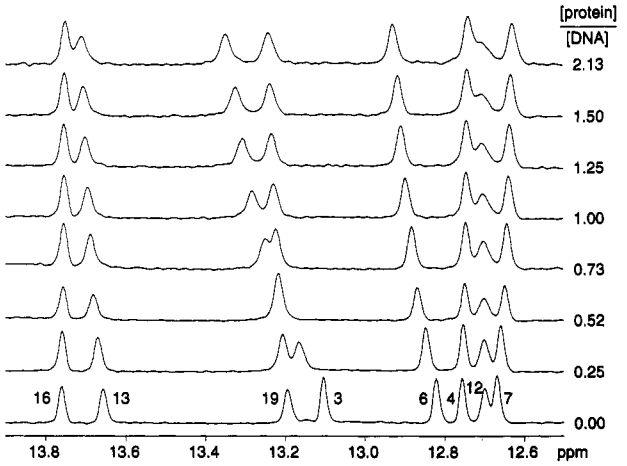

FIGURE 1: Spectral changes in DNA imino resonances upon addition of GAL4 protein. Shown are ^{1}H NMR spectra between 12.5 and 13.9 ppm. Numbered resonances indicate the assignment of imino protons to each nucleotide of the DNA. Aliquots of GAL4 protein were titrated into the half-site DNA giving molar protein:DNA ratios ranging from 0.00 to 2.31. Buffer conditions are 0.1 M NaCl/20 mM phosphate, pH 7.0, at 25 °C.

these parts of GAL4 are not required for the specific interaction of GAL4 to DNA (Gadhavi et al., 1990; Marmorstein et al., 1992).

DNA assignments were made following standard procedures (Feigon et al., 1982; Gronenborn & Clore, 1985; Hare et al., 1983; Scheek et al., 1983; Wüthrich, 1986) and are given in Table 1. Analysis of NOE intensities from 150-, 300-, and 450-ms NOESY spectra and measurement of coupling constants from E-COSY spectra show that the free duplex generally adopts a B-DNA conformation in solution, in agreement with earlier studies (Hansen et al., 1991). Imino proton signals of the DNA were monitored as protein was added to form the protein–DNA complex (Figure 1). The spectra taken in the course of titration show an increase in line width as the higher molecular weight protein–DNA complex becomes the predominant molecular species. Several resonance lines change chemical shift and exhibit fast exchange between free and bound resonance positions on the NMR chemical shift time scale. Therefore, with the knowledge of the chemical shift assignments for imino protons of free DNA, chemical shifts in the protein–DNA complex could be determined easily by extrapolating the change in resonance position to 100% bound.

Summing the changes in chemical shift for all observed imino protons and plotting the result against the protein:DNA ratio yield a binding curve. Analysis of the curve, assuming an equilibrium between one monomer of GAL4(65) and one DNA duplex, gives a equilibrium dissociation constant of 161 ± 12 μM. The data can be recast into Scatchard plot form and are presented in Figure 2. To reduce scatter, the data are summed for all observed imino protons at each point in the titration. Since the weighting of individual points is different from that in the chemical shift vs ratio plot, Scatchard analysis yields a K_d of 173 ± 3 μM. Analysis for individual imino protons that shift more than 0.04 ppm between free and bound forms (3, 6, 7, 13, and 19) gives a narrow range of binding constants (between 69 ± 15 and 175 ± 14 μM, $\bar{x}$ = 138 ± 38 μM). Analyses of the two resonances that shift little between free and bound forms have a much greater error associated with the determined binding parameters (e.g., analysis for G_4 indicates a K_d of 325 ± 277 μM and a chemical shift difference of 0.014 ± 0.004 ppm between free and bound forms). Only 1:1 binding of protein to DNA fits the observed change in

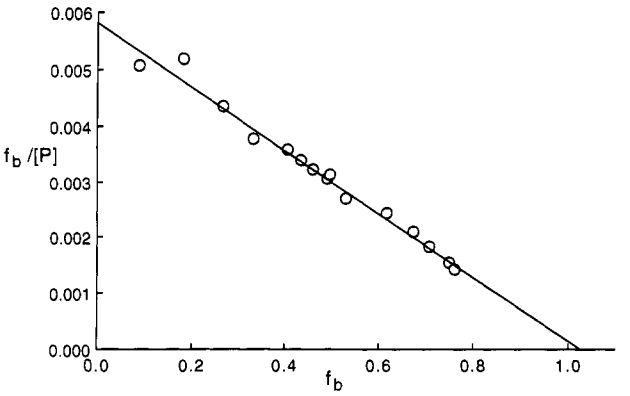

FIGURE 2: Scatchard analysis of the binding of GAL4 protein to its half-site DNA. Using an equation that describes the binding of one monomer of GAL4(65) to one DNA duplex (see Experimental Procedures), the line drawn through the points is calculated with an equilibrium dissociation constant of 172 μM. The data are summed for all observed imino protons at each point in the titration. The initial DNA concentration was 0.5 mM, and the final concentration was 0.36 mM.

