

NMR analysis of CYP1(HAP1) DNA binding domain–CYC1 upstream activation sequence interactions: recognition of a CGG trinucleotide and of an additional thymine 5 bp downstream by the zinc cluster and the N-terminal extremity of the protein

Anne-Lise Vuidepot, François Bontems*, Michel Gervais¹, Bernard Guiard¹, Evelyne Shechter¹ and Jean-Yves Lallemand

Groupe de RMN, DCSO, Ecole Polytechnique, F91128 Palaiseau, France and ¹Centre de Génétique Moléculaire, Laboratoire propre du Centre National de la Recherche Scientifique associé à l'Université Pierre et Marie Curie, F91190 Gif-sur-Yvette, France

Received April 23, 1997; Revised and Accepted June 16, 1997

ABSTRACT

The DNA binding domain of the yeast transcriptional activator CYP1(HAP1) contains a zinc-cluster structure. The structures of the DNA binding domain–DNA complexes of two other zinc-cluster proteins (GAL4 and PPR1) have been studied by X-ray crystallography. Their binding domains present, besides the zinc cluster, a short linker peptide and a dimerization element. They recognize, as homodimers, two rotationally symmetric CGG trinucleotides, the linker peptide and the dimerization element playing a crucial role in binding specificity. Surprisingly, CYP1 recognizes degenerate forms of a direct repeat, CGGnnnTAnCGGnnnTA, and the role of its linker is under discussion. To better understand the binding specificity of CYP1, we have studied, by NMR, the interaction between the CYP1(55–126) peptide and two DNA fragments derived from the CYC1 upstream activation sequence 1B. Our data indicate that CYP1(55–126) interacts with a CGG and with a thymine 5 bp downstream. The CGG trinucleotide is recognized by the zinc cluster in the major groove, as for GAL4 and PPR1, and the thymine is bound in the minor groove by the N-terminal region, which possesses a basic stretch of arginyl and lysyl residues. This suggests that the CYP1(55–126) N-terminal region could play a role in the affinity and/or specificity of the interaction with its DNA targets, in contrast to GAL4 and PPR1.

INTRODUCTION

The CYP1 protein is an oxygen-dependent transcriptional activator of the yeast *Saccharomyces cerevisiae* (1–3). Its DNA binding domain is located in the N-terminal region of the molecule (2,3) and belongs to the zinc-cluster family, which is characterized by the presence of two zinc ions complexed by the sulfur atoms of six cysteines (4,5). On the basis of amino acid

alignment, several proteins of this family, including CYP1, could be grouped in a subclass in which the zinc cluster is connected to a dimerization element by a short linker peptide (6).

The interactions with DNA of several members of this subclass have been studied. They have been shown to recognize DNA sequences containing two rotationally symmetrical or two directly repeated CGG trinucleotides separated by a variable number of base pairs (6). Analysis of the crystal structures of the DNA–DNA binding domain complexes of GAL4 (7) and PPR1 (8) has shown that binding of the CGG trinucleotides is ensured by a highly conserved helix of each zinc cluster. In this context, the linker element has been proposed to play an essential role. Indeed, the specificity of the interaction seems to be determined by the fit of the distance between the two zinc-cluster domains in the dimer, imposed by the structure of the linkers, to the distance between the two CGGs in the DNA target (9). The GAL4 and PPR1 linkers have the same length, yet they recognize two rotationally symmetrical CGGs spaced by 11 and 6 bp respectively. GAL4 linkers are completely spread on the DNA and contribute to the stability of the complex by non-specific interactions with phosphate groups located between the two CGGs. The PPR1 linkers adopt a structure that brings the two zinc-cluster regions into close proximity. This induces an asymmetrical disposition of the two monomers stabilized by the presence of contacts between the zinc cluster of one monomer and the linker peptide of the other.

CYP1 seems to behave rather differently. In contrast to known GAL4 and PPR1 targets, which nearly always present a perfect inverted repeat of the CGG motif (10,11), CYP1 targets do not really present a consensus sequence (12 and references therein). Some of them may be read either as inverted repeats: (CCG_{n7}CGG) or as direct repeats: (CGG_{n6}CGG) or even do not present any CGG repeat. In fact, selection experiments (13) and analysis of various CYP1 binding sites (12) suggested that the natural targets of CYP1 are degenerate forms of the optimal sequence, CGGnnnTAnCGGnnnTA. All these observations lead to three important structural questions: (i) does the CYP1 zinc cluster recognize CGG trinucleotides as do the zinc clusters of

*To whom correspondence should be addressed. Tel: +33 1 69 33 48 32; Fax: +33 1 69 33 30 10; Email: francois.bontems@polytechnique.fr



Figure 1. (a) Comparison of the CYP1(55–126) fragment with those used in the NMR and crystallographic study of PPR1 (7) and GAL4 (9,28). (b) Sequences and numbering of the 16 and 11 bp fragments of *CYC1*-UAS1B.

PPR1 and GAL4?; (ii) does the TA doublet play a direct role in the interaction?; (iii) in the case of a direct involvement of the TA doublets in binding, what part of the CYP1 DNA binding domain interacts with the additional TA doublets? To address these questions, we have undertaken a study of the interactions between the CYP1(55–126) peptide, whose structure was previously determined (14), and a DNA fragment derived from *CYC1* upstream activation sequence 1B (*CYC1*-UAS1B). Our data indicate that the CGG trinucleotide is recognized by the zinc-cluster domain in the major groove, as in the case of GAL4 and PPR1, while the N-terminal part of the peptide, which is unstructured in GAL4 (7) and PPR1 (8) complexes, interacts in the minor groove with a thymine located 5 nt downstream of the CGG at the position corresponding to an adenine in the optimal sequence proposed by Zhang and Guarente (13).

MATERIALS AND METHODS

Sample preparation

The 55–126 fragment of the CYP1 DNA binding domain, whose structure of the (60–100) region has been previously determined by NMR spectroscopy (14), was expressed and purified as previously described (5). Cells were grown on minimal medium supplemented with [¹⁵N]ammonium chloride, leading to uniformly ¹⁵N-labeled samples.

All single-strand oligonucleotides were synthesized with a model 7500 DNA synthesizer (Milligen). The two duplexes (16 and 11 bp respectively) were prepared by mixing an equimolar amount of each strand (determined by absorbance at 260 nm), which were subsequently heated to 95°C and annealed by slow cooling at room temperature.

The resulting samples were dialyzed against 50 mmol/l NaH₂PO₄/K₂HPO₄, 100 or 200 mmol/l NaCl buffer, pH 6.0. Finally, 10% D₂O and 0.03% NaN₃ (to prevent bacterial growth) were added. D₂O samples were prepared by freeze drying H₂O samples and redissolving them in pure D₂O.

