Volume 299, number 3, 205–208 FEBS 10791 March 1992

Determination of the DNA binding site of the GAL4 protein

A photo-CIDNP study

Yasuharu Serikawa, Masahiro Shirakawa and Yoshimasa Kyogoku

Institute for Protein Research, Osaka University, Suita, Osaka 565, Japan

Received 6 January 1992

After the assignment of ^{1}H NMR signals of aromatic side chains by means of specific deuteration, we analyzed the DNA binding site of GAL4 by measuring photo-CIDNP spectra. The results showed that Trp^{36} is involved in both the specific interaction with UAS_G and non-specific DNA binding. This residue is located inside the Cys-rich region, but outside the putative Zn-finger. The photo-CIDNP spectrum also showed that the side chains of Tyr^{40} and His^{53} are not exposed on the surface of the protein.

NMR; Photo-CIDNP; GAL4; UAS_G; Zinc finger; DNA binding

1. INTRODUCTION

GAL4, a transcriptional activator protein that binds specifically to 17 base-pair DNA sequences named UAS_G [1], is responsible for induction of the genes coding for galactose-metabolizing enzymes of *Saccharomyces cerevisiae* [2]. Pan and Coleman [3] reported the production of a DNA binding domain, GAL4(149*), comprising residues 1–147 plus two additional residues, and Serikawa et al. [4] were successful in producing GAL4 fragments of 147, 110 and 96 amino-terminal residues that could be expressed in *E. coli*. All of them contain a stoichiometric amount of Zn(II) and specifically bind to UAS_G.

The DNA binding domain of GAL4 has a Cys-rich region (residues 11–38), there being six cysteine residues [5], and it had long been supposed that the domain contained a cysteine–zinc finger comprising a zinc ion bound between two pairs of cysteine residues [6]. However, Pan and Coleman proposed that the DNA binding domain does not contain such a zinc finger but forms a $Zn(II)_2Cys_6$ binuclear cluster [7,8], although coordination positions were revised later [9]. A similar cluster model was proposed by Gadhavi et al. [10]. From the results of replacement of amino acid segments of the DNA binding domain by means of in vitro mutagenesis with analogous segments of another transcriptional activator, PPR1, Corton and Johnson assumed that the 14-amino acid sequence adjacent to the zinc finger is required for the DNA sequence recognition specificity [11]. However, there is no direct physicochemical evidence for the involvement of any specific amino acid residue for or region of the DNA binding domain in the complex formation with UAS_G.

Here we prepared protein fragments, GAL4(110) and GAL4(147) corresponding to N-terminal regions of the 110 and 147 residues, respectively [4]. After the assignment of ^{1}H NMR signals of aromatic side chains by means of specific deuteration, we used the photo-CIDNP (chemically induced dynamic nuclear polarization) method [12] to clarify whether tryptophan, tyrosine and histidine residues of the DNA binding domain of GAL4 are involved in the DNA binding. GAL4(110) contains one histidine, one tryptophan, one tyrosine and three phenylalanines as aromatic residues. Among them, Trp^{36} is located within the Cys-cluster (residues 11–38), but Tyr and His are not.

2. MATERIALS AND METHODS

2.1. *Protein preparations*

Non-labeled GAL4(110) and GAL4(147) were expressed in *E. coli* strain TG1 cells harboring pGT110 and pGT147, respectively [4], which were grown in YT broth. For the preparation of deuterated amino acid-substituted GAL4(110), *E. coli* strain KMBL1788 (thyA301, bio87, argA, endA101, pheA, metE72) was used as a host for pGN110 [4] because of its requirement for phenylalanine. (1,2,3,4,5-$^2H_5$)Phenylalanine and (3,5-$^2H_5$)tyrosine were purchased from Cambridge Isotope Co. M9-Glucose containing 20 amino acids, and 50 mg/l each of thymine and biotin was used as a medium. For the deuteration of phenylalanine residues of GAL4(110), 50 mg/l of (1,2,3,4,5-$^2H_5$)-L-phenylalanine was added instead of its non-deuterated form, and for the deuteration of both phenylalanine and tyrosine residues, (2,4-$^2H_2$)-L-tyrosine was used together with deuterated phenylalanine. The purification procedure was described previously [4].

