

J-Bio NMR 098

## The solution structure of the human retinoic acid receptor- $\beta$ DNA-binding domain

R.M.A. Knegtel<sup>a</sup>, M. Katahira<sup>a</sup>, J.G. Schilthuis<sup>b\*</sup>, A.M.J.J. Bonvin<sup>a</sup>, R. Boelens<sup>a</sup>,  
D. Eib<sup>a</sup>, P.T. van der Saag<sup>b</sup> and R. Kaptein<sup>a,\*\*</sup>

<sup>a</sup>*Department of Chemistry, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands*

<sup>b</sup>*Hubrecht Laboratory, Netherlands Institute for Developmental Biology, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands*

Received 15 September 1992

Accepted 28 September 1992

*Keywords:* Retinoic acid receptor; DNA binding domain; Zinc finger; Solution structure; Nuclear magnetic resonance

---

### SUMMARY

The three-dimensional structure of the DNA-binding domain of the human retinoic acid receptor- $\beta$  (hRAR- $\beta$ ) has been determined by nuclear magnetic resonance spectroscopy in conjunction with distance geometry, restrained molecular dynamics and iterative relaxation matrix calculations. A total of 1244 distance restraints were obtained from NOE intensities, of which 448 were intra-residue and 796 inter-residue restraints. In addition 23  $\chi$  and 30  $\phi$  dihedral angle restraints were obtained from J-coupling data. The two 'zinc-finger' regions of the 80-amino acid residue protein are followed by two  $\alpha$ -helices that cross each other perpendicularly. There is a short stretch of  $\beta$ -sheet near the N-terminus. The  $\alpha$ -helical core of the protein is well determined with a backbone root-mean-square deviation (r.m.s.d.) with respect to the average of 0.18 Å and 0.37 Å when the side chains of residues 31, 32, 36, 61, 62, 65 and 69 are included. The r.m.s.d. for the backbone of residues 5–80 is 0.76 Å. For the first finger (residues 8–28), the r.m.s.d. of the backbone is 0.79 Å. For the second finger (residues 44–62) the r.m.s.d. is 0.64 Å. The overall structure is similar to that of the corresponding domain of the glucocorticoid receptor, although the C-terminal part of the protein is different. The second  $\alpha$ -helix is two residues shorter and is followed by a well-defined region of extended backbone structure.

---

### INTRODUCTION

All-trans-retinoic acid (RA), a derivative of vitamin A, plays a crucial role in vertebrate development and differentiation (Thaller and Eichele, 1987). Retinoic acid acts by binding to nuclear retinoic acid receptors (RARs). RARs are members of a superfamily of ligand-inducible

---

\* Present address: Ludwig Institute for Cancer Research, Courtauld Building, 91 Riding House Street, London W1P 8BT, U.K.

\*\* To whom correspondence should be addressed.

*Supplementary material available from the corresponding author:* A list of restraints used in the calculations.



glucocorticoid and oestrogen receptors are inverted repeats with a 3 base pair (bp) spacing between half sites. The vitamin D<sub>3</sub> receptor, thyroid hormone receptor, and RAR, however, have direct repeat response elements with a 3-, 4- or 5-bp spacing between half sites, respectively (Umesono et al., 1991). Although different spacer requirements are found, the consensus sequences for the glucocorticoid and oestrogen response elements are similar to those of the vitamin D<sub>3</sub>, thyroid hormone, and retinoic acid response elements, which suggests a similar mode of binding to DNA, involving residues located near the two C-terminal cysteines in the first zinc finger (Danielsen et al., 1989; Mader et al., 1989; Umesono and Evans, 1989; Luisi et al., 1991; Umesono et al., 1991). Recently it was shown that the vitamin D<sub>3</sub> receptor, thyroid hormone receptor and RAR preferably bind to their response elements as heterodimers with retinoid X receptors (Yu et al., 1991; Kliewer et al., 1992; Leid et al., 1992; Zhang et al., 1992), in contrast to the glucocorticoid and oestrogen receptors, which form homodimers upon DNA binding.