Interaction experiments

We have studied the evolution of the NMR spectra of the 16 and 11 bp DNA fragments in the presence of increasing amounts of protein (protein/16 bp and protein/11 bp experiments) and the evolution of the protein spectra upon addition of the 16 bp fragment (16 bp/protein experiment). Addition of increasing amounts of protein to a 16 bp fragment sample was done twice, using two different salt concentrations (100 and 200 mmol/l NaCl).

Before each experiment, the DNA and protein samples were dialyzed against the same buffer (50 mmol/l NaH₂PO₄/K₂HPO₄, 100 or 200 mmol/l NaCl, pH 6.30 for the CYP1–16 bp complexes and pH 6.0 for the CYP1–11 bp complex). The added species was fractionated and each fraction lyophilized. After recording the first reference spectrum, the sample was carefully removed from the NMR tube, mixed with the lyophilized fraction and returned to the NMR tube. This process was repeated for each addition. Four spectra (at ratios of 0.25:2, 0.5:2, 0.75:2 and 1:2) were registered for the 16 bp/protein experiment, leading to a final concentration of 1.5 mmol/l DNA and 3 mmol/l CYP1. The reverse protein/16 bp experiment was carried out by recording five spectra (at ratios of 0.33:1, 0.66:1, 1:1, 1.33:1 and 1.66:1) in a first experiment and four spectra (0.25:1, 0.50:1, 0.75:1 and 1:1) in the second, using a 1.5 mmol/l DNA concentration in both cases. Similarly, three spectra (0.25:1, 0.5:1 and 1:1) were realized for the protein/11 bp experiment, leading to a final concentration of 2 mmol/l for both the protein and the DNA.

NMR processing

Spectra were collected on a Bruker AMX-600 spectrometer equipped with a gradient ¹³C/¹⁵N/¹H triple resonance probe.

The DNA fragments were assigned using NOESY (15,16), HOHAHA (17) and DQF-COSY (18) experiments recorded in H₂O and D₂O at 30 (16 bp) and 25°C (11 bp). The NOESY mixing time was set to 100 or 300 ms, while the HOHAHA spin lock duration was set to 50 or 70 ms, using a MLEV17 sequence

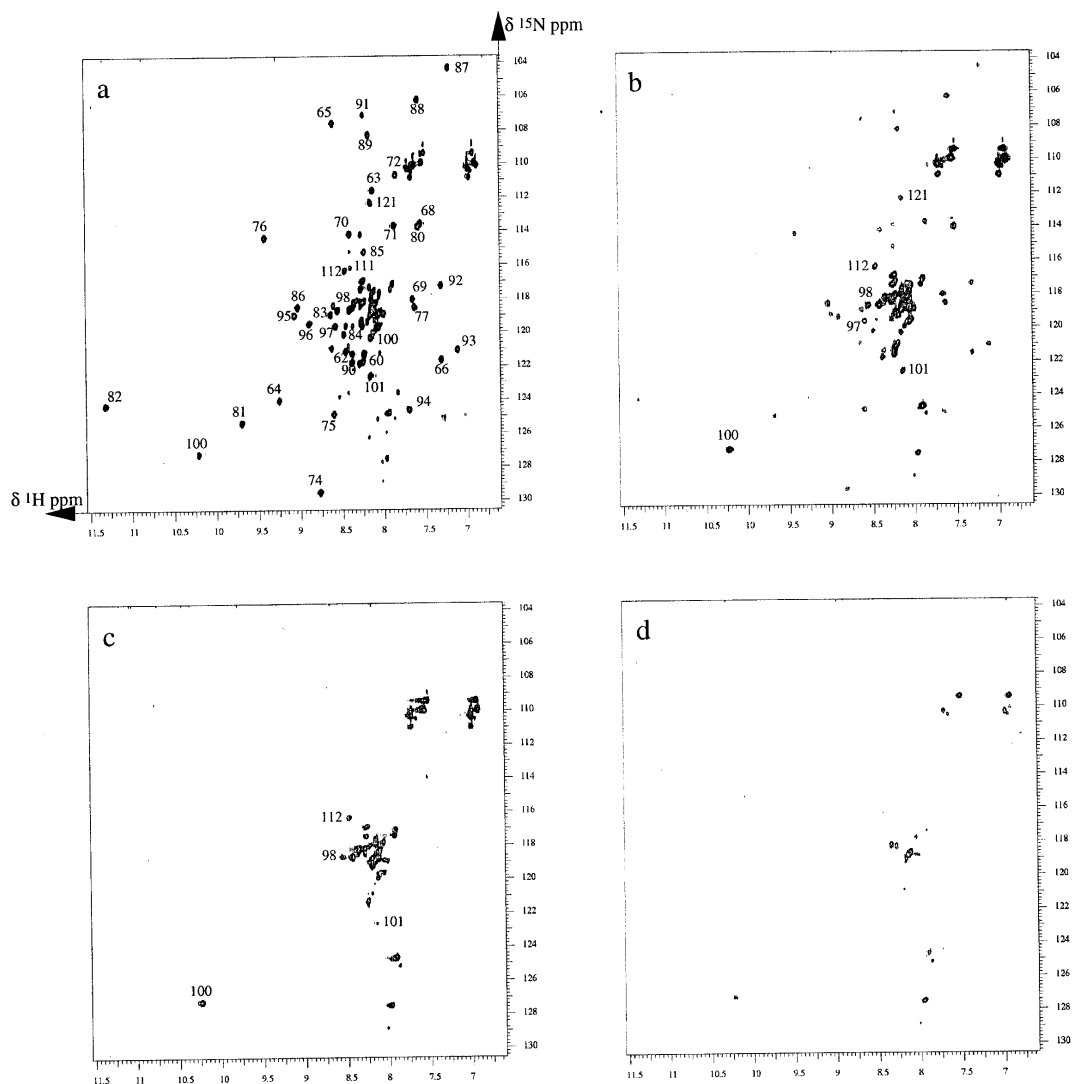


Figure 2. Comparison of the HSQC spectra of the CYP1(55–126) protein registered in the presence of increasing amounts of the 16 bp fragment. (a) Reference; (b) 0.25:2 DNA:protein ratio; (c) 0.5:2 DNA:protein ratio; (d) 1:2 DNA:protein ratio. The previously assigned correlations are labeled in the reference spectrum (a). Some correlations of the unstructured 55–62 and 97–126 extremities are still detectable in the (c) or even (d) experiments. Those which have been assigned are reported. They concern the C-terminal region (97–126) of the protein.