imino chemical shift positions. Although other regions of the DNA spectrum could possibly be monitored as GAL4 is added, only the imino protons are sufficiently disperse in a one-dimensional spectrum so that each resonance can be followed individually throughout the titration, and only this region of the DNA spectrum (in H_2O solution) is clear of protein resonances. Parenthetically, the measured K_d for GAL4(65) binding to a full UAS is about 1 μM (S. D. Liang, R. Marmorstein, S. C. Harrison, and M. Ptashne, unpublished results). For fully dimeric GAL4(147), the K_d is about 2×10^{-11} M (Carey et al., 1989).

Since the resonances for individual imino protons could be followed from free DNA to the complex with protein, their assignment in the protein–DNA complex was trivial. Assignment of the remaining DNA ^{1}H resonances was more demanding because of resonance overlap in one-dimensional spectra. Therefore, a two-dimensional NOESY spectrum was taken near the midpoint of the titration, and again at the end of the experiment. The sample was then lyophilized and resuspended in D_2O for further NMR analysis using NOESY and TOCSY spectra.

Assignment of DNA resonances within the protein–DNA complex (in equilibrium with free protein and DNA) was carried out by extending the known imino assignments to adenine H2 and cytosine H4 protons, assuming Watson–Crick base-pairing interactions (Gronenborn & Clore, 1985). In turn, cytosine H4 protons gave cross-peaks to cytosine H5, which, in turn, had cross-peaks to H6 protons. Having chemical shift positions for the cytosine H6 and H5 protons as a base, assignment to the remaining protons was obtained using standard procedures for right-handed DNA in D_2O solution, as carried out previously for the unbound DNA.

In a solution of GAL4(65) and the half-site DNA, an equilibrium exists between free protein, free DNA, and the protein–DNA complex. Since the lifetime of the complex is much shorter than the frequency difference between free and bound resonances (i.e., it shows fast exchange behavior), the population-averaged resonance is observed. Assignments for a protein–DNA complex can be obtained in a facile manner by altering the ratio of protein to DNA. This changes the populations of free and bound forms of the molecule and consequently produces shifts in resonance positions that circumvent problems of resonance overlap. For DNA resonances that shift upon addition of protein, the addition of

Table 1: ^{1}H NMR Chemical Shifts (in ppm) for GAL4 Binding-Site DNA[a]