Structural information on the human RAR- $\beta$  DNA-binding domain might give more insight in its DNA-binding properties, which include both half-site sequence recognition and its preference for direct repeats with a 5-bp spacing. An NMR study was undertaken to elucidate the solution structure of the RAR- $\beta$  DNA-binding domain. Both nitrogen-15 and proton resonance assignments for all but the first three residues were obtained (Katahira et al., 1992). Here, we report the determination of the solution structure, by NMR, of a 80-residue protein fragment containing the human RAR- $\beta$  DNA-binding domain (Fig. 1).

## MATERIALS AND METHODS

### *Bacterial expression and production of hRAR- $\beta$ DNA-binding domain*

The DNA-binding domain of the human retinoic acid  $\beta$  receptor was obtained with a bacterial protein expression system, using an inducible gene coding for the T7 RNA polymerase in combination with transcription of the target gene under control of a strong T7 promoter (Studier and Moffat, 1986). A pET3 vector was used, which has a *NdeI* site overlapping the translation start site and a *BamHI* site downstream of it. Using oligodirected mutagenesis (Kunkel et al., 1987), a *NdeI* site in the hRAR- $\beta$  cDNA (De Thé et al., 1987) with the ATG codon in frame directly upstream of codon 75 (Pro) was introduced and 154 (Asp) was changed into a stop codon, followed by a *BamHI* site. By inserting the 270-bp *NdeI/BamHI* fragment of the human RAR $\beta$  cDNA into the pPET3 vector, all T7 coding sequences were eliminated from the vector, except for the initiator ATG codon, and a plasmid was created that directed the synthesis of a 80-amino acid protein fragment (Fig. 1) containing the complete DNA-binding region of hRAR- $\beta$ . The hRAR- $\beta$  DNA-binding domain has a calculated molecular weight of 9.4 kD. BL21(DE3)pLysS bacteria transformed with the plasmid construct under control of the lacUV promoter synthesized high levels of a protein with the expected mobility, as estimated by SDS-PAGE (results not shown), upon induction with isopropyl  $\beta$ -D-thiogalactopyranoside (final concentration 0.1 mM). Bacteria were harvested by centrifugation 3 h after induction.

### *Purification of bacterially expressed hRAR- $\beta$ DNA-binding domain*

The hRAR- $\beta$  DNA-binding domain was isolated by using an adapted version of the protocol described for the glucocorticoid receptor (Freedman et al., 1988). Bacterial pellets were resuspended in lysis buffer (50 mM TrisHCl, pH 8.0, containing 10% glycerol, 4 mM CaCl<sub>2</sub>, 40 mM

MgCl<sub>2</sub>, 0.5 mM PMSF, 1 mM EDTA, 5 mM DTT, and 250 mM NaCl) at 5% of the culture volume. Cells were lysed by freezing them at -80 °C and thawing them quickly in a waterbath at 37 °C. DNase I was added to the lysate to a final concentration of 10 µg/ml, and the lysate was incubated at room temperature for 15 min. Bacterial debris was pelleted by centrifugation and discarded. Polymyxin P was then added over 10 min at 4 °C to a final concentration of 0.2% under constant stirring, and the precipitate was removed by centrifugation. Ammonium sulphate ((NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>) was added to the supernatant to give 30% saturation, and the precipitated bacterial proteins were pelleted by centrifugation.

The (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> concentration in the supernatant was adjusted to 80% saturation and the precipitated protein was collected by centrifugation. The supernatant was discarded and the pellet was resuspended in TGEDZ<sub>50</sub> (50 mM TrisHCl, pH 8.0, containing 10% glycerol, 1 mM EDTA, 5 mM DTT, 50 µM ZnCl<sub>2</sub>, and 50 mM NaCl) and dialysed overnight against the same buffer. The dialysate was loaded onto an Accel Plus CM cation exchange column (Millipore) (column dimensions 0.75 × 18 cm). The column was eluted using a gradient of 50 mM to 1 M NaCl in TGEDZ<sub>50</sub> over 60 min. The peak fractions were pooled and used for the NMR experiments. The TGEDZ<sub>50</sub> buffer was replaced with a solution containing 200 mM NaCl, 0.1 mM DTT and 5% <sup>2</sup>H<sub>2</sub>O, and the protein was concentrated to 2–4 mM, using an Amicon YM-3 filter membrane. The presence of zinc in the hRARβ DNA-binding domain was confirmed by atomic absorption spectroscopy.

