Eur Biophys J (1996) 24: 405–412

© Springer-Verlag 1996

ARTICLE

Paresh Gadhavi · Peter J. Morgan · Peter Alefounder
Stephen E. Harding

A physico-chemical investigation of the self-association of the DNA binding domain of the yeast transcriptional activator GAL 4

Received: 22 January 1996 / Accepted: 24 April 1996

Abstract It has previously been suggested that the DNA binding domain (residues 1 to 147) of the yeast transcriptional activator GAL 4 exists in solution in dimeric form, with the region responsible for dimerisation somewhere between residues 74 and 147. In this study limited proteolysis and carboxy-terminal deletions of the DNA binding domain (residues 1 to 147) of the yeast transcriptional activator GAL 4 followed by subsequent characterization by equilibrium sedimentation in the analytical ultracentrifuge have been used to define more precisely the regions required for DNA binding and protein self-association. Sedimentation equilibrium analyses confirmed that the 'hydrophobic region' of the protein (residues 54–97, which contains a larger proportion of α-helix), is essential for dimerisation, with an apparent dissociation constant $K_{D,app}$, of $\approx$50 μM for the 1–94 residue peptide and $\approx$20 μM for the 1–147 residue peptide. Our studies do not rule out the possible formation of small amounts of additional higher order complexes.

Key words GAL 4 fragments · Apparent dissociation constants

Introduction

The yeast transcriptional activator GAL 4, like many other eukaryotic gene regulatory proteins, consists of separable sequence specific DNA binding and transcriptional activating domains (Ptashne 1988). The DNA binding domain has been localized to the amino terminal 147 residues of the protein, which is sufficient for sequence specific DNA binding to the upstream activation sequence (UAS_G) of the *GAL* genes in *Saccharomyces cerevisiae* (Keegan et al. 1986). Atomic absorption, extended X-ray absorption fine structure (EXAFS) (Povey et al. 1990) and ^{113}Cd NMR (Pan and Coleman 1989) spectroscopic studies of the DNA binding domain of GAL 4 show that it contains two zinc ions per molecule bound via the six cysteines in the sequence (residues 11 to 38). Following a complete assignment of the two dimensional ^{1}H NMR spectra of both the zinc and the cadmium forms of GAL 4, the metal ion co-ordination was deduced for the cadmium substituted form. Two dimensional ^{1}H-^{113}Cd NMR experiments showed that each cadmium ion is co-ordinated by four cysteine residues where two of the cysteines (residues 11 and 28) act as bridging ligands in a Zn_2Cys_6 (C_6) two metal ion cluster (Gadhavi et al. 1991; Gardner et al. 1991). More recently, the complete three dimensional solution structure of the zinc and cadmium forms of GAL 4 were reported (Kraulis et al. 1992; Baleja et al. 1992) which confirmed that the protein belongs to a distinct class of zinc domain protein.

Although the region containing the C_6 two metal ion cluster is required for DNA binding by the GAL 4 protein it is not sufficient; studies have shown that a 14 amino acid sequence adjacent to and carboxy terminal to the cysteine-rich region is required for sequence specific recognition of DNA (Corton and Johnston 1989). In support of this, sequence comparisons between GAL 4 and the LAC 9 transcriptional activator from *Kluyveromyces lactis*, which also binds to the UAS_G in vivo and in vitro, showed marked conservation of residues not only within the cysteine-rich sequence but also in residues adjacent to and carboxy terminal to it (Witte and Dickson 1988). This suggested that both the cysteine-rich zinc binding and the adjacent carboxy terminal sequences act inseparably in DNA binding. These results have now been confirmed and extended following X-ray diffraction analysis of the GAL 4 (1–65) protein-DNA complex (Marmorstein et al. 1992). The palindromic nature of the GAL 4 DNA binding sites (I to IV) in the UAS_G first suggested that the protein may bind to DNA

P. Gadhavi · P. Alefounder
Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, UK

P. J. Morgan · S. E. Harding (✉)
National Centre for Macromolecular Hydrodynamics, Department of Applied Biochemistry and Food Science, University of Nottingham, Sutton Bonington, LE12 5RD, UK
(e-mail: SCZsteve@szn1.nott.ac.uk)

as a dimer. More recently, DNA binding studies, using amino terminal fragments of GAL4 (GAL4 (1–74) and (1–147)) showed that the region between residues 74 and 147 in the DNA binding domain was important for dimerisation (Carey et al. 1989). Similar findings were also reported for LAC9 using a DNA binding fragment (residues 85–228) that formed glutaraldehyde cross-linked dimers in solution and which bound sequence specifically to DNA (Halvorsen et al. 1990). Recent studies have shown that in the C_6 class of fungal transcriptional activator protein, three regions can be distinguished from each other: the zinc-binding region, which provides most of the sequence-specific DNA binding activity, flanked at its C-terminal end by a linker region that in turn connects to a region involved in dimerisation. When each of these regions from the GAL4, PPR1 and PUT3 proteins were interchanged in "domain swap" experiments, and the chimeric proteins were assayed by gel retardation analysis, it was shown that the linker region also imparted a degree of specificity to the association by directing the zinc binding region to its target site (Reece and Ptashne 1993).

In this paper we examine further the intact DNA binding domain of GAL4 (residues 1 to 147); up to now most of the structural studies have focused on the sequence specific DNA binding aspects only. Initially we used limited proteolysis of GAL4 (1–147), both on its own and when complexed to DNA, to define the domain structure. Subsequently we expressed and purified different fragments of GAL4 each containing the cysteine-rich metal binding region (residues 10 to 40), but differing in their carboxy termini. Gel retardation assays were then used to define the regions required for specific DNA binding and dimerisation. Equilibrium sedimentation analysis was then used to estimate the extent and strength of the dimerisation in the various fragments and finally CD spectroscopy of the different fragments suggested that the dimerisation domain contains a larger proportion of α-helix. The results suggest that the DNA binding domain of GAL4 contains a dimerisation region that may be reminiscent of the leucine zipper or helix-loop-helix classes of eukaryotic gene regulatory protein.