posi-		H6/H8		1′		2′		2″		3′		4′		5′,5″		NH[b]		AH2 CH5 TH7	
tion	base	δ	Δ	δ	Δ	δ	Δ	δ	Δ	δ	Δ	δ	Δ	δ	Δ	δ	Δ	δ	Δ
1	C	7.65	+0.04	5.98	0.00	2.06	−0.13	2.49	−0.03	4.66	−0.01	4.11	−0.04	3.72,3.76	−0.04,−0.10			5.96	−0.00
2	C	7.49	+0.33	5.54	−0.42	2.03	+0.24	2.35	−0.09	4.82	+0.01	4.09	−0.02	4.01,*c*	−0.00,*c*	8.67,6.94	−0.12,−0.04	5.67	+0.07
3	G	7.84	+0.03	5.49	+0.55	2.65	−0.11	2.71	+0.04	4.96	+0.07	4.37	+0.09	3.95,4.07	−0.02,+0.02	13.12	+0.31		
4	G	7.76	+0.04	5.53	−0.33	2.59	−0.03	2.72	−0.20	4.98	−0.01	4.35	−0.04	4.14,4.19[e]	*f*,+0.02	12.77	−0.01		
5	A	8.00	+0.17	6.01	+0.16	2.55	+0.08	2.84	+0.13	5.02	+0.08	4.38	+0.09	4.13,4.17	−0.01,−0.03			7.67	+0.04
6	G	7.51	+0.07	5.48	+0.08	2.45	+0.01	2.60	+0.03	4.92	+0.07	4.30,*c*	+0.01	4.15,*c*	+0.03,*c*	12.84	+0.14		
7	G	7.57	+0.04	5.64	+0.03	2.50	−0.04	2.69	−0.01	4.94	0.00	4.35,*c*	−0.01	4.15,4.22[e]	*f*,−0.04	12.68	−0.05		
8	A	8.10	+0.04	6.19	+0.05	2.60	+0.03	2.84	+0.05	4.99	+0.04	4.41	−0.005	4.13,4.18	+0.07,+0.01			7.94	+0.04
9	C	7.34	−0.01	5.98	0.00	2.10	+0.01	2.41	−0.01	4.76	+0.03	4.18,*c*	0.00	4.12,*c*	+0.05,*c*	8.14,6.70	*f*	5.30	0.00
10	T	7.51	+0.01	6.22	0.00	2.25	+0.01	2.25	+0.01	4.53	0.00	4.05	0.00	4.13,4.04[e]	*f*,−0.01			1.67	+0.03
11	A	8.04	+0.01	5.97	0.00	2.50	+0.01	2.63	−0.01	4.82	0.00	4.20	−0.01	3.69,3.69	0.00,0.00			7.90	+0.02
12	G	7.88	+0.01	5.92	+0.01	2.67	+0.01	2.79	+0.01	4.95	+0.01	4.40	0.00	4.12,4.12	0.00,0.00	12.73	+0.01		
13	T	7.31	+0.01	6.08	+0.01	2.22	+0.01	2.55	−0.01	4.88	0.00	4.28	0.00	4.16,4.23	−0.05,−0.01	13.69	+0.07	1.33	+0.03
14	C	7.58	0.00	6.00	−0.04	2.29	+0.12	2.52	−0.04	4.82	+0.01	4.23	−0.01	*d*	*d*	8.31,6.74	+0.02,+0.08	5.62	+0.01
15	C	7.56	+0.01	5.85	−0.20	2.12	+0.12	2.49	−0.03	4.74	−0.21	4.17	+0.01	*d*	*d*	8.31,6.87	−0.18,−0.10	5.59	−0.01
16	T	7.47	+0.26	6.05	+0.18	2.20	0.00	2.55	+0.04	4.87	−0.05	4.21	+0.12	4.10,*c*	0.00,*c*	13.78	−0.02	1.62	−0.03
17	C	7.55	−0.28	5.90	−0.16	2.12	−0.22	2.38	−0.18	4.82	−0.07	4.14	+0.03	4.07,*c*	*f*,*c*	8.46,6.80	−0.16,+0.61	5.67	+0.17
18	C	7.43	+0.04	5.46	+0.53	1.93	−0.13	2.24	+0.11	4.80	+0.11	4.07	+0.13	4.04,4.17[e]	*f*,−0.01	8.70,6.90	+0.61,+0.79	5.66	+0.03
19	G	7.83	+0.12	5.60	−0.42	2.63	−0.04	2.70	−0.13	4.95	−0.01	4.31	−0.09	3.96,4.04	−0.01,0.00	13.22	+0.06		
20	G	7.80	+0.12	6.13	+0.12	2.51	+0.09	2.34	+0.05	4.64	+0.04	4.18	+0.07	4.10,4.12[e]	*f*,0.00				

[a] Chemical shifts ±0.02 ppm are relative to HOD at 4.76 ppm (also the carrier frequency at 25C) for the free, unbound DNA. Beside each chemical shift is the change when fully bound to GAL4 protein. 5′ and 5″ protons are not assigned stereospecifically. [b] Assignments are the for the H1 imino protons of guanine, the H3 imino protons of thymine, or the hydrogen-bonded/non-hydrogen-bonded amino protons on N4 of cytosine. A and G amino protons and NH protons of terminal nucleotides were not observed. [c] Unassigned for free and bound forms. [d] Unassigned for free and bound forms. In the free form, resonances are between 4.13 and 4.22 ppm. [e] Assignment derived from the bound complex, assignment in the free form has an error of ±0.04 ppm. [f] Unassigned for the bound form.

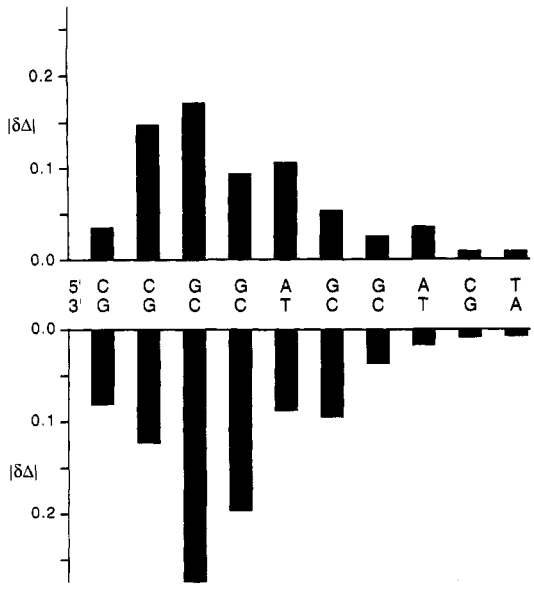

FIGURE 3: Chemical shift changes in ^{1}H NMR resonances of DNA upon addition of GAL4 protein. Total chemical shift changes are summed for all assigned protons of each nucleotide (except 5′ and 5″ protons) and then divided by the number of measurements for each nucleotide (7, 9, 7, and 8 for A, C, G, and T, respectively).