(19). The spectral width covered 4807 Hz in D_2O (carrier frequency offset 10780 Hz) and 9259 Hz in H_2O (carrier frequency offset 12200 Hz). All experiments consisted of 512 FIDs of 64 or 128 scans, each corresponding to 2048 time domain points. To avoid saturation of the exchangeable protons, water suppression was achieved using a Watergate (20) or a jump–return (21) sequence.

Protein–DNA interactions were followed using HSQC and NOESY spectra. HSQC spectra were recorded with spectral widths of 4130 and 1812 Hz in the proton and ^{15}N dimensions respectively. Two hundred and fifty six FIDs of 96 scans were accumulated. NOESY spectra were obtained as previously mentioned. The CYP1(55–126)–11 bp complex (corresponding to the last point of the protein/11 bp experiment) was further analyzed by recording a HMQC jump–return spectrum (22) and a series of NOESY spectra with different mixing times. In addition, NOE correlations between the protein and the DNA were looked for by mean of a series of 1D NOE difference experiments (23).

All experiments were processed off-line on Silicon Graphics or Sun workstations using GIFA software (24). Before Fourier transformation, all data were apodized with an appropriate shifted sine bell in two dimensions and zero filled to a $1K \times 2K$ real matrix ($512 \times 1K$ for HMQC and HSQC experiments). Analyses were carried out using GIFC software (25).

Model building of the 11 bp–CYP1(60–95) complex

Considering the similarities of the interactions observed in our complex with those described in the case of GAL4, a model was built using the GAL4–DNA complex (PDB entry pdb1d66.ent) as template.

The 11 bp DNA fragment was generated with the Biosym Builder module. The coordinates of the CYP1(60–95) fragment were taken from the first record of the pdb1pyc.ent database entry.

The axes of the 11 bp and of the GAL4 complex DNA fragments were superimposed in a manner that maximized the

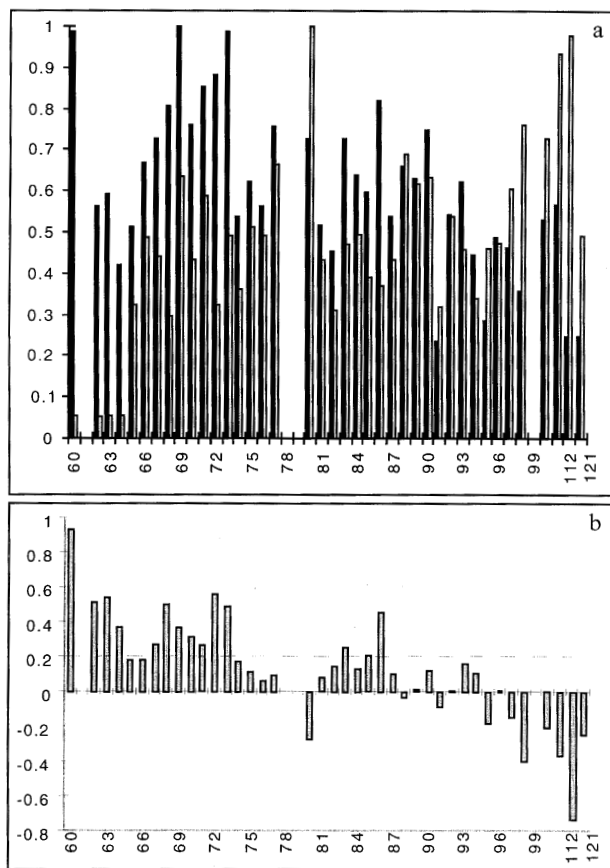


Figure 3. (a) Comparison of the relative intensities of the ^{15}H - ^1H correlations of the protein in the reference spectra (in black) and in the 0.25:2 DNA:protein ratio spectra (in grey). The value of the most intense assigned correlation has been taken as the reference in each spectrum. (b) Differences between the relative intensities.

match between the CGG triplets of the two molecules. The backbone of the CYP1(60–95) fragment was superimposed on that of the GAL4 zinc cluster using the regions conserved in both molecules. Then the CYP1(60–95) region was slightly re-oriented to optimize the match between the recognition helices.

The obtained structure was further minimized in the presence of the NOE constraints. The heavy atoms of the DNA and of the CYP1(64–95) backbone were constrained by a harmonic potential to avoid large deformations ($50 \text{ kcal/mol}/\text{\AA}^2$ on the DNA, $10 \text{ kcal/mol}/\text{\AA}^2$ on the protein backbone). A series of 10 minimizations was performed using the same force constant on the NOESY constraints ($10 \text{ kcal/mol}/\text{\AA}^2$) but with an increasing weight (from 0.1 to 1) on the van der Waals potential. All minimizations were performed without electrostatic potential.

The complex was built and analyzed using the Biosym Insight interface. All minimizations were realized using X-PLOR software (26).

RESULTS

The CYP1(55–126) fragment we have studied (Fig. 1) is similar to those used in the crystallographic and NMR studies of GAL4 (7,27,28) and PPR1 (8). It is formed of the zinc cluster subdomain (residues 64–95) linked to the 16 first amino acids of the putative

dimerization helix (residues 111–126) by the linker peptide (residues 96–110). In addition, nine residues (55–63) are present at the N-terminus. The linkers of the GAL4 and PPR1 fragments are shorter (nine residues instead of 14), but the lengths of the dimerization helices are similar (16 residues for CYP1 and GAL4, 17 for PPR1). The main difference is the presence in the case of the PPR1 fragment of a 23 amino acid tail at the C-terminus. The CYP1(55–126) fragment was previously studied alone in solution, leading to determination of the structure of the zinc cluster region, the other parts of the molecule remaining unstructured (14).

The two DNA targets are part of *CYC1* upstream activation sequence 1B (*CYC1*-UAS1B) (Fig. 1). The first, corresponding to the 16 bp (GCCGGGGTTTACGGAC) sequence, was chosen to promote fixation of two molecules of protein, eventually as a dimer. The second fragment of 11 bp (CCGGGGTTTAC) was used to look in more detail at the interactions between the protein and one of the CGG trinucleotides.

^1H resonance assignment of the 16 and 11 bp DNA fragments

The main difficulty of nucleic acid assignment is the poor dispersion of the spectra, amplified, in our case, by the lack of symmetry of the fragments. Despite these difficulties, the assignment of all proton resonances, with the exception of a few $\text{H}5'/\text{H}5''$ protons, was obtained using the well-described standard procedure (29).