Materials and methods

Protein expression and purification

The following fragments of the GAl4 protein were expressed under the control of the *tac* promoter in the Escherichia coli strain JM101: GAL4 (1–49), (1–54), $(1–65)^{+1}$, (1–94) and $(1–147)^{+2}$. The sub-genes encoding the GAL4 $(1–65)^{+1}$, (1–94) and $(1–147)^{+2}$ fragments were kind gifts from Professor M. Ptashne (Harvard). The sub-genes encoding the GAL4 (1–49) and (1–54) fragments were constructed by the introduction of stop codons into the plasmids expressing GAL4 (1–147) using site directed mutagenesis (Amersham). All the constructs were verified by sequencing in M13. One litre cultures were grown in 2TY medium (16 g/l tryptone, 10 g/l yeast extract and 5 g/l NaCl) to an optical density at 600 nm of 0.5–0.8, before addition of isopropyl-β-D-thiogalactopyranoside and zinc acetate to 1 mM each. Cells were harvested after 4–5 hours further growth and the pellets were stored at –70 °C until required. Cell pellets were re-suspended in 40 ml of buffer A (10 mM triethanolamine-hydrochloride pH 7.5, 0.1 M NaCl) and zinc chloride was added to either 0.5 mM (for GAL4 (1–94) and $(1–147)^{+2}$) or 5 mM (for GAL4 (1–49), (1–54), and $(1–65)^{+1}$). Proteinase inhibitors phenylmethylsulphonylfluoride (PMSF) and L-1-chloro-3-[4-tosylamido]-7-amino-2-heptanone HCl (TLCK) were added, prior to and after lysis in a french press, to a final concentration of 2 mM and 100 µg/ml respectively. The lysate was clarified by addition of DNase 1 and $MgCl_2$ to 1 µg/mL and 5 mM respectively, followed by centrifugation at 100,000 *g* for 30 minutes. The supernatant was then loaded directly onto an S-Sepharose (Pharmacia) fast flow column (1.6 cm by 5 cm), and washed with buffer A to remove any unbound protein. The fragment of the GAL4 protein was then eluted using a gradient of 0.1 M to 1.0 M NaCl. Fractions were either pooled and dialysed against buffer A prior to a final purification using a MonoS (HR 5/5) FPLC™ column and a similar gradient or were gel filtered on a Sephacryl S 100 HR (Pharmacia) column. After concentration by ultra-filtration through an Amicon YM2 membrane the protein was dialysed against either 0.5 mM or 5 mM zinc chloride (see above). The samples, over 95% pure as judged by Tricine-SDS-PAGE, were stored at –70 °C until required. Protein concentrations were determined by amino acid analysis from which the molar extinction coefficients at 280 nm were determined (8620 to 8900 M^{-1} cm^{-1} for GAL4 (1–49) to (1–94) respectively).

Limited proteolysis

Limited proteolysis of GAL4 $(1–147)^{+2}$ was carried out in either 10 mM triethanolamine hydrochloride pH 7.5 for bovine α-chymotrypsin (56 units/mg) and trypsin (10000 units/mg) (Sigma) or 50 mM ammonium bicarbonate pH 7.8 for *Staphylococcus aureus* V8 endoproteinase Glu-C (550 units/mg, ICN Laboratories). In each case, the samples also contained 50 mM NaCl, 5 mM $MgCl_2$, and 1 mM $ZnCl_2$. Proteolysis was stopped with the addition of PMSF to 10 mM for trypsin and α-chymotrypsin or glacial acetic acid to 20% (v/v) for V8 endoproteinase. Experiments conducted in the presence of DNA used 20 µM protein, a consensus GAL4 DNA binding site (5′-GGGTCGGAG-TACTGTCCTCCGACTG-3′) and salmon testis DNA (Pharmacia) at 20 µM and 1 mg/mL respectively.

Gel retardation assays

A GAL4 consensus DNA binding site (5′-CGGAA-GACTCTCCTCCG-3′), cloned into the *Bam*H1 site of pUC18, was used for gel retardation assays (Carey et al.

1989). The DNA was labelled at the 5′ end, after cutting the plasmid with *Hin*dIII and treatment with calf intestinal phosphatase (Boehringer), with [γ-^{32}P]ATP (Amersham) and poly-nucleotide kinase (Northumbrian Biologicals). The labelled DNA was then cut again with *Eco*R 1 and purified by preparative PAGE. The concentration of the labelled DNA was determined from the known specific activity of the [γ-^{32}P]ATP (5000 Ci/mmol).

Protein and ^{32}P labelled DNA were diluted to 20 µl in a binding buffer (10 mM Hepes pH 7.0, 100 mM NaCl, 5 mM $MgCl_2$, 20 µM $ZnCl_2$, and 10% (v/v) glycerol). The final mixture contained 10000 cpm of ^{32}P labelled DNA and 0.5 µg of poly-dIdC (Pharmacia). Protein-DNA mixtures were incubated at room temperature for 10 to 20 minutes, and then separated by electrophoresis on an 8% polyacrylamide gel (19 cm by 17 cm by 1 mm) with a Tris-borate (45 mM) pH 8.3 running buffer. The gel was fixed in 25% (v/v) acetic acid and 10% (v/v) methanol and then vacuum dried at 80 °C. The dried gel was autoradiographed on X-ray film (Fuji) for 24 hours at –70 °C using intensifying screens.