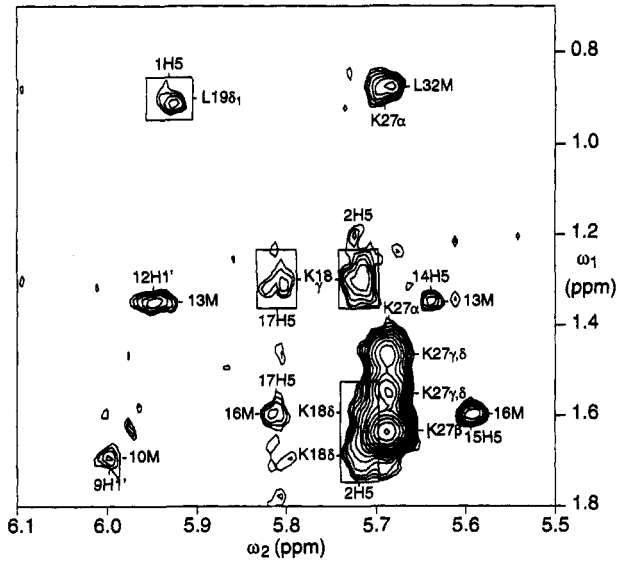

FIGURE 4: Expanded region of a NOESY spectrum in D_2O showing contacts between the DNA and the protein. Labels starting with numerals are DNA; those starting with one-letter amino acid codes are protein. In particular, for the intermolecular NOEs, which are boxed, labels along ω_2 indicate DNA resonances, and labels along ω_1 show protons from amino acid residues of GAL4 protein. The spectrum is recorded for a 0.4 mM solution of the complex at a protein:DNA ratio of 2.31 in 0.1 M NaCl/20 mM phosphate, pH 7.0, 25 °C.

different amounts of protein acts as a chemical shift reagent. ^{1}H NMR assignments were obtained for the individual protein and DNA compnents (each 0% bound), a complex with [GAL4]:[DNA] ratio = 0.86 (corresponding to 57% bound protein and 49% bound DNA), and a complex with [GAL4]:[DNA] ratio = 2.31 (corresponding to 33% bound protein and 76% bound DNA). The different positions for the DNA at the various protein:DNA ratios (0, 49, and 76% bound) solved any overlap problems for the imino, cytosine amino, base H2, H5, H6, and H8, and sugar 1′, 2′, 2″, 3′, and 4′ protons. In addition, a few 5′ and 5″ resonances could be assigned. By extrapolationn to 100% bound, chemical shift assignments were obtained for the DNA within the protein–DNA complex. Chemical shift changes from free DNA are reported in Table 1.

The greatest chemical shift changes in the DNA occur in the nucleotides in the base pair triplet 5′d(CGG)3′·5′d(CCG)-3′ located 1 base pair from one end of the half-site DNA (Figure 3). In the full-site 17 base pair DNA, this corresponds to the 3 base pairs at the extreme ends of the site. From the 11 known GAL4-binding sites, these (and only these) base pairs occur 10 or 11 times at this position, and would therefore appear to be the major determinant of specificity in the recognition of DNA by GAL4 (Marmorstein et al., 1992). In the crystal structure of the GAL4(65)–DNA complex (Marmorstein et al., 1992), the small zinc-binding module lies in the major groove of the DNA in contact with these three base pairs. Some large downfield chemical shift changes are observed for cytosine N4 amino protons on bases 17 and 18 that indicate replacement of the usual hydrogen-bond-accepting water molecules by carbonyls from amino acid residues of the protein (MacMillan et al., 1993). In the case of GAL4, the crystal structure of the complex shows exactly this—the carbonyl of Lys-17 accepts a hyrogen bond from one of the N4 protons on cytosine-18, and the carbonyl of Lys-18 accepts hydrogen bonds from the remaining N4 proton on cytosine-18, as well as the non-Watson–Crick hydrogen-bonded N4 proton on cytosine-17.