The presence of the low field imino proton resonances confirmed the double helical structure of the 16 bp sequence and analysis of the intra-residual and sequential ($\text{H}6-\text{H}8$)/ $\text{H}2'/\text{H}2''$ correlations argued in favor of an overall B-type conformation. The non-exchangeable protons were assigned using the T8 T9 T10 triplet on one strand and the unique T18 C19 sequence on the other as a starting point. All but the G1 and G17 imino protons, which are located at the extremities of the DNA fragment and are probably in very fast exchange with the solvent, were identified. Assignment of the cytosine amino protons was also quite straightforward. In contrast, none of the adenine and guanine NH_2 protons could be observed in the recorded spectra.

The 11 bp sequence corresponds to the C2G31–C12G21 region of the 16 bp sequence (the numbering of which will be used for both nucleotides). Its assignment was thus mainly derived from that of the 16 bp sequence by superimposing the NOESY spectra of the two molecules. Most of the correlations present in both spectra were found at similar positions. The main variations of chemical shifts concerned the protons of the base pairs located at the termini of the 11 bp sequence, namely C2 and G31 on one strand and C12, G21 and T22 on the other.

Addition of the 16 bp DNA fragment to CYP1(55–126)

Figure 2 shows the comparison of four HSQC spectra of the protein recorded at increasing DNA:protein ratios. Clearly, the correlations broaden markedly and decrease in intensity but undergo only very small chemical shift variations, suggesting that the complex has an intermediate exchange rate. However, the variations of the intensities are very inhomogeneous. At a 0.25:2 16 bp:protein ratio most of the correlations are still present, with a more important decrease for the zinc-cluster protons. This is amplified at a ratio of 0.5:2, where the cluster disappears. Finally, at a 1:2 16 bp:protein ratio nearly all correlations are absent.

Comparison of the relative intensities between the spectra recorded at ratios of 0:2 and 0.25:2 (Fig. 3) allows delineation of three regions. The N-terminal fragment (60–64), the first helix

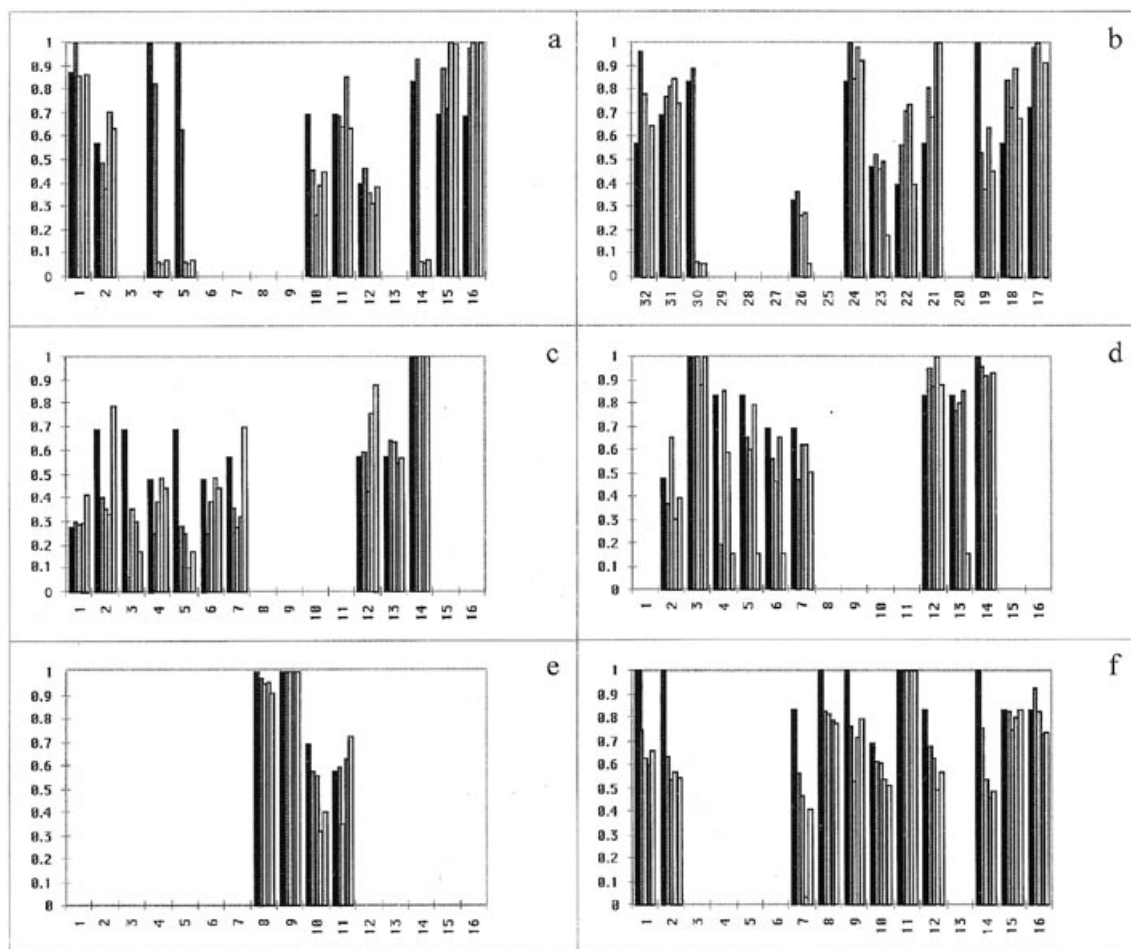


Figure 4. Comparison of the relative intensities of the 16 bp NOESY spectra recorded in the presence of increasing protein:DNA ratios (reference, 0.25:1, 0.5:1, 0.75:1 and 1:1, from black to light grey). For each series and each spectrum the value of the most intense correlation has been taken as the reference. (a) H6H8/H1' correlations of the 1–16 strand; (b) H6H8/H1' correlations of the 17–32 strand; (c) CG base pair imino/low field amino correlations; (d) imino/high field amino correlations; (e) AT base pair imino-H2 correlations; (f) cytosine H5-H6 and thymine CH₃-H6 correlations.

(65–71) and the beginning of the following strand (72–73) are characterized by a noticeable diminution in the relative intensities, leading to the disappearance of the Ile60 and Leu62–Cys64 residues. On the other hand, the cluster C-terminal region Tyr95–Gln98 and the Trp100, Ala101, Asn111, Asp112 and Val121 residues, which belong to the linker and the dimerization element, show an increase in relative intensity. Finally, the central part of the cluster remains stable, with the exception of the His80 intensity increase.