Mass spectrometry

Electrospray ionisation mass spectra were recorded using a Fison/VG Bio-Q quadrupole mass spectrometer. The solvent flow consisted of a mixture of 50% (v/v) methanol and 1% (v/v) acetic acid in water, containing 2 µg/mL gramicidin-S as internal standard. The flow rate was 2–4 µl/min and positive ions were detected. Calibrations were performed using PEG 600 for the mass/charge range 400 to 1500 m/z. Each sample, containing 50% (v/v) methanol, 1% (v/v) acetic acid and 20 to 100 pmole/µl of protein, was analysed by 10 µL injections via a loop injector into the source flow. Cone extraction voltages (B 1) used were between 55 and 105 V. All spectra were recorded in multiple scan acquisition mode. Molecular weights were determined using the manufacturer's software.

CD spectroscopy

CD spectra were recorded using a Jobin Y von Dichrograph model CD 6, calibrated with an aqueous solution of (+)-10-camphorsulphonic acid (Aldrich), in 1 mm cells (0.3 ml), scanning between 195 and 260 nm (using a 1 nm spacing, 2 nm bandwidth and a 2s integration time). Each sample (5 to 15 µM) was scanned twice and the resulting spectrum was background corrected by subtraction of a spectrum of the solvent (5 mM $ZnCl_2$).

Sedimentation equilibrium

A Beckman Optima XL-A analytical ultracentrifuge was used for low speed equilibrium sedimentation analysis (Creeth and Harding, 1982). Solute distributions were recorded at equilibrium via their absorption at 278 nm; equilibrium was attained after ~24 hours. The final solute distribution data were analysed on an IBM 3084Q computer using the 'MSTARA' routine (Harding et al. 1992). Weight average molecular weights (M_w) were calculated using the procedures outlined in Creeth and Harding (1982) and partial specific volumes were calculated from the amino acid sequences of the proteins using the "consensus" formula of Perkins (1986).

Other methods

Tricine-SDS-PAGE was carried out as described by Schägger and Von Jagow (1987); gels were stained with Coomassie Brilliant Blue R-250. Protein samples were purified on a C_{18} HPLC column (Brownlee Aquapore GP 18 with 3.2 mm by 1.5 cm guard and 2.1 mm by 3 cm column cartridges), eluting in a gradient of 0–100% acetonitrile containing 0.1% (v/v) trifluoroacetic acid.

Results

Limited proteolysis of the DNA binding domain of GAL 4

A fragment containing the DNA binding domain of GAL 4, GAL 4 $(1-147)^{+2}$, was expressed in *E. coli* and was purified by cation exchange chromatography and gel filtration; the latter step removed a mild contaminating proteolytic activity. Limited proteolysis at 30 °C, in the presence and absence of DNA, using α-chymotrypsin, V 8 endoproteinase Glu-C and trypsin, was then carried out. Analysis by Tricine-SDS-PAGE showed that, in the absence of DNA, both α-chymotrypsin and trypsin (each at 1% (w/w) enzyme to protein) gave fragments of between 5 and 4 kDa, respectively (see Fig. 1 a). In the presence of the GAL 4 DNA binding site and an excess of salmon testis DNA (3 : 1 (w/w) DNA to protein) the production of these smaller fragments was almost completely inhibited, resulting in larger fragments of ~13 kDa (see Fig. 1 b). Interestingly, for α-chymotrypsin, protection against proteolysis was noticeably more complete in the presence of the specific DNA binding site. The major proteolytic fragments were purified by reversed phase HPLC and were identified by amino terminal sequencing, amino acid analysis and electrospray ionization mass spectrometry. Both α-chymotrypsin and trypsin cut just before (at Leu_4 and Lys_2, respectively) and just after (at Leu_{49} and Lys_{43}, respectively) the zinc binding region (see Fig. 2). Partial cleavage sites were also identified, for example at Lys_{45}, Arg_{46} and Arg_{51} for trypsin and at Leu_3 for α-chymotrypsin. Away from the zinc binding region, other cleavage sites were identified between residues 97 and 122 releasing a number of hydrophobic polypeptides corresponding to GAL 4 sequences between residues 49 and 122. These peptides, which are largely insoluble at neutral pH, were also identified after purification by reversed phase HPLC.