2.2. *NMR measurements*

^{1}H NMR spectra at 500 MHz were recorded with a JEOL GX-500s

Correspondence address: Y. Kyogoku, Institute for Protein Research, Osaka University, Suita, Osaka 565, Japan.

spectrometer equipped with a 5 mm sample probe using 16,384 data points and a spectral width of 6,000 Hz. Proton chemical shifts are expressed relative to an internal reference, 2,2-dimethyl-2-silapentane sulfonate (DSS). ^{1}H photo-CIDNP spectra were measured for a 1 mM protein solution in 0.5 ml of 20 mM phosphate buffer and 0.3 M KCl at p^{2}H 7.0. Spectra were recorded using a specially designed probe with a 3 mm quartz rod to introduce laser light. The sample tube was irradiated in the probe for 0.3 s with 488 nm line (0.7 w) from an NEC GLG-3300 argon ion laser for data acquisition. Alternately, 'light' and 'dark' free induction decays were collected and subtraction yielded the photo-CIDNP difference spectrum. A light–dark cycle took 40 s and 32 cycles were accumulated.

3. RESULTS AND DISCUSSION

First, we assigned the aromatic resonances by means of selective deuteration of phenylalanine and tyrosine side chains, pH titration and NOE (nuclear Overhauser effect) measurements. The aromatic region of the NMR spectra of non-deuterated GAL4(110) and (1,2,3,4,5-$^2H_5$) phenylalanine-substituted GAL4(110) are shown in Fig. 1a and b, respectively. GAL4(110) contains three

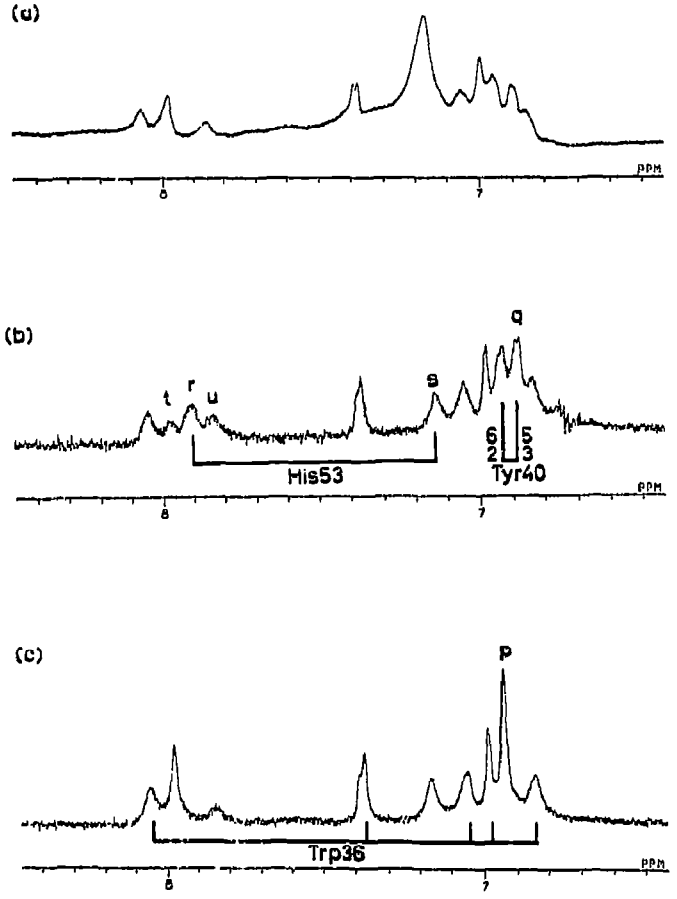

Fig. 1. 500 MHz ^{1}H NMR spectra of aromatic protons of GAL4(110). (a) Intact GAL4(110). (b) (1,2,3,4,5-$^2H_5$) phenylalanine-substituted GAL4(110). (c) (1,2,3,4,5-$^2H_5$)-phenylalanine, (3,5-$^2H_2$)tyrosine-substituted GAL4(110). The solution comprised 4 mg of each GAL4(110) in a ^2H_2O solution containing 20 mM phosphate buffer and 300 mM KCl, p^{2}H 7.0. Proton chemical shifts were recorded relative to an internal reference, 2,2-dimethyl-2-silapentane sulfonate (DSS).