The decrease in the relative intensities in the 67–73 region, in particular of Arg68, Arg70, Lys71, Val72 and Lys73, demonstrates the importance of this region for the interaction. These amino acids are conserved or type conserved in the GAL4 and PPR1 zinc clusters and have been shown to be involved in contact with DNA (9,7). This suggests that the CYP1 zinc cluster binds to DNA in the same manner as GAL4 and PPR1. The decrease in Lys86 intensity also suggests that this residue participates in contacts either with DNA or with the second protein. Indeed, a Lys86→Ile mutation induces a loss of affinity of CYP1 for the *CYC1* and *CYC7* UAS (30). More surprising are the variations observed at both ends of the cluster. The increases in the Trp100, Ala101,

Asn111, Asp112 and Val121 relative intensities, together with the observation that the chemical shifts of most of the non-assigned correlations stay unmodified throughout the experiment, suggest that the linker and the dimerization helix remain unstructured. On the other hand, the disappearance of the Ile60, Leu62, Ser63 and Cys64 correlations suggests that these amino acids acquire structure in the presence of DNA. This latter result, confirmed by analysis of the protein–11 bp complex (see later), indicates that the N-terminal region of CYP1(55–126) behaves differently from that of GAL4, which remains unstructured in the presence of DNA (7). Finally, the particular behavior of His80 does not seem to be related to any direct interaction. In fact, its intensity increase almost certainly reflects mobility of the residue, located in a loop at the junction of the two half-domains of the zinc cluster.

Addition of CYP1(55–126) to the 16 bp DNA fragment

Addition of the protein to the 16 bp fragment sample also resulted in a marked broadening of the resonances, leading to the disappearance of nearly all DNA signals at a ratio of 1.66:1.

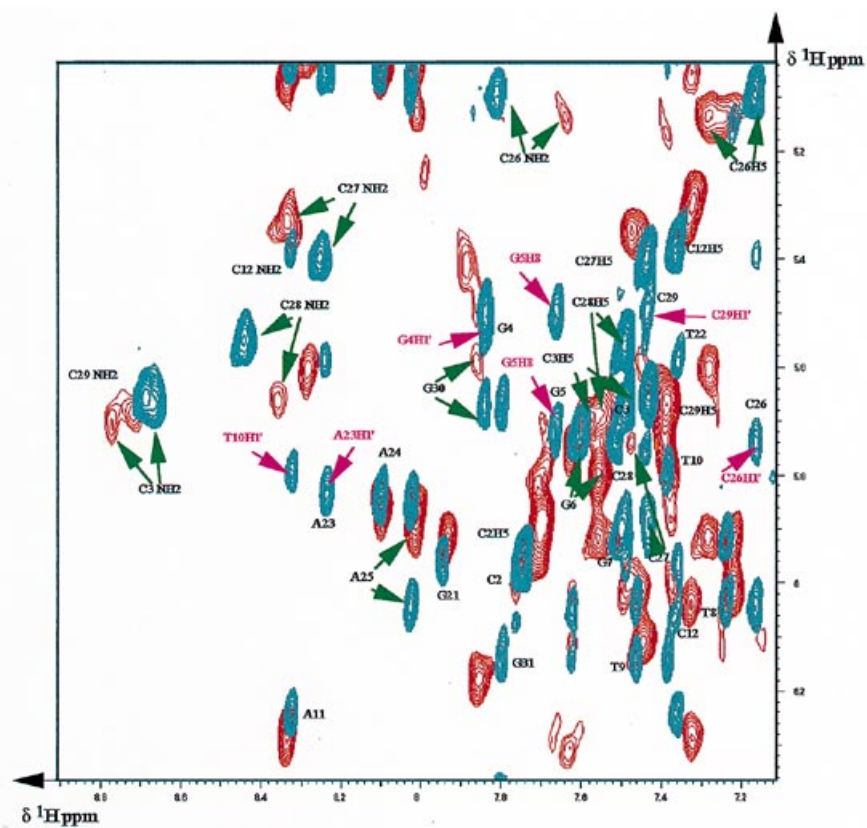


Figure 5. Superimposition of the NOESY H6H8/H1'H5 regions of the free (blue) and complexed (red) 11 bp DNA fragment. The assignments are reported in black. The correlations present only in the free form and the correlations undergoing large chemical shift variations are indicated by pink and green arrows respectively.

However, a closer look at the imino region shows that the rate of disappearance depends on the proton considered. Using the 2D NOESY series, it appears that the first resonances to disappear belong to G4C29 (imino and amino resonances) and T10A23 (T10 imino and H1', A23 H2) base pairs and C26 (H1'H2'H2''). On the other hand, the resonances of the base pairs located at the two termini (G1C32, G31 and C16G17, T18) together with those of the central region (T8A25, T9) remain visible until the end of the experiment.

Unfortunately, the decrease in the intensities was too fast to allow a detailed analysis. So the experiment was repeated, focusing on the first half of the curve (protein:16 bp ratios of 0.25:1, 0.50:1, 0.75:1 and 1:1), with a higher salt concentration in the hope of accelerating the exchange rate. Indeed, many correlations of the DNA (followed on NOESY spectra) and of the protein (detected on an HSQC spectrum) remain visible at the ratio of 1:1.

The mean intensity, calculated on all analyzed correlations, is 60% for the 0.25:1 and 35% for the 0.5:1 and 1:1 protein:16 bp ratios. As shown in Figure 4a and b (H6,H8-H1' correlations), the main variations concern the G30, G4 and G5 bases (which correspond to the first CGG triplet), G14 (which belongs to the second CGG) and the A23 and C26 bases. The C12G21 base pair (corresponding to the cytosine of the second CGG) seems unaffected, while the C3 protons (which belong to the first CGG) disappear rapidly, suggesting a binding difference between the two CGGs. A similar observation can be made in Figure 4c and d (CG imino-amino correlations), which shows that the correlations concerning the first CGG triplet disappear faster than those

concerning the second. The effect of binding on the T10A23 base pair is also visible through the imino-H2 (Fig. 4e) and to a lesser extent through the CH₃-H6 (Fig. 4f) correlations.