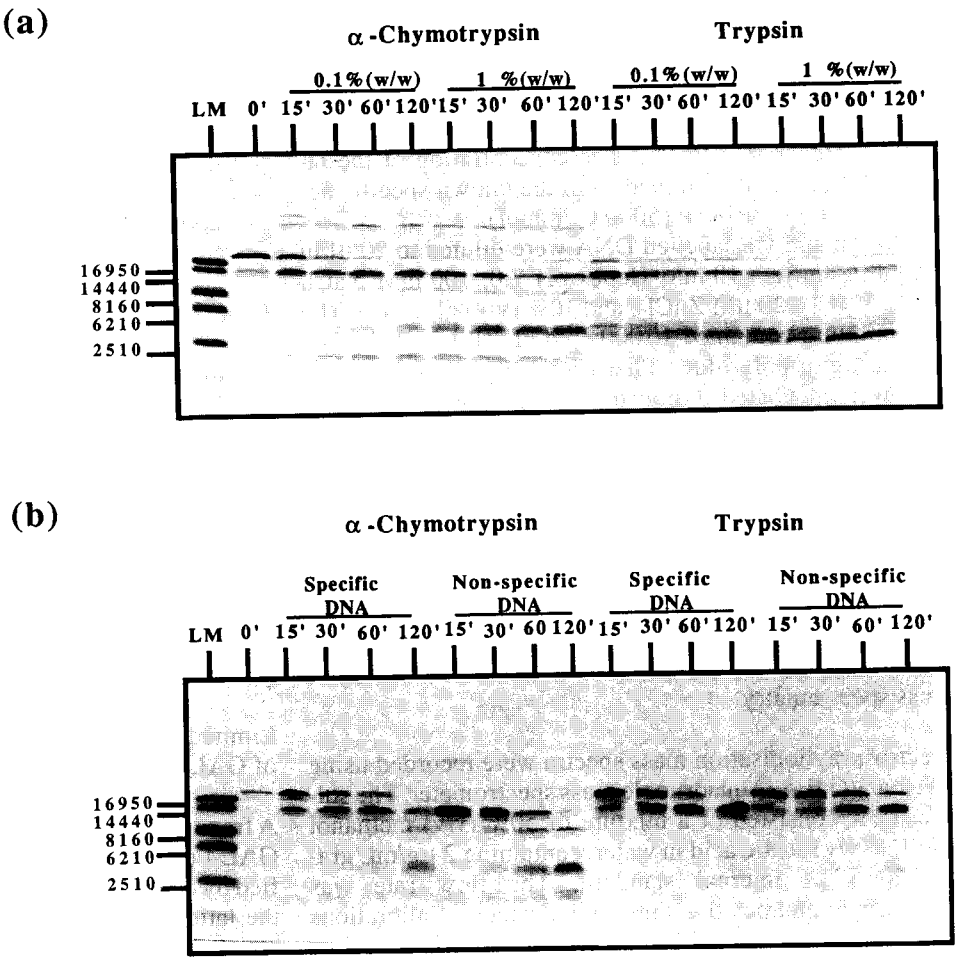


Fig. 1 Limited proteolysis of GAL4 $(1-147)^{+2}$ at 30 °C in the absence (**a**) and presence (**b**) of DNA. Each lane contained 100 pmols (≈1.7 μg) of protein. In (**b**), reaction mixtures containing the GAL4 DNA binding site also contained 0.5 mg/ml of salmon testis DNA. Samples were withdrawn at the time indicated. The reaction was stopped by the addition of PMSF to 10 mM after which the solution was mixed and boiled with an equal volume of 4% (w/v) SDS PAGE sample buffer and analysed by 4–16% (T) Tricine-SDS-PAGE. The lane marked LM contained 5 μg of low molecular weight protein markers (Sigma)

A summary of these analyses, in the form of a proteolytic map, is shown in Fig. 2. The results suggest that the protein contains at least two structured regions between residues 5 and 43 and residues 52 and 97. They show that the region of the protein between residues 43 and 51 is protected when the protein binds DNA (see Fig. 1 b).

Residues 1 to 54 of GAL4 are required for sequence specific DNA binding activity

The results of the limited proteolysis experiments enabled the rational expression and purification of a number of GAL4 fragments, each containing the cysteine-rich zinc binding region but differing in their carboxy termini. Four different fragments, GAL4 (1–49), (1–54), $(1-65)^{+1}$ and (1–94) were expressed and assayed for sequence specific DNA binding activity.

Gel retardation assays showed that both GAL4 (1–54) and $(1-65)^{+1}$ formed specific protein-DNA complexes whereas GAL4 (1–49) did not (see Fig. 3 a). Although both GAL4 (1–54) and $(1-65)^{+1}$ bound sequence specifically, the affinity of binding was markedly lower than that seen for either GAL4 (1–94) or $(1-147)^{+2}$, which appear

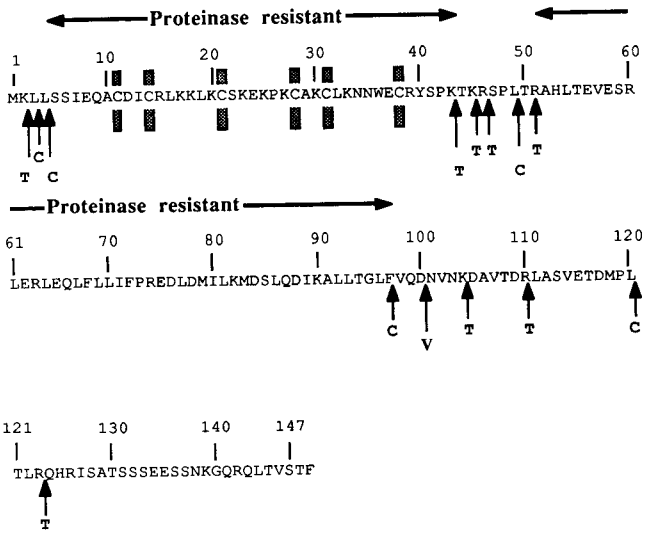


Fig. 2 A proteolytic map of the DNA binding domain of GAL4, GAL4 $(1-147)^{+2}$, using α-chymotrypsin (C), V 8 endoproteinase Glu-C (V) and trypsin (T) in the absence of DNA. The long arrows mark the limit cleavage sites (observed at 120 minutes) while the shorter arrows mark the partial cleavage sites (observed between 15 and 60 minutes) during the digestion. Cleavage sites were deduced using amino acid analyses, amino terminal sequencing and electrospray ionization mass spectrometry of HPLC purified samples