After completion of the DNA assignments in D_2O solution, there remain a few cross-peaks in the NOESY spectrum that cannot be explained by intramolecular DNA–DNA or protein–protein cross-peaks. Several of these NOEs are shown in Figure 4. In ω_2, these resonances are from the assigned DNA protons since the resonance position in ω_2 changes between free and bound forms concurrent with the percentage of DNA bound in the two spectra taken at different protein to DNA ratios.

The identities in ω_1 of the extra NOE cross-peaks could not be from protons of the assigned bound DNA and are therefore due to protein. Resonance positions in ω_1 for spectra taken at different protein to DNA ratios were extrapolated back to 0% bound protein. One cross-peak occurs between cytosine-1 H5 and a methyl group of the protein that resonates at 0.87 ppm in the free protein (Figure 4). The protein resonance corresponds to one of the methyl groups of Leu-19 and definitely not one of the other methyl groups in the metal-binding module of GAL4, such as Leu-16 or Leu-32. On the basis of the NMR data alone, however, we cannot rule out that this cross-peak is from a methyl from an amino acid residue in the C-terminal tail region of GAL4. As discussed above, this is unlikely since available crystallographic and genetic data indicate that the metal-binding subdomain (residues 10–40) is the part of the GAL4 that is responsible for sequence-specific recognition. Likewise, protons of Lys-18 are observed to make intermolecular contact to the DNA. In total 11 NOEs are seen between protein and DNA (Table 2). They involve four residues of the protein (A10, K18, L19, and K20) and four residues of the DNA (C1, C2, T16, and C17). All are consistent with the contacts defined from the crystal structure of the GAL4(65)–full-site DNA complex (Marmorstein et al., 1992). Our model for the metal binding of GAL4–half-site DNA complex is presented in Figure 5.

Complete reassignment of protein resonances within the protein–DNA complex was much more difficult than reassigning the DNA resonances. Although many resonances of the complex exhibit fast exchange behavior on the NMR chemical shift time scale, several show substantial broadening. In particular, the amide protons of Lys-17, Lys-18, Leu-19,

Table 2: Observed Intermolecular Contacts between GAL4(65) and Half-Site DNA

GAL4(65)	DNA	GAL4(65)	DNA
Ala-10 β	T_{16} 2′	Lys-18 δ's	C_2 H6
Lys-18 α	C_{17} H5	Lys-18 ϵ's	C_2 H5
Lys-18 β's	C_2 H5	Lys-18 ϵ's	C_2 H6
Lys-18 γ's	C_2 H5	Leu-19 δ_1M[a]	C_1 H5
Lys-18 γ's	C_{17} H5	Lys-20 α[b]	C_{17} H6
Lys-18 δ's	C_2 H5		

[a] Stereospecificity of assignment taken from the crystal structure of the GAL4 protein–full-site DNA complex (Marmorstein et al., 1992). [b] Tentative assignment.

FIGURE 5: Schematic representation for the complex between the DNA-binding domain of GAL4 and DNA. This structural model is adapted from the crystallographic work of Marmorstein (1992) to show half of their dimeric GAL4–DNA complex and is drawn using the MOLSCRIPT program (Kraulis, 1991). The nucleotides are numbered in the DNA duplex. Amino acid residues that interact with DNA are labeled starting with the one-letter amino acid code. The two central metal ions ligated by GAL4 are shown as spheres.

Table 3: Protein ^{1}H Chemical Shift (in ppm) Assignments[a]

	% protein bound			
proton	0.00	0.33	0.57	1.00
Ala-10 β	1.52	1.54	1.54	1.57
Lys-18 α	3.88	3.81	3.72	3.63
Lys-18 β	2.11, 1.85	2.06, 1.96	2.00, 2.00	1.94, 2.13
Lys-18 γ	1.33	1.26	1.21	1.12
Lys-18 δ	1.66, 1.71	1.64, 1.71	*b*	1.58, 1.71
Lys-18 ϵ	3.03	2.98	*b*	2.88
Leu-19 δ_1	0.87	0.88	0.90	0.91
Lys-20 α	4.31	4.29	*b*	4.25

[a] ^{1}H NMR chemical shift assignments are shown for residues of the protein involved in intermolecular NOEs. Assignments for fully bound protein are determined by extrapolation of observed resonance positions at 0, 33, and 57% bound to 100%. [b] Not observable at this protein:DNA ratio in H_2O solution.

and Cys-21 could not be observed, some of which are implicated to be in contact with the DNA from the crystal structure (Marmorstein et al., 1992). Therefore, we report here only resonance chemical shifts for residues of the protein that show intermolecular NOE cross-peaks to DNA (Table 3).