Ha and co-workers (12) have recently proposed an optimal sequence for CYP1 binding sequences formed by the repetition of two CGG and two TA motifs, CGGnnnTAnCGGnnnTA. The DNA fragment we have used in the present study was prepared from the wild-type *CYC1-UAS1B* and presents a CGGnnnTTn-CGG motif. The TA doublet of the optimal sequence is replaced by a TT in the first half-site and is absent in the second. As expected, CYP1 binding perturbs the proton signals of the two CGGs, but also those of the T10A23 base pair, which corresponds to the second base pair of the TT doublet. Interestingly, the main effect is seen on the T10 imino and A23 H2 protons, which are located in the minor groove of the DNA, while the CH₃ and H6 protons (in the major groove) are only weakly affected. The importance of this additional interaction is also indirectly assessed by the non-equivalence of the two half-sites. As shown by both protein/16 bp experiments, the H6,H8-H1' and imino-amino correlations of the first CGG triplet disappear more rapidly than those of the second, suggesting that CYP1(55-126) presents a higher affinity for the first half-site, which possesses the TT doublet, than for the second, which does not. Strikingly, the T9A24 base pair, whose importance was also suggested by Zhang and Guarente (13), seems unaffected by binding of CYP1 in our experiments. It is impossible to rule out that this could result from the use of a truncated CYP1 fragment. However, another explanation may be considered. It has been shown that the

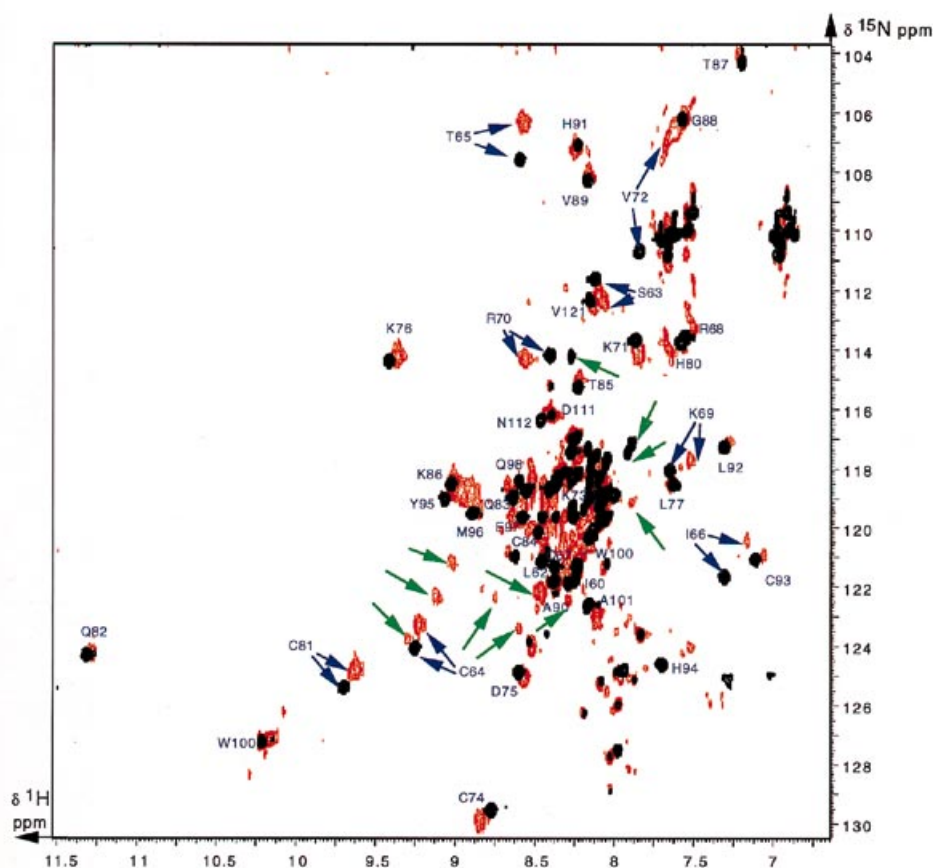


Figure 6. Superimposition of two CYP1(55–126) ^{15}N - ^1H HMQC spectra recorded in the absence (black) and presence (red) of the 11 bp fragment. As evidenced by the arrows, many correlations undergo large chemical shift variations. Some of them, in green, correspond to non-assigned residues. The others, in blue, concern residues of the zinc cluster thought to be involved in interaction with the DNA and also the N-terminal Ser63–Ile66 segment.

binding of GAL4 is sensitive to the nature of the base pairs in the middle of the site, even in the absence of any specific contacts (31). Similarly, we can imagine a structural role for the first base pair that would favor, for example, correct orientation of the second.

Addition of CYP1(55–126) to the 11 bp half-site DNA fragment

As previously, all resonances broaden as the higher molecular weight complex becomes the predominant species. But, in addition, we also observe the displacement of some correlations (Fig. 5), suggesting that several protons now show a fast exchange rate. This phenomenon concerns both the DNA and the protein.

DNA evolution was followed using the imino-amino, H5/H6 and H6-H8/H1' correlation regions, where no protein signal was present. Many correlations disappear, in particular those of G4, G5 and T10 on one strand and A23, C28 and C29 on the other. We also observe a large chemical shift variation of the C3, C26, C27 and G30 resonances. As expected, these modifications concern the C3G30, G4C29, G5C28 motif and its surrounding bases (C26 and C27), but also the T10A23 base pair. The T8/A25 and T9/A24 imino-amino correlations remain visible throughout the experiment.

Similarly, superimposition of the HMQC spectra of the free and complexed protein (Fig. 6) shows that some residues undergo a large chemical shift variation, in particular the 63–66 and the 69–72 regions, together with Cys81. Others do not seem to be influenced by the interaction, i.e. the 75–80, 82–93, 95–101

fragments and Asp111, Asn112 and Val121. These results are confirmed by the evolution of the correlation intensities observed in the NOESY spectrum of the complex (data not shown). In addition, some cross-peaks 'appear' in the spectrum. These new cross-peaks result from large displacements of correlations previously located in the crowded central region of the spectrum and corresponding to non-assigned protons of the unstructured regions in the free form of CYP1.

These observations agree with our previous experiments. Even in the presence of a half-site (11 bp), the CYP1(55–126) fragment interacts with the CGGnnnnT sequence. Our data also show clearly that a part of the protein acquires structure upon DNA binding. Considering that the DNA target contains only one half-site, that Trp100, Ala101, Asp111, Asn112 and Val121 remain unaffected and that the Ser63–Ile66 segment undergoes a large chemical shift variation, it seems clear that the N-terminal part of the CYP1(55–126) fragment is concerned in the interaction. This confirms our previous hypothesis and strongly suggests that the T10/A23 base pair is recognized by the N-terminal region of the CYP1(55–126) protein.