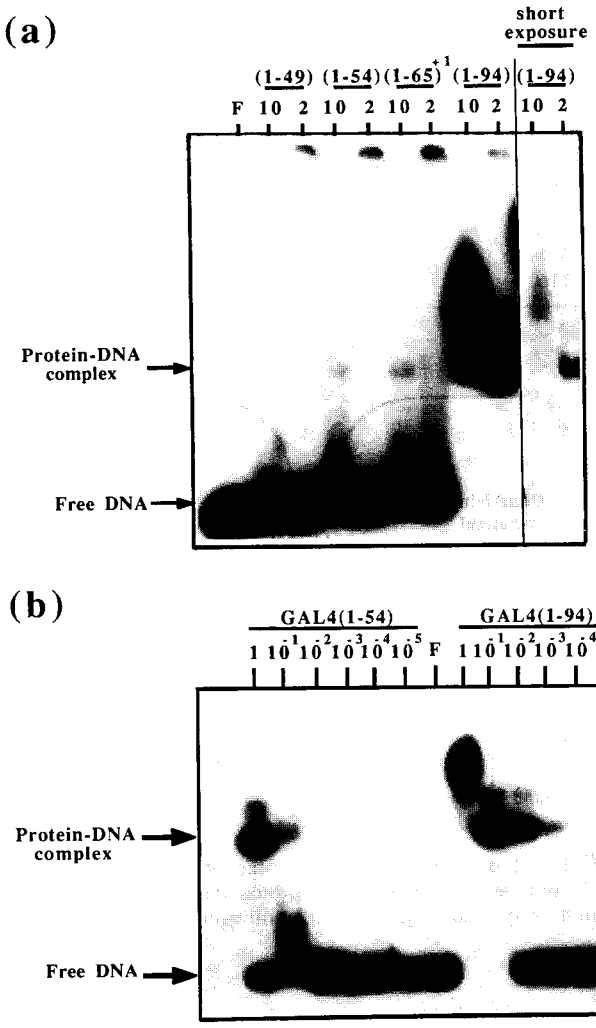


Fig. 3 Gel retardation assays of (**a**) the different amino terminal fragments of the DNA binding domain of GAL 4 (the numbers above the lanes indicate the protein concentration in μM in the binding reactions) and (**b**) of GAL 4 (1 – 54) and (1 – 94) at various protein concentrations. In the latter the numbers above each lane indicate the dilution of the protein relative to the starting concentration (25 μM). Reactions were analysed by 8% non-denaturing PAGE with a 45 mM Tris-borate pH 8.3 buffer. Lanes marked F contain free DNA only. DNA loading concentration throughout ≈5 nM

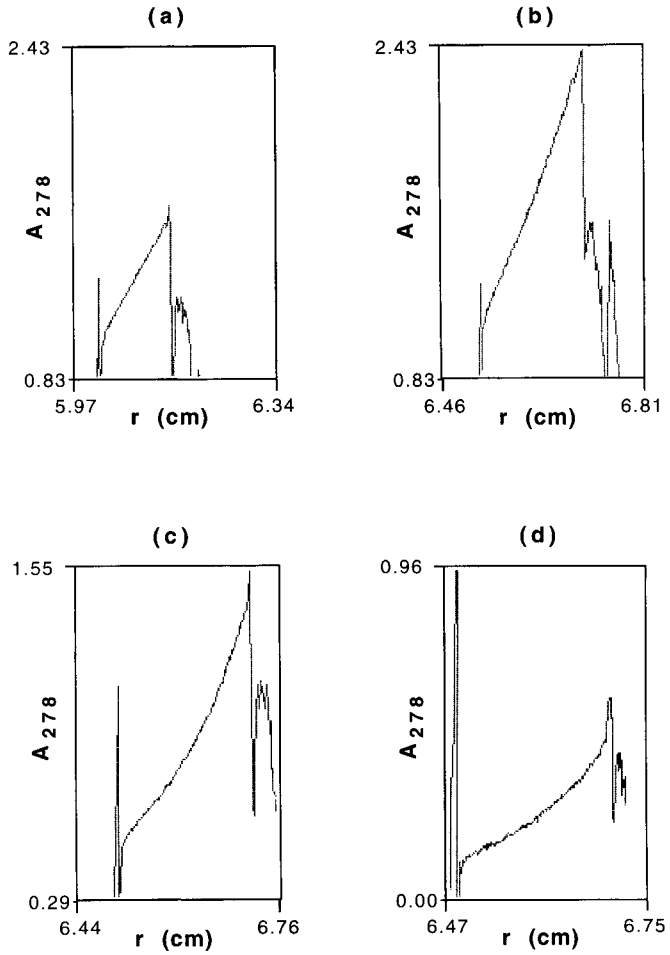


Fig. 4 The solute distribution at equilibrium (recorded using absorption optics at 278 nm and at 20 °C with a 12 mm optical path length cell and ≈0.2 – 0.3 cm solution column lengths) of (**a**) GAL 4 (1 – 54) at a loading concentration of ~100 μM and 28,000 rpm (**b**) GAL 4 $(1-65)^{+1}$ at a loading concentration of ~100 μM and 28,000 rpm, (**c**) GAL 4 (1 – 94) at a loading concentration of 60 μM and 20,000 rpm and (**d**) GAL 4 $(1-147)^{+2}$ at a loading concentration of 24 μM and 18,000 rpm

to bind with very similar affinity (data not shown). The relative difference in binding affinities between GAL 4 (1 – 54) and (1 – 94) was quantified in a protein titration study, which showed that GAL 4 (1 – 94) formed a specific protein-DNA complex at a hundred-fold lower concentration of protein (see Fig. 3 b).