#### *DNase I protection assays*

A synthetic, double-stranded DNA fragment containing the sequence of interest (see text and legend to Fig. 2) was radioactively labelled at one end, using Klenow enzyme and α-<sup>32</sup>P-labelled dCTP (3000 Ci/mmol, Amersham), to a specific activity of about 3 × 10<sup>3</sup> dpm/ng. Purified RAR-DNA-binding domain protein was mixed with the radioactive probe in 10 µl of binding buffer (a solution containing 4% Ficoll 400, 15 mM KCl, 5 mM Hepes, 0.5 mg/ml BSA, 10 ng/ml pdIdC, 1 mM DTT, 0.1 mM EDTA, 0.5 mM MgCl<sub>2</sub>, and 5 µM ZnCl<sub>2</sub>, pH 7.9) and incubated for 1 h at room temperature. The specified amount (see Fig. 2 legend) of DNaseI (Boehringer Mannheim) was then added and the mixture was incubated for 30 min at 25 °C. The reaction was stopped by adding one volume of 5 mM EDTA, 0.1% SDS and cooling to 0 °C. The mixture was extracted with phenol, and DNA was subsequently precipitated and then analysed by electrophoresis on a 12% polyacrylamide gel under denaturing conditions. Besides samples from the protection assays, an equivalent amount of a purine-specific sequence reaction (Maxam and Gilbert, 1980) of the same fragment was run on the gel as sequence-specific size markers.

#### *NMR experiments*

NMR samples of 2–4 mM were dissolved in 95/5% <sup>1</sup>H<sub>2</sub>O/<sup>2</sup>H<sub>2</sub>O mixtures containing 200 mM NaCl and 0.1 mM DTT, at pH 6.5–6.8. <sup>1</sup>H NMR and <sup>15</sup>N NMR spectra were recorded at 288–305 K on Bruker AMX500 and AM600 spectrometers, operating at 500 and 600 MHz, respectively. Two dimensional (2D) NOE (Jeener et al., 1982; States et al., 1982), HMQC (Bax et al., 1983), DQF COSY (Rance et al., 1983) and PE COSY (Müller, 1987) spectra were recorded in a phase-sensitive mode, using TPPI (Marion and Wüthrich, 1983). The water signal was suppressed by irradiation in the relaxation delay and in the mixing time. The 2D spectra consisted of 400 FIDs (700 in the HMQC experiment and 600–700 in COSY experiments) of 2048 data points (4096 in the case of DQF-COSY and PE-COSY experiments) of 64–160 scans each. They were

recorded and processed as described previously (Katahira et al., 1992). NOE cross peaks were integrated using the 'TRITON' package and initial build-up rates were obtained by fitting to exponential or linear functions.

#### *Stereospecific assignments and torsion angle restraints*

Backbone NH-H $\alpha$  vicinal coupling constants were estimated from multiplets in DQF-COSY (Rance et al., 1983) spectra in H<sub>2</sub>O and from the HMQC (Bax et al., 1983) spectrum recorded with high resolution in t<sub>1</sub>. Residues were judged to have a  $\phi$ -torsion angle of  $-120^\circ \pm 30^\circ$  when its  $^3J_{\text{HN}\alpha}$  was larger than 9 Hz and were restrained to such values in the structure calculations. Side-chain H $\alpha$ -H $\beta$  coupling constants were estimated from PE-COSY and DQF-COSY spectra recorded in D<sub>2</sub>O. The J-coupling data were compared with the intensities of corresponding H $\alpha$ -H $\beta$  and NH-H $\beta$  NOE cross peaks in the 50-ms and 100-ms mixing time 2D NOE spectra. Since accurate determination of  $^3J_{\alpha\beta}$  coupling constants proved difficult due to cancellation effects in the anti-phase multiplets, rotamers were assigned to be *gauche* when  $^3J_{\alpha\beta}$  was clearly smaller than 5 Hz, and *trans* when  $^3J_{\alpha\beta}$  was clearly larger than 9 Hz. Stereospecific assignments of  $\beta$ -protons of AMX spin systems and  $\gamma$ -methyl groups of valine residues were obtained from  $^3J_{\alpha\beta}$  data and intra-residue NOEs involving H $\beta$  (H $\gamma$  for valines), H $\alpha$  and NH protons (Zuiderweg et al., 1985). For structure calculations, the  $\chi_1$ -angles were restrained to values of  $60^\circ$ ,  $180^\circ$  or  $-60 \pm 30^\circ$ .