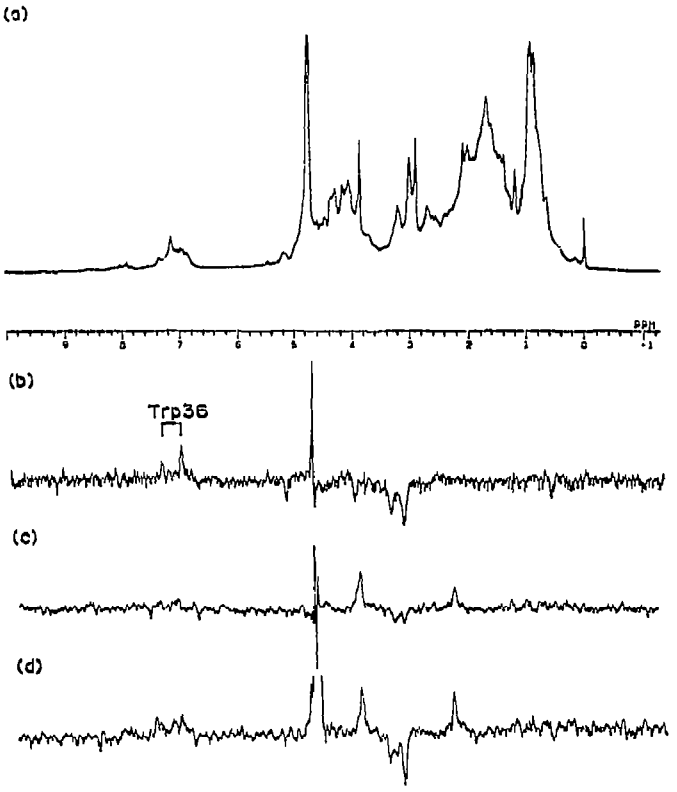

Fig. 2. Photo-CIDNP spectra of GAL4(110) and its complex with the UAS_G 17-mer oligonucleotide. Each solution comprised 1 mM GAL4(110) in 20 mM phosphate buffer, p^{2}H 7.0. (a) Dark spectrum and (b) photo-CIDNP difference spectrum of GAL4(110) itself in 0.3 M KCl, (c) 1:1 GAL4(110) dimer/UAS_G in 0.3 M KCl, and (d) 1:1 GAL4(110) dimer/UAS_G in 1.0 M KCl.

phenylalanine residues, at positions 65, 68 and 97. Some of the signals due to phenylalanine residues overlap at 7.2 ppm, as shown in Fig. 1a. Elimination of the signals of the ring protons of the phenylalanine residues resulted in much better separation in the aromatic region (6.5–8.5 ppm) of the spectrum of GAL4(110) (Fig. 1b). The resonance assignment of the unique tyrosine residue at position 40 was achieved by measuring a spectrum for (1,2,3,4,5-$^2H_5$)phenylalanine- and (3,5-$^2H_2$)tyrosine-substituted GAL4(110) (Fig. 1c). The signal denoted by q in Fig. 1b could be clearly assigned to 3,5-proton resonances of the unique tyrosine residue, Tyr^{40}, since it was missing in Fig. 1c. Because of the elimination of scalar coupling, and dipole interaction between 3,5-protons and 2,6-protons of Tyr^{40}, the resonance of 2,6-protons of Tyr^{40} at 6.93 ppm, denoted by p in Fig. 1c, became singlet and sharper than that seen in Fig. 1b. Identification of the C-2 and C-4 proton resonances of the unique histidine residue, His^{53}, was straightforward on p^{2}H titration (data not shown). Two signals assigned to His^{53} (denoted by r and s in Fig. 1b) gave a pH titration curve indicating the pKa value to be 6.7. The residual peaks at 8.05, 7.38, 7.06, 6.98 and 6.85 ppm are connected by NOE assigned to the unique Trp