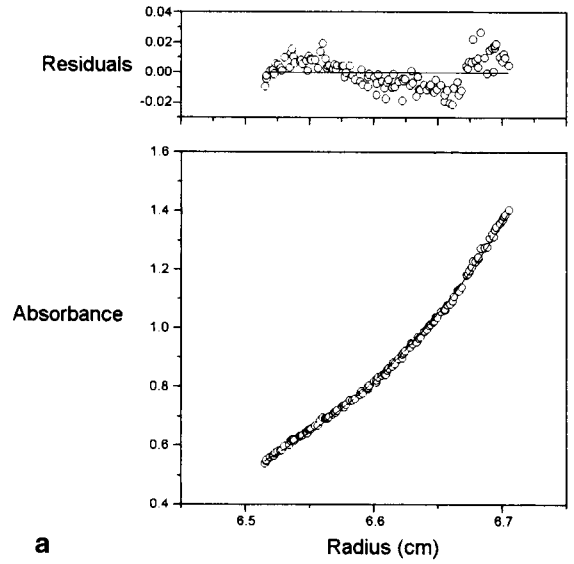
The GAL 4 protein contains a dimerisation domain between residues 65 and 94

The gel retardation assays of the various GAL 4 fragments suggested that the sequence between residues 65 and 94 is responsible for increasing the affinity rather than the specificity of binding. Previous studies have shown that the GAL 4 protein contains a dimerisation domain somewhere between residues 74 and 147 (Carey et al. 1989); we suspected, therefore, that the structured region (defined by limited proteolysis) between residues 52 to 97 might constitute this domain.

Molecular weight determinations using low speed equilibrium sedimentation analysis were carried out for each protein fragment in solution and Fig. 4 compares the equilibrium solute distributions recorded using the absorption optical system at 280 nm. The smaller GAL 4 fragments (GAL 4 (1 – 54) and $(1-65)^{+1}$) are clearly monomeric in solution (Table 1). In marked contrast, the larger fragments (GAL 4 (1 – 94) and $(1-147)^{+2}$) give weight average molecular weights between the monomeric (M_1) values – as evaluated from mass spectrometry – and the expected values for a dimer in both cases. The simplest interpretation of the data would therefore suggest that these larger frag-

Table 1 Monomer and weight average molecular weight measurements for the four amino-terminal fragments of GAL4. Monomer molecular weight (M_1) values were determined by electrospray ionization mass spectrometry using a Fison/VG Bio-Q mass spectrometer. Absolute, weight averaged, molecular weight (M_w) values were determined by low speed sedimentation equilibrium analysis for proteins in 1 to 5 mM zinc chloride solution using a Beckman Optima XL-A analytical ultracentrifuge. Partial specific volumes ($\bar{v}$) were calculated from the amino acid sequences of the fragments using the consensus formula of Perkins (1986)

Fragment	c (mg/ml)	$\bar{v}$ (ml/g)	M_1 (g/mol)	M_w (g/mol)	$K_{2,app}$ (M^{-1})
Zn-GAL 4 (1 – 54)	0.63	0.746	6308	6200 ± 1000	0
Zn-GAL 4 (1 – 65)	0.78	0.744	7798	6700 ± 1000	0
Zn-GAL 4 (1 – 94)	0.66	0.754	11 061	17 000 ± 1000	2×10^4
Zn-GAL 4 (1 – 147)	0.41	0.743	17 116	26 000 ± 1000	5×10^4



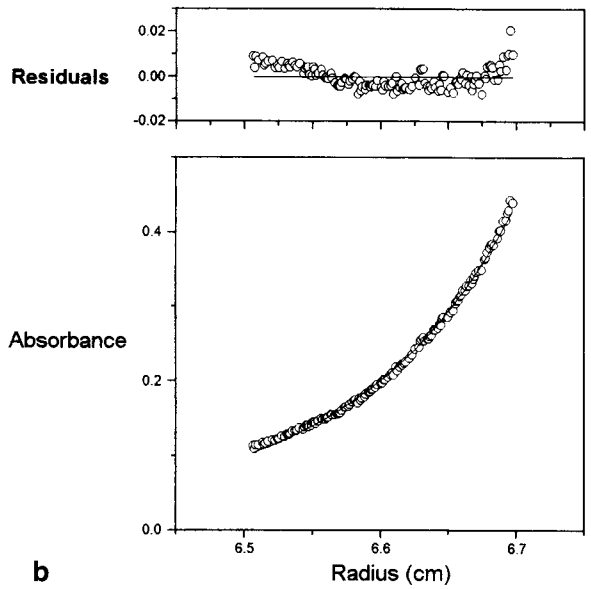


Fig. 5 Association equilibria from sedimentation equilibrium optical records for (**a**) GAL4 (1–94) at a loading concentration of 60 μM and 20,000 rpm and (**b**) GAL4 $(1-147)^{+2}$ at a loading concentration of 24 μM and 18,000 rpm. Lines fitted (not easily seen – coincidental with the experimental points) are for a monomer-dimer association equilibrium fitted according to eq. III27 of Kim et al. (1977). Apparent dissociation constants (**a**): ≈50 μM and (**b**) ≈20 μM

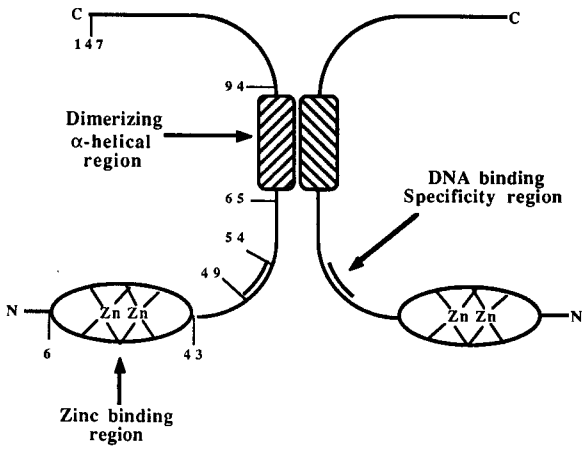


Fig. 6 Possible schematic structure of the DNA binding domain of GAL4 from the analysis of limited proteolysis experiments, gel retardation assays and sedimentation equilibrium studies

ments exist as an equilibrium between a monomer and a dimer, although the possible presence of smaller amounts of trimers or higher oligomers cannot be excluded from these data alone.