Several intermolecular contacts predicted by the crystal structure of GAL4 with the full-site DNA, which should result in intermolecular NOE cross-peaks, cannot be discerned in our NOESY spectrum. This is most likely due to rapid exchange with solvent or resonance overlap. No intermolecular contacts are seen from the protein to labile protons of the DNA, despite, for example, the close approach of the non-hydrogen-bonded H4 proton on C4 to one of the methyl groups of Leu-19. Presumably, the lifetime of the labile proton of the DNA is too short to allow the buildup of an NOE. Not all expected intermolecular NOEs can be distinguished in crowded spectral regions. For example, we do not see the strong NOE cross-peak that would be predicted for a T16 3′ to Ala-10 methyl group contact. Not only does the T16 3′ resonate near the residual water signal, it also overlaps with the Ala-10 α proton, which would give a strong cross-peak to the Ala-10 methyl group. Other NOEs might be expected to appear, for example, between the backbone 4′ and 5′ protons of G12 and T13 of the DNA and the methyl groups of Leu-49 and Thr-50 of the protein. However, in this case, the backbone protons of the DNA overlap with the α protons of the protein. In addition, although we have not assigned the methyl groups of Leu-49 and Thr-50 for the free, unbound protein, the available assignments (Gadhavi et al., 1990; Pan & Coleman, 1991; Shirakawa et al., 1993) in dicate that they would overlap with other leucine and threonine residues in this part of the protein. Since the protein assignments for intermolecular cross-peaks were made by extrapolation to the unbound form, the identity for the amino acid residue would be ambiguous.

We do see additional stabilization of the protein structure beyond the core residues 10–40. Previously, assignments were obtained for Ile-7, Pro-42, and Lys-43 in free GAL4. These residues gave strong intraresidue TOCSY peaks, but weak NOE peaks, indicating that they are mobile. In the complex, these residues give NOE cross-peaks with intensities comparable to amino acids within the central core of GAL4. In addition, there are now clearly a large number of NOEs from α protons (4.3–4.4 ppm) to methyls in a region of the spectrum that is blank for the free protein. From the chemical shifts published for the N- and C-terminal ends of Gal4(65) (Gadhavi et al., 1990; Pan & Coleman, 1991; Shirakawa et al., 1993), these are due to leucines in the part of the protein that is now more rigid in the protein–DNA complex.

There are several avenues left open to characterize more fully the protein–DNA complex for which we have some initial results and comments. The rate of dissociation of the protein–DNA complex can be determined from chemical exchange broadening. The line widths of DNA resonances change at different ratios of protein to DNA, especially for resonances with large chemical shift differences between free and bound forms. For example, the G3 imino resonance (at 13.12 ppm) is 11.5 Hz wide when 0% bound, is 22.5 Hz wide when 50% bound (the resonance appears at 13.16 ppm), and is 21.5 Hz wide when 75% bound (at 13.25 ppm). Although other line-broadening mechanisms may be taking place, an estimate of the dissociation rate can be obtained as follows. If we assume that the fully bound G3 imino resonance is 22 Hz wide, the natural line width is 16.8 Hz at 50% bound. Therefore, we have 5.7 Hz of line broadening due to chemical exchange, which corresponds to an off rate, k_d, of 3300 s^{-1} (Sykes & Scott, 1972). Using our measured equilibrium constant, K_d, of 160 μM, the on rate, k_a, is 2×10^7 M^{-1} s^{-1}. Although the rate of association is slower than that expected from diffusion control (ca. 10^9 M^{-1} s^{-1}), it is within the range observed for enzyme–substrate and protein–nucleic acid interactions (Fersht, 1985). The GAL4 protein–DNA system may require some desolvation for binding to occur, or the binding could be a multistep process.

The titration data from chemical shift changes in the DNA as GAL4 is added indicate a binding stoichiometry of 1:1. However, this does not rule out other complexes, perhaps of

different stoichiometry, making a minor contribution to the species of molecules in the NMR tube. For example, resonances of the G12 and T13 imino protons broaden at protein:DNA ratios greater than 1 (Figure 1), although some intensity from the terminal imino proton of G20 may be appearing near the G12 resonance. At high protein concentrations (and low temperature), GAL4(65) may be transiently forming dimers, as suggested by the α-helical structure seen for GAL4(62) (Serikawa et al., 1992) that is consistent with the α-helical coiled-coil dimerization element observed in the crystal structure of the GAL4(65)–full-site DNA complex (Marmorstein et al., 1992). Since the spacing between the CGG base pair recognition triplets is critical for GAL4 (S. D. Liang, R. Marmorstein, S. C. Harrison, and M. Ptashne, unpublished results) and since our half-site DNA contains the middle base pair of the 17 base pair binding site, two of these half-sites cannot simultaneously bind to a GAL4 dimer in exactly the same manner as one.