Intermolecular NOE constraints and a model of the complex

Using 2D NOESY and 1D NOE difference experiments, we were able to observe 20 intermolecular contacts between the protein and the 11 bp fragment (Table 1). They concern three protein

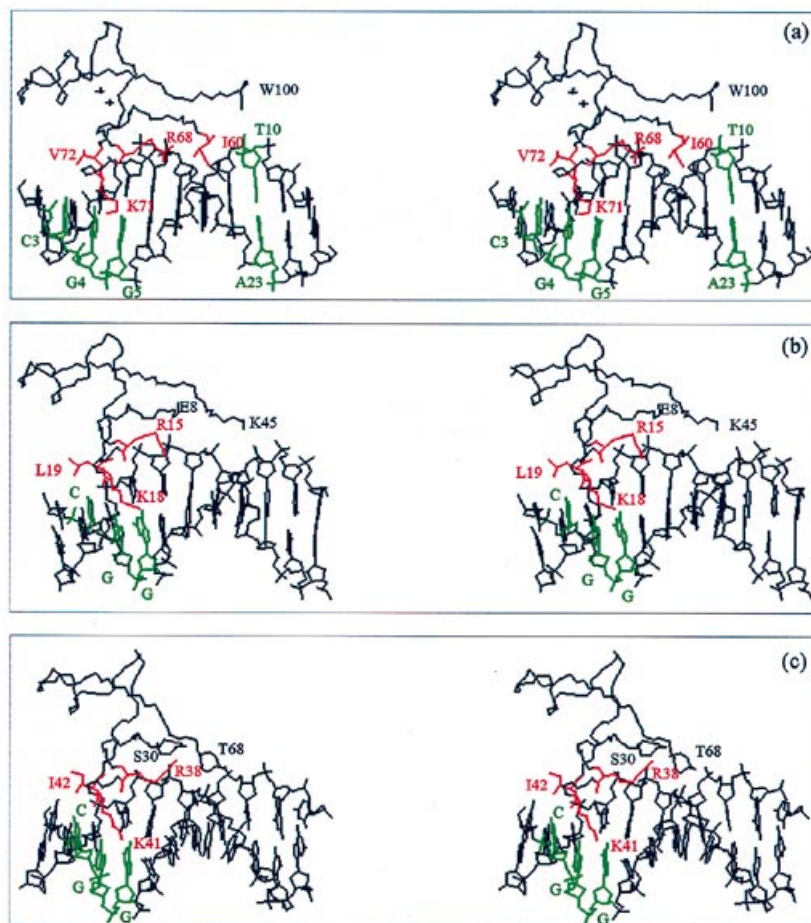


Figure 7. Comparison of the stereoview model structure of the CYP1(60–96) fragment in interaction with the 11 bp DNA fragment (a) with the crystallographic structures of GAL4 zinc cluster–DNA (b) and PPR1 zinc cluster–DNA (c) complexes. For clarity, only one half-site has been represented in the latter two cases. The CGG triplets together with the T10A23 base pair thought to interact with the CYP1(55–126) N-terminal region are in green. The CYP1 Arg68, Lys71 and Val72 residues, whose positions are defined by the observation of protein–DNA NOEs, and the corresponding GAL4 and PPR1 residues are displayed in red. Similarly, the position of the CYP1 Ile60 side chain is also displayed in red.

residues (Arg68, Lys71 and Val72) and five DNA bases (C2, C3, C27, C28 and C29) and demonstrate an interaction between the region of the cluster first helix (Lys69–Val72) and the CGG (C3–G5 and C28–G30) motif.

Many of these contacts are similar to those observed for GAL4 (Lys71H^a/C28H5, Lys71H^b/C3H5, Lys71H^c/C3H5 and Val72CH₃/C2H5) (28). We thus decided to build a preliminary model of the complex using the relative protein/DNA disposition observed in the case of GAL4 and our intermolecular NOEs (Fig. 7). After refinement this model displays no bad contacts and a unique distance violation $>0.5 \text{ \AA}$ (between the Lys71 H^γ and the C29 H5), which may be due to the fact that we kept the DNA structure rigid during the minimization.

The structure, in agreement with all the data we have previously obtained, supports the idea that the CYP1 zinc cluster domain recognizes the CGG trinucleotides, as do those of GAL4 and PPR1. However, the N-terminal region of GAL4 appears unstructured and rather far from the DNA. In contrast, we observe that the CYP1 Ile60–Cys64 fragment has a conformation that brings the Ile60 residue into the minor groove of the DNA near C26 (which may affect its intensity variations). This may look rather strange, considering the absence of any constraint between

the N-terminal region of the protein and the DNA. In fact, it appears that this particular Ile60–Cys64 peptide structure is present in nearly all free CYP1 structures. This results from the presence of Pro61, which restrains the available conformational space, and of several NOE restraints observed in the free structure, between Leu62, Ser63 and Cys64 and their surroundings (in particular Cys67, Arg68, Cys74, Tyr95 and Met96).

DISCUSSION

The manner in which GAL4 and PPR1 recognize DNA seems rather well understood today. Nearly all their known UASs contain two rotationally symmetrical CGG trinucleotides (10,11) and all the specific interactions occur between these two CGGs and the two zinc clusters of a protein dimer (7,8,31). In the case of CYP1 the picture is more complicated. The various UASs have very heterogeneous sequences. They correspond to a direct repetition of a CGG trinucleotide, but with many variations (for example *CYC1*-UAS1A and -B, CGGn6CGG; *CYB2*-UAS1, AAGGn6CGG; *CYC7*, CGCn6CGC) (12 and references therein). It has been shown recently that the two CGG trinucleotides of the latter two targets are not functionally equivalent (32). In addition, several experiments indicate the existence of contacts outside the

CGGs. Methyl interaction experiments (33) have shown that CYP1 interacts not only with the CGG (*CYC1*) or CGC (*CYC7*) trinucleotides in the major groove, but also with a stretch of As in the minor groove covering 6 or 7 bp. More recently, selection (13) and mutation (12) experiments have led to the conclusion that CYP1 recognizes degenerate forms of the CGGnnnTAn-CGGnnnTA optimal sequence.

Table 1. Summary of the intermolecular NOEs observed between the CYP1(55–126) protein and the 11 bp *CYC1*-UAS1B DNA

Protein	DNA
NOEs from the NOESY spectra	
Lys71 H ^α	Cyt3 H5
Lys71 H ^α	Cyt28 H5
Lys71 H ^β	Cyt28 NH ₂ a
Lys71 H ^{β'}	Cyt28 NH ₂ b
Lys71 H ^{β'}	Cyt3 H5
Lys71 H ^{β'}	Cyt29 NH ₂ a
Lys71 H ^{β'}	Cyt29 NH ₂ b
Lys71 H ^γ	Cyt3 H5
Lys71 H ^γ	Cyt29 H5
Lys71 H ^γ	Cyt28 NH ₂ b
Lys71 H ^γ	Cyt3 NH ₂ b
Arg68 H ^β	Cyt27 H5
Arg68 H ^γ	Cyt27 H5
NOEs from the 1D NOE difference experiments	
Val72 CH ₃	Cyt3 H5
Val72 CH ₃	Cyt29 H5
Val72 CH ₃	Cyt3 NH ₂ a
Val72 CH ₃	Cyt2 H5
Val72 CH ₃	Cyt2 H6
Val72 CH ₃	Cyt29 NH ₂ a
Val72 CH ₃	Cyt29 NH ₂ b