The strengths of the dimerisation can best be described in terms of apparent dimerisation constants (or the corresponding molar dissociation constants). If we make the reasonable assumption that the concentrations, c are small enough so that non-ideality effects are not significant, the apparent dimerisation constant (on a weight concentration scale) $k_{2,app}$ can be estimated from (Deonier and Williams 1970)

$$k_{2,app} = \{1/4c\}\ \{[R^2/(2\text{-}R)^2] - 1) \qquad (1)$$

where $R = M_w/M_1$, and with the apparent molar dimerisation constant given by $K_{2,app} = (M_1/2).k_{2,app}$. The values obtained ($2\times10^4$ M^{-1} and 5×10^4 M^{-1}) for the 1–94 and 1–147 fragments are in good agreement with values obtained (≈1.5×10^4 M^{-1} and ≈3.6×10^4 M^{-1}) from direct modelling of the concentration distribution data according to Eq. (III27) of Kim et al. (1977) (Fig. 5). In terms of apparent molar dissociation constants, $K_{D,app}$ these values are ≈50 μM and ≈20 μM for the 1–94 and 1–147 fragments respectively. As before, this analysis does not exclude the possible presence of small amounts of trimer or higher or-

der association products, and this may account for the small deviation in the residuals of Fig. 5.

Discussion

Many gene regulatory proteins consist of distinct structural elements, or 'motifs', which can recognise and interact with either nucleic acids or other proteins. GAL 4, a yeast transcriptional activator, is no exception. The protein contains several independent functional domains, one for sequence specific DNA binding and others for activation of transcription, (Ptashne 1988).

In this paper we show that the DNA binding domain (residues 1 to 147) can be subdivided further using limited proteolysis and deletion analysis. The results show that the domain consists of two regions resistant to proteolysis, one between residues 5 and 43, containing the two metal ion cluster, and the other, between residues 52 and 97, that is markedly hydrophobic. It is interesting to note that despite the large number of lysine (9 out of 43) and arginine residues (3 out of 43) in the metal ion cluster, no tryptic cleavage within this region was observed; the three dimensional structure of GAL 4 (10–40) shows that this region is highly structured in solution (Kraulis et al. 1992; Baleja et al. 1992). The sequence connecting the two proteinase resistant regions was found to overlap a sequence that had earlier been shown to be important for sequence specific recognition of DNA (Corton and Johnston 1989; Reece and Ptashne 1993). Consistent with this, in the presence of both specific and non-specific DNA, proteolytic cleavages were not observed within this sequence. Furthermore, gel retardation assays of various carboxy terminal deletion fragments showed that residues 50 to 54 are required for providing sequence specific DNA binding activity, confirming previous work which suggested that residues 48 to 56, which are completely conserved between GAL 4 and LAC 9, are involved (Witte and Dickson 1988; Corton and Johnston 1989). These results are also consistent with the crystal structure of the GAL 4 (1–65) protein-DNA complex which shows that this extended region is in close contact with the DNA (Marmorstein et al. 1992).

Even though the amino terminal 54 residues are sufficient for sequence specific DNA binding, our studies also showed that the region between residues 65 and 94 contributes to the overall affinity of binding (by a factor of ~100). Results from equilibrium sedimentation analyses show that both GAL 4 (1–94) and $(1-147)^{+2}$ dimerise to similar extents and indicate that the sequence between residues 65 and 94 in GAL 4 is required for this process. Clearly, the presence of residues 65 to 94 enhances the ability of the protein to dimerise in the absence of DNA, perhaps by stabilising the overall structural integrity of the region between residues 52 to 97. This would appear to be consistent with an increase in the proportion of α-helical content of the 1–94 and 1–147 fragments: the smaller 65 residue peptide is known from NMR structural studies (Kraulis et al. 1992; Baleja et al. 1992) to have ≈20% α-helix (with little secondary structure in the region of residues 49 to 65): comparative CD spectra have indicated more pronounced minima at 208 and 220 nm (indicative of greater α-helical content [Johnson 1990]) for the longer fragments.

The absence of secondary structure in the region of residues 49–65 in the absence of DNA contrasts with the situation in the presence of DNA where this region becomes structured; residues 40–49 form an extended region whilst residues 50–65 form the beginnings of an α-helical coiled coil dimerisation element (Marmorstein et al. 1992). The fact that in GAL 4 (1–94) no proteolytic cut sites are observed between residues 52 and 65 indicates that this region is structured even in the absence of DNA. It appears that the structure of the dimerisation element seen in the crystal structure of the GAL 4 (1–65) protein-DNA complex (residues 50–65) can be stabilised either by binding to DNA or by the addition of the remainder of the dimerisation domain (residues 65 to 97).

The presence of a dimerisation domain adjacent to the DNA binding region in a gene regulatory protein is also seen in a variety of other proteins, particularly those containing the 'basic-leucine zipper' (bZIP) motif or 'basic helix-loop-helix' motif. For example GCN 4, a yeast transcriptional activator, forms a homodimer via a leucine-rich, and α-helical, polypeptide which forms a coiled coil. The helix-loop-helix proteins form a parallel four helix bundle. In each case dimerisation allows the amino terminal DNA binding elements (basic motif) to make sequence specific contacts with the DNA. For both classes of protein, however, removal of the leucine zipper destroys the ability of the protein to bind to its recognition sequence because a stable interaction with DNA requires a dimer (for a recent review, see Ellenberger, 1994). Although removal of the dimerisation region in GAL 4 did not abolish DNA binding, the affinity was reduced a hundred-fold, indicating that the region stabilises the interaction between the protein and the DNA. In a sense, the zinc binding region (residues 5 to 43, defined by proteolysis) flanked by a highly flexible stretch of polypeptide may be analogous to the 'basic motif' in the leucine zipper or helix-loop-helix proteins (see, e. g. Wendt et al. 1995). It is followed by a region containing a significant proportion of α-helix and hydrophobic residues, which is responsible for dimerisation: values for the apparent dissociation constants (≈50 μM and ≈20 μM for GAL 4 fragments 1–94 and 1–147 respectively) suggest however a weaker interaction compared with other coiled-coil systems (Ellenberger 1994; Wendt et al. 1995). Despite this, the binding is still sufficient to increase the affinity for DNA ~100 fold, and it could well be that any stronger protein dimerisation would be largely redundant.