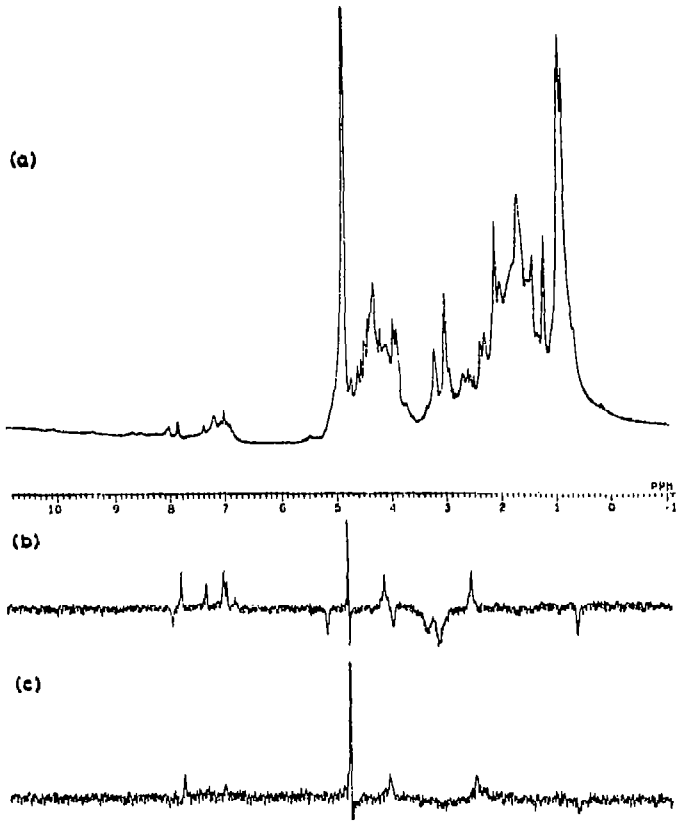

Fig. 3. Photo-CIDNP spectra of GAL4(147) and its complex with the CRP binding site 22-mer oligonucleotide. Each solution comprised 1 mM GAL4(147) in 20 mM phosphate buffer and 0.3 M KCl, p²H 7.0. (a) Dark spectrum and (b) photo-CIDNP difference spectrum of GAL4(147) itself, and (c) 1:1 GAL4(110) dimer/CRP binding site. The spectral conditions were the same as given under Materials and Methods, and 64 cycles were accumulated.

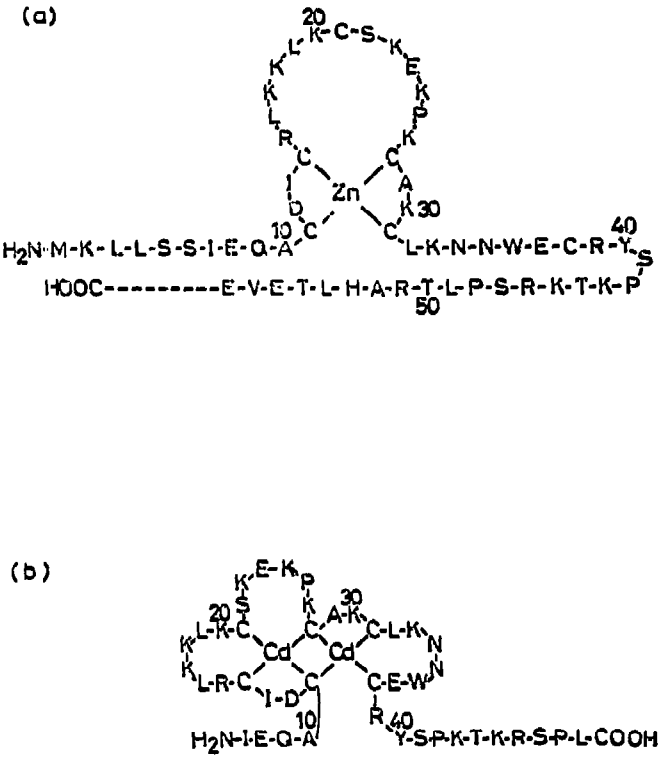

Fig. 4. Schematic drawing of the amino acid sequence of the DNA binding domain of GAL4. (a) The putative Zn-finger. (b) $Cd(II)_2Cys_6$ cluster model [9,10].