Ideally, we would use a high concentration of protein and DNA to improve the signal to noise ratio of the NOESY spectrum. In addition, labeling the protein with ^{13}C would be useful for distinguishing intermolecular NOEs from intramolecular NOEs in heavily overlapped regions of the NMR spectrum (Ikura & Bax, 1992; Meadows et al., 1993). Given NOE cross-peaks of greater intensity for residues C-terminal to Pro-42 in the GAL4(65)–half-site DNA complex, assignments within the protein–DNA complex could be completed.

The rather weak binding of GAL4(65) to the half-site binding-site DNA results in a rapid rate of dissociation and a short lifetime for any particular protein–DNA complex. Since the lifetime of the complex is short compared to the chemical shift differences between free and bound forms, the complex shows fast exchange behavior. One disadvantage of fast exchange is that the observed NOE is an average between free and bound forms. Another is that a large molar excess of protein must be used to saturate the DNA (and *vice versa*). Finally, the transient nature of the complex may allow other processes to occur when the components are not bound. For example, amide exchange rates for the protein may not be expected to show selective retardation of individual protons upon binding to DNA (Mau et al., 1992). The great advantage of fast exchange is that resonance positions change smoothly as the protein:DNA ratio is altered, which in turn may be used as an aid in chemical shift assignment for the protein–DNA complex.

A general understanding of the way in which proteins interact with DNA has been gained from indirect techniques, such as DNA footprinting, and more detailed pictures have come from crystallographic analysis of protein–DNA complexes (Alber, 1993; Beese et al., 1993; Clark et al.,1993; Ferré-D'Amaré et al., 1993; Harrison, 1991; Kim et al., 1993a,b; Marmorstein et al., 1992; Morikawa, 1993; Pabo & Sauer, 1992; Steitz, 1990; Wolberger, 1993). In recent years, a number of investigations have demonstrated the increasing capability of NMR to examine protein–DNA interactions in solution (Kaptein, 1993). These include the Lac repressor (Chuprina et al., 1993), the glucocorticoid receptor (Remerowski et al., 1991), the *Antp* homeodomain (Otting et al., 1992; Qian et al., 1993), and the GATA-1 DNA-binding domain (Omichinski et al., 1993). This study of the GAL4–DNA complex contributes to progress in the field. We have observed intermolecular NOE cross-peaks confirming the specific interaction of GAL4 with DNA in solution. We have also obtained an accurate measurement of the binding affinity of the GAL4 monomer to a half-site DNA recognition sequence, along with other information concerning the dynamics of the binding reaction. As more powerful NMR techniques become available, similar experiments can be performed on larger GAL4 protein–DNA complexes to provide more insight into the recognition problem.

ACKNOWLEDGMENT

We thank Dr. Thanabal for advice on data collection and maintenance of the NMR spectrometers and Dr. R. Marmorstein for coordinates of the protein–DNA complex and a gift of the 10 base pair duplex DNA. The titration analysis program CRVFIT was kindly provided by R. Boyko and the MOLSCRIPT program by Dr. P. Kraulis.