In conclusion, the sequence specific contacts between GAL 4 and the DNA binding site appear to be due both to the proteolytically resistant metal binding region (residues 5 to 43) and to a proteolytically labile linker region (residues 43 to 54). The region comprising residues 52 to 97, contains a larger α-helical content, facilitates dimer formation thus markedly increasing the affinity of binding.

A possible schematic model for this is shown in Fig. 6. More precise knowledge of the structure of this region will however require X-ray crystallographic or NMR studies.

Acknowledgements We would like to thank Ernest Laue, Jean Thomas, Caroline Hill, Len Packman, Phil Evans and Jane Povey for their help and useful discussions during the course of this work and Juliette Jacoby for amino acid analyses. We also thank Professor Mark Ptashne (Harvard University) for the kind and generous gifts of expression plasmids encoding GAL4 $(1-65)^{+1}$, (1–94) and $(1-147)^{+2}$. This work was supported by the Science and Engineering Research Council (now Biotechnology and Biological Sciences Research Council) and its Cambridge Centre for Molecular Recognition, and the Engineering and Physical Science Research Council.

References

Baleja JD, Marmorstein R, Harrison SC, Wagner G (1992) Solution structure of the DNA-binding domain of CD2-GAL4 from *Saccharomyces cerevisae*. Nature 356: 450–453

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

Corton JC, Johnson SA (1989) Altering DNA-binding specificity of GAL4 requires sequences adjacent to the zinc finger. Nature 340: 724–727

Creeth JM, Harding SE (1982) Some observations on a new type of point average molecular weight. J Biochem Biophys Meth 7: 25–34

Deonier RC, Williams JW (1970) Self-association of muramidase (lysozyme) in solution at 25°, pH 7.0, and I = 0.20. Biochemistry 9: 4260–4267

Ellenberger T (1994) Getting a grip on DNA recognition – Structure of the basic region leucine-zipper, and the basic region helix-loop-helix DNA-binding domains. Curr Op Structural Biology 4: 12–21

Gadhavi PL, Raine AR, Alefounder PR, Laue ED (1990) Complete assignment of the ^{1}H NMR spectrum and secondary structure of the DNA binding domain of GAL4. FEBS Lett 276: 49–53

Gadhavi PL, Davis AL, Povey JF, Keeler J, Laue ED (1991) Polypeptide-metal cluster connectivities in Cd(II) GAL4. FEBS Lett 281: 223–226

Gardner KH, Pan T, Narula S, Rivera E, Coleman JE (1991) Structure of the binuclear metal-binding site in the GAL4 transcriptional factor. Biochemistry 30: 11292–11302

Halvorsen Y-DC, Nandabalan K, Dickson R (1990) LAC9 DNA-binding domain coordinates 2 zinc atoms per monomer and contacts DNA as a dimer. J Biol Chem 265: 13283–13289

Harding SE, Horton JC, Morgan PJ (1992) MSTAR: A FORTAN program for the model independent molecular weight analysis of macromolecules using low speed or high speed sedimentation equilibrium. In: Harding SE, Rowe AJ and Horton JC (eds) Chapter 15, Analytical ultracentrifugation in Biochemistry and Polymer Science, Royal Society of Chemistry, UK

Johnson WC (1990) Protein secondary structure and circular dichroism – a practical guide. Proteins Struct Function Gen 7: 205–214

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

Kim H, Deonier RC, Williams JW (1977) The investigation of self-association reactions by equilibrium ultracentrifugation. Chem Rev 77: 659–690

Kraulis PJ, Raine ARC, Gadhavi PL, Laue ED (1992) Structure of the DNA-binding domain of zinc GAL4. Nature 356: 448–450

Marmorstein R, Carey M, Ptashne M, Harrison SC (1992) DNA recognition by GAL4 – Structure of a protein DNA complex. Nature 356: 408–414

Pan T, Coleman JE (1989) Structure and function of the Zn(II) binding-site within the DNA-binding domain of the GAL4 transcription factor. Proc Nat Acad Sci USA 86: 3145–3149

Perkins SJ (1986) Protein volumes and hydration effects. The calculations of partial specific volumes, neutron-scattering match-points and 280 nm-absorption coefficients for proteins and glycoproteins from amino-acid sequences. Eur J Biochem 157: 169–180

Povey JF, Diakun GP, Garner CD, Wilson SP, Laue ED (1990) Metal-ion coordination in the DNA-binding domain of the yeast transcriptional activator GAL4. FEBS Lett 266: 142–146

Ptashne M (1988) How eukaryotic transcriptional activators work. Nature 335: 683–689

Reece RJ, Ptashne M (1993) Determinants of binding-site specificity among yeast C6 zinc cluster proteins. Science 261: 909–911

Schägger H, Von Jagow G (1987) Tricine sodium dodecyl-sulfate polyacrylamide-gel electrophoresis for the separation of proteins in the range from 1-kDa to 100-kDa. Analyt Bioch 166: 368–379

Wendt H, Berger C, Baici A, Thomas RM, Bosshard HR (1995) Kinetics of folding of leucine-zipper domains. Biochemistry 34: 4097–4107

Witte M, Dickson R (1988) Mol Cell Biol 8: 3726–3733