residue at position 36. Thus we could assign all the aromatic side chains except those of phenylalanine residues. These assignments are consistent with those of the side chain protons of the smaller GAL4 DNA binding domains, GAL4(62), determined by 2D and 3D NMR [13], GAL4(62*) by Pan and Coleman [14], and GAL4(4-49) by Gadhavi et al. [15]. The weak signals at around 8 ppm (peaks t and u) are probably due to unexchanged amide protons, because they disappeared after the sample had been exposed to a deuterium oxide solution at an elevated temperature (data not shown).

Photo-CIDNP difference spectra of GAL4(110) are shown in Fig. 2b. The three absorption peaks at 6.83, 6.98 and 7.38 ppm correspond to the ring protons of Trp^{36}. Besides the tryptophan resonances, two additional absorption peaks, at 7.03 and 7.85 ppm, were observed in the photo-CIDNP difference spectrum of GAL4(147) (Fig. 3b). They should be due to His^{125}, which is present in GAL4(147), but not in GAL4(110). These peak patterns of the histidine and tryptophan residues observed in the CIDNP difference spectra are quite consistent with the typical photo-CIDNP pattern of amino acid side chains reported by Kaptein [12]. The two emission peaks at around 3.1 and 3.4 ppm in both spectra (Figs. 2b and 3b) are supposed to be due to C_β protons of the tryptophan residue, and the peak at 5.2 ppm to be due to a C_α proton. The emission peak at 7.9 ppm, and the two absorption peaks at 2.6 and 4.2 ppm in Figs. 2b and 3b are due to the flavin dye. The origins of the two negative signals at 4.01 and 0.64 ppm are unknown. From the results it can be said that Trp^{36} and His^{125} are exposed on the surface of the DNA binding domain of GAL4. On the other hand, the side chains of His^{53} and Tyr^{40} are buried in the protein.

On the addition of an equimolar amount of the UAS_G 17-mer [1], which consists of two deoxyribonucleotides, d(CGGAAGACTCTCCTCCG) and d(CGGAGGAGAGTCTTCCG), to GAL4(110) (double-stranded 17-mer DNA/GAL4(110) dimer = 1/ 1), the photo-CIDNP signals due to Trp^{36} significantly decreased in intensity (Fig. 2c). We observed only a very weak residual signal for the C-2 proton of Trp^{36}. It is clear that Trp^{36} is involved in the complex formation and remarkably shielded from the flavin dye. With an increase in the salt concentration we can expect dissociation of the complex. As shown in Fig. 2d, the signals of Trp^{36} reappeared at higher salt concentrations, although their intensities were not as high as for the protein itself.

To determine whether or not this phenomenon is spe-

cific to the complex formation with UAS_G, a deoxyribonucleotide duplex of similar length was added to a solution of GAL4(147), which has an identical DNA binding property to GAL4(110) [4]. Here we used a 22-mer duplex corresponding to the CRP binding site in the *lac* promoter, of which the sequence is d(TAATGTGAGTTAGCTCACTCAT)/d(ATGAGTGAGCTAACTCACATTA) [16] and which does not exhibit homology of greater than a three base sequence with UAS_G. As shown in Fig. 3c, the photo-CIDNP signals due to Trp^{36} significantly decreased in the same way as in the case of the UAS_G complex, although the CIDNP signal due to His^{125}, which does not exist in GAL4(110), was not affected by the DNA binding as such. This means that Trp^{36} is also involved in non-specific interactions.

Corton and Johnston [11] showed that most of the Cys-rich region of GAL4 (residues 11–38, the exception being Lys^{23}) including Trp^{36} can be replaced by an analogous zinc-cluster region from another yeast activator protein, PPR1, without alteration of its DNA binding specificity. Therefore, the Cys-rich region itself might not determine the DNA specificity, although the region is essential for DNA binding, as shown by a point mutation study [17]. According to the results of Corton and Johnston [11], it is the 14 amino acid region (residues 39–52) adjacent to the Cys-rich region that determines the DNA specificity. Trp^{36} is located within the Cys-rich region (residues 11–38) and adjacent to the 14 amino acid region (Fig. 4). Our results are quite consistent with these observations. Trp^{36} is involved in both base sequence-specific and non-specific interactions with DNA.