Our data confirm binding of the CGG trinucleotides and of, at least, a TA base pair five residues downstream in the case of the *CYC1*-UAS1B target. They show that the CGG trinucleotide is recognized by the zinc cluster in a manner similar to that found for GAL4 and PPR1, but also that the additional TA base pair is bound in the DNA minor groove by the CYP1(55–126) N-terminal region. The sequence of this region (Arg55–Lys–Arg–Asn–Arg–Ile–Pro–Leu–Ser63) contains a stretch of four basic residues. This stretch is even longer in the whole protein (Ser50–Ser–Lys–Ile–Lys–Arg–Lys–Arg–Asn–Arg–Ile–Pro–Leu–Ser63). Similar basic regions have been described at the N-termini of the λ repressor (34) and, more recently, of the GAGA protein (35). They have been demonstrated to play a critical role in DNA binding. In addition, saturation mutation experiments conducted on the CYP1(55–125) fragment (30) have led to the characterization of several mutants that modulate the activity and/or affinity of CYP1 for the *CYC1*- and *CYC7*-UAS. They concern the linker region but also the N-terminal (Lys54–Ile66) fragment.

Thus, considering our NMR results together with all the data in the literature, we propose a model of CYP1–*CYC1*-UAS1B interaction in which the CGG trinucleotide is recognized by the zinc cluster domain of the protein and a supplementary region is recognized by the basic residue-rich Lys52–Ile–Lys–Arg–Lys–Arg–Asn–Arg59 fragment, the zinc cluster and the basic residue-rich region being linked by the Ile60–Ser63 tetrapeptide. Interestingly, an analysis of the 47 zinc cluster protein N-terminus

sequences contained in the Swissprot databank reveals that 36 of them (but neither GAL4 nor PPR1) contain at least one basic residue-rich region, suggesting that the puzzling behavior of CYP1 may not be an exception.

See supplementary material available in NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Dardel and Dr Timmerman for helpful discussions and critical comments on the manuscript and Mrs D.Menay for the synthesis of the oligonucleotides. This work was supported by the French Association Pour la Recherche contre le Cancer (ARC).

REFERENCES

- Verdière, J., Creusot, F. and Guerineau, M. (1985) *Mol. Gen. Genet.*, **199**, 524–526.
- Creusot, F., Verdière, J., Gaisne, M. and Slonimski, P.P. (1988) *J. Mol. Biol.*, **204**, 263–276.
- Pfeifer, K., Kim, K.S., Kogan, S. and Guarente, L. (1989) *Cell*, **56**, 291–301.
- Johnston, M. (1987) *Microbiol. Rev.*, **51**, 458–476.
- Timmerman, J., Guiard, B., Shechter, E., Delsuc, M.-A., Lallemand, J.Y. and Gervais, M. (1994) *Eur. J. Biochem.*, **225**, 593–599.
- Schjerling, P. and Holmberg, S. (1996) *Nucleic Acids Res.*, **24**, 4599–4607.
- Marmorstein, R., Carey, M., Ptashne, M. and Harrison, S.C. (1992) *Nature*, **356**, 408–414.
- Marmorstein, R. and Harrison, S.C. (1994) *Genes Dev.*, **8**, 2504–2512.
- Reece, R.J. and Ptashne, M. (1993) *Science*, **261**, 909–911.
- Giniger, E., Varnum, S.M. and Ptashne, M. (1985) *Cell*, **40**, 767–774.
- Roy, A., Exinger, F. and Losson, R. (1990) *Mol. Cell. Biol.*, **10**, 5257–5270.
- Ha, N., Hellauer, K. and Turcotte, B. (1996) *Nucleic Acids Res.*, **24**, 1453–1459.
- Zhang, L. and Guarente, L. (1994) *Genes Dev.*, **8**, 2110–2119.
- Timmerman, J., Vuidepot, A.-L., Bontems, F., Lallemand, J.-Y., Gervais, M., Shechter, E. and Guiard, B. (1996) *J. Mol. Biol.*, **259**, 792–804.
- Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) *J. Chem. Phys.*, **71**, 4546–4553.
- Kumar, A., Ernst, R.R. and Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun.*, **95**, 1–6.
- Braunschweiler, L. and Ernst, R.R. (1983) *J. Magn. Resonance*, **53**, 521–528.
- Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K. *Biochem. Biophys. Res. Commun.*, **117**, 458–479.
- Bax, A. and Davis, D.G. (1985) *J. Magn. Resonance*, **65**, 355–360.
- Piotto, M., Saudek, V. and Sklenar, V. (1992) *J. Biomol. NMR*, **2**, 661–665.
- Plateau, P. and Guéron, M. (1982) *J. Am. Chem. Soc.*, **104**, 7310–7311.
- Szewczak, A.A., Kellog, G.W. and Moore, P.B. (1993) *FEBS Lett.*, **327**, 261–264.
- Dhingra, M.M., Sarma, M.H., Gupta, G. and Sarma, R.H. (1983) *J. Biomol. Struct. Dyn.*, **1**, 417–428.
- Delsuc, M.-A. (1989) In Skilling, J. (ed.), *Entropy and Bayesian Methods*. Kluwer, Dordrecht, The Netherlands, pp. 285–290.
- Rouh, A., Delsuc, M.-A., Bertran, G. and Lallemand, J.-Y. (1993) *J. Magn. Resonance Ser. A*, **102**, 357–359.
- Brünger, A.T. (1988) *X-PLOR Manual*. Yale University Press, New Haven, CT.
- Baleja, J.D., Marmorstein, R., Harrison, S.C. and Wagner, G. (1992) *Nature*, **356**, 450–453.
- Baleja, J.D., Mau, T. and Wagner, G. (1994) *Biochemistry*, **33**, 3071–3078.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley and Sons, New York, NY.
- Turcotte, B. and Guarente, L. (1992) *Genes Dev.*, **6**, 2001–2009.
- Liang, S.D., Marmorstein, R., Harrison, S.C. and Ptashne, M. (1996) *Mol. Cell. Biol.*, **16**, 3773–3780.
- Naït-Kaoudjt, R., Roy, W., Guiard, B. and Gervais, M. (1997) *Eur. J. Biochem.*, **244**, 301–309.
- Pfeifer, K., Prezant, T. and Guarente, L. (1987) *Cell*, **49**, 19–27.
- Clarke, N.D., Beamer, L.J., Goldberg, H.R., Berkower, C. and Pabo, C.O. (1991) *Science*, **254**, 267–270.
- Ominchinski, J.G., Pedone, P.V., Felsenfeld, G., Gronenborn, A.M. and Clore, G.M. (1997) *Nature Struct. Biol.*, **4**, 122–132.