REFERENCES

Alber, T. (1993) *Curr. Biol. 3*, 182–184.
Baleja, J. D., Marmorstein, R., Harrison, S. C., & Wagner, G. (1992) *Nature 356*, 450–453.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
Beese, L. S., Derbyshire, V., & Steitz, T. A. (1992) *Science 260*, 352–355.
Bram, R. J., & Kornberg, R. D. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 43–47.
Carey, M., Kakidani, H., Leatherwood, J., Mostashari, F., & Ptashne, M. (1989) *J. Mol. Biol. 209*, 423–432.
Chasman, D. I., & Kornberg, R. D. (1990) *Mol. Cell. Biol. 10*, 2916–2923.
Chuprina, V. P., Rullmann, J. A. C., Lamerichs, R. M. J. N., van Boom, J. H., Boelens, R., & Kaptein, R. (1993) *J. Mol. Biol. 234*, 446–462.
Clark, K. L., Halay, E. D., Lai, E., & Burley, S. K. (1993) *Nature 364*, 412–420.
Feigon, J., Wright, J. M., Leupin, W., Denny, W. A., & Kearns, D. R. (1982) *J. Am. Chem. Soc. 104*, 5540–5541.
Ferguson-Miller, S., & Koppenol, W. H. (1981) *Trends Biochem. Sci. 6*, iv–vii.
Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B., & Burley, S. K. (1993) *Nature 363*, 38–45.
Fersht, A. (1985) in *Enzyme structure and mechanism*, Freeman, New York.
Gadhavi, P. L., Raine, A. R. C., Alefounder, P. R., & Laue, E. D. (1990) *FEBS Lett. 276*, 49–53.
Gardner, K. H., Pan, T., Narula, S., Rivera, E., & Coleman, J. E. (1991) *Biochemistry 30*, 11292–11302.
Giniger, E., Varnum, S. M., & Ptashne, M. (1985) *Cell 40*, 767–774.
Gronenborn, A. M., & Clore, G. M. (1985) *Prog. Nucl. Magn. Reson. Spectrosc. 17*, 1–33.
Hansen, A., van Hoy, M., & Kodadek, T. (1991) *Biochem. Biophys. Res. Commun. 175*, 492–499.
Hare, D. R., Wemmer, D. E., Chou, S.-H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319–336.
Harrison, S. C. (1991) *Nature 353*, 715–719.
Hore, P. J. (1983) *J. Magn. Reson. 55*, 283–300.
Hull, W. E. (1975) Ph.D. Thesis, Harvard University, Cambridge, MA.
Ikura, M., & Bax, A. (1992) *J. Am. Chem. Soc. 114*, 2433–2440.
Johnston, M. (1987) *Microbiol. Rev. 51*, 458–476.
Kaptein, R. (1993) *Curr. Opin. Struct. Biol. 3*, 50–56.
Kim, B., & Little, J. W. (1992) *Science 255*, 203–206.
Kim, J. L., Nikolov, D. B., & Burley, S. K. (1993a) *Nature 365*, 520–527.
Kim, Y., Geiger, J. H., Hahn, S., & Sigler, P. B. (1993b) *Nature 365*, 512–520.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.
Kraulis, P. J., Raine, A. R. C., Gadhavi, P. L., & Laue, E. D. (1992) *Nature 356*, 448–450.
Ma, J., & Ptashne, M. (1987) *Cell 51*, 113–119.

MacMillan, A. M., Lee, R. J., & Verdine, G. L. (1993) *J. Am. Chem. Soc. 115*, 4921–4922.
Marmorstein, R., Carey, M., Ptashne, M., & Harrison, S. C. (1992) *Nature 356*, 408–414.
Mau, T., Baleja, J. D., & Wagner, G. W. (1992) *Protein Sci. 1*, 1403–1412.
Meadows, R. P., Nettesheim, D. G., Xu, R. X., Olejniczak, E. T., Petros, A. M., Holzman, T. F., Severin, J., Grubbins, E., Smith, H., & Fesik, S. W. (1993) *Biochemistry 32*, 754–765.
Morikawa, K. (1993) *Curr. Opin. Struct. Biol. 3*, 17–23.
Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J., & Gronenborn, A. M. (1993) *Science 261*, 438–446.
Otting, G., Qian, Y. Q., Billeter, M., Mueller, M., Affolter, M., Gehring, W. J., & Wüthrich, K. (1990) *EMBO J. 9*, 3085–3092.
Pabo, C. O., & Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053–1095.
Pan, T., & Coleman, J. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3145–3149.
Pan, T., & Coleman, J. E. (1991) *Biochemistry 30*, 4212–4222.
Qian, Y. Q., Otting, G., & Wüthrich, K. (1993) *J. Am. Chem. Soc. 115*, 1189–1190.
Remerowski, M. L., Kellenbach, E., Boelens, R., van der Marel, G. A., van Boom, J. H., Maler, B. A., Yamamoto, K. R., & Kaptein, R. (1991) *Biochemistry 30*, 11620–11624.
Sadowski, I., Ma, J., Triezenberg, S., & Ptashne, M. (1988) *Nature 335*, 563–564.
Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., & van Boom, J. H. (1983) *J. Am. Chem. Soc. 105*, 2914–2916.
Serikawa, Y., Shirakawa, M., & Kyogoku, T. (1992) *FEBS Lett. 299*, 205–208.
Shirakawa, M., Fairbrother, W. J., Serikawa, Y., Ohkubo, T., Kyogoku, Y., & Wright, P. E. (1993) *Biochemistry 32*, 2144–2153.
Steitz, T. A. (1990) *Q. Rev. Biophys. 23*, 205–280.
Stone, G., & Sadowski, I. (1993) *EMBO J. 12*, 1375–1385.
Sykes, B. D., & Scott, M. D. (1972) *Annu. Rev. Biophys. Bioeng. 1*, 27–50.
Wolberger, C. (1993) *Curr. Opin. Struct. Biol. 3*, 3–10.
Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, Wiley, New York.