In the analyses of smaller GAL4 DNA binding domains, GAL4(62) [13] and GAL4(7–49) [15], two regions for residues Asp^{12}–Lys^{19} and Lys^{30}–Trp^{36} were found to be α-helical. This structure must be retained in GAL4(110) and GAL4(147). Trp^{36} seems to be located on the α-helix which may contribute to DNA binding. As far as we know this is the first direct physicochemical evidence for the involvement of any specific amino acid residue of GAL4 in DNA binding. A photo-CIDNP experiment was previously performed to study the interaction of the Cro repressor from bacteriophage k with the operator DNA and non-specific DNA, and provided clear evidence for the involvement of tyrosine and histidine residues in the interaction [18].

Tyr^{40} is located in the 14 amino acid region adjacent to the Cys-rich region (residues 39–52), which was shown to determine the DNA specificity of GAL4 [11], and His^{53} is located adjacent to this region. The CIDNP experiment, however, showed that both of them are buried in the protein and are thus not involved in the interaction directly.

REFERENCES

[1] Giniger, E., Varnum, M.S. and Ptashne, M. (1985) Cell 40, 767–774.
[2] Oshima, Y. (1981) in: Molecular Biology of the Yeast *Saccharomyces*, vol 1 (J. Strathem, E. Jones and J.K. Broach eds.) pp. 159–180. Cold Spring Harbor, NY.
[3] Pan, T. and Coleman, J.E. (1989) Proc. Natl. Acad. Sci. USA 86, 3145–3149.
[4] Serikawa, S., Shirakawa, M., Matuo, H. and Kyogoku, Y. (1990) Protein Eng. 3, 267–272.
[5] Laughon, A. and Gesteand, R.F. (1984) Mol. Cell Biol. 4, 260–267.
[6] Rhodes, D. and Klug, A. in: Nucleic Acids and Molecular Biology, vol 2 (F. Eckstein and D.M.J. Lilley eds.) Springer, Berlin.
[7] Pan, T. and Coleman, J.E. (1990) Proc. Natl. Acad. Sci. USA 87, 2077–2081.
[8] Pan, T. and Coleman, J.E. (1990) Biochemistry 29, 3023–3029.
[9] Gardner, K.H., Pan, T., Narula, S., Rivera, E. and Coleman, J.E. (1991) Biochemistry 30, 11292–11302.
[10] Gadhavi, P.L., Davis, A.L., Povey, J.F. and Laue, E.D. (1991) FEBS Lett. 281, 223–226.
[11] Corton, J.C. and Johnston, S.A. (1989) Nature 340, 724–727.
[12] Kaptein, R. (1982) in: Biological Magnetic Resonance. vol. 4 (L.J. Berliner and J. Reuben eds.) pp. 145–191, Plenum Press, New York.
[13] Shirakawa, M., Fairbrother, W.J., Serikawa, Y., Ohkubo, T., Kyogoku, Y. and Wright, P.E. (1991) Biochemistry (submitted).
[14] Pan, T. and Coleman, J.E. (1991) Biochemistry 30, 4212–4222.
[15] Gadhavi, P.L., Raine, A.R.C., Alefounder, P.R. and Laue, E.D. (1990) FEBS Lett. 276, 49–53.
[16] Lee, S.J., Shirakawa, M., Akutsu, H., Kyogoku, Y., Shiraishi, M., Kitano, K., Shima, M., Ohtsuka, E. and Ikehara, M. (1987) EMBO J. 6, 1129–1135.
[17] Johnston, M. and Dover, J. (1987) Proc. Natl. Acad. Sci. USA 84, 2401–2405.
[18] Shirakawa, M., Lee, S.J., Akutsu, H., Kyogoku, Y., Kitano, K., Shin, M., Ohtsuka, E. and Ikehara, M. (1985) FEBS Lett. 181, 286–290.