#### *Distance restraints*

NOE cross peaks were peak-picked from the 2D NOE spectra recorded with 50, 100, 150 and 200 ms mixing times in H<sub>2</sub>O and D<sub>2</sub>O and integrated by simple addition after base-plane subtraction. For distance geometry (DG), initial NOE build-up rates were used to estimate distances and a 10% error was added to account for integration and experimental errors. For calibration, six different tyrosine H $\delta$ -He NOE cross peaks were used, representing a distance of 2.48 Å. Appropriate pseudo-atom corrections were added to all upper bounds involving methylene, methyl or aromatic groups, according to Wüthrich (1986), except for methyl groups, for which a pseudo-atom correction of 0.3 Å was used (Koning et al., 1990). All lower bounds were set to 2.0 Å. For four phenylalanine and four tyrosine residues, part of the inter-residue NOEs could be assigned specifically to protons on one particular side of the aromatic ring on the basis of structures obtained from DG by using pseudo-atoms for these protons, in which the rings already showed a definite orientation with respect to surrounding residues.

The NOE data were supplemented with hydrogen-bond constraints from the previously identified helices (residues 27–37 and 61–70) and  $\beta$ -sheet regions (residues 7–9 and 23–25) (Katahira et al., 1992), restraining the distances between the carbonyl oxygen of residue (i) and the amide proton of residue (i+4) within a range of 1.85–2.30 Å. On the basis of homology with the glucocorticoid receptor DNA-binding domain, six sulphur–sulphur constraints per zinc finger were applied to cysteines 8, 11, 25 and 28 in the first finger and 44, 50, 60 and 63 in the second finger, assuming a tetrahedral coordination of the zinc and a 2.35 Å zinc–sulphur bond distance. Structures generated without these constraints had similar zinc coordination and  $\alpha$ -helical regions although less well determined backbone conformations.

#### *Structure calculations*

The metric matrix distance geometry program, derived from the original DG program (Havel

et al., 1983), was modified to enable it to include torsion angle information. Torsion angle restraints were translated to lower and upper bound distance restraints, as well as to chiral restraints on the atoms involved as described by Havel and Wüthrich (1985). Thirty structures were generated and optimized against distances and chiralities in 500 steps, using conjugate gradient minimization. Of the 30 structures, 29 with a correct chirality at the C-atoms were selected and subjected to further optimization, using distance-bound driven dynamics (DDD) (Scheek and Kaptein, 1988) running for 500 steps at 300 K and 500 steps at 1 K. The time step used in the DDD calculation was 20 ps and a temperature coupling constant of 200 ps was applied. The constraint force was  $10^4 \text{ kJ mol}^{-1} \text{ nm}^{-4}$ .

For restrained energy minimization (REM) and restrained molecular dynamics (RMD) calculations, zinc atoms were placed at the average positions of the cysteine sulphur atoms involved, and zinc–sulphur restraints of 2.35 Å were added to the restraints list. Since the GROMOS force field does not contain sulphur–zinc bonding with tetrahedral geometry, the van der Waals repulsion term for the zinc–sulphur bond was removed from the force field to avoid large repulsion energies caused by imperfections in the force field, and the tetrahedral geometry was retained by sulphur–sulphur and sulphur–zinc distance restraints. The 29 structures obtained were energy-minimized in the GROMOS standard vacuum force field, using a force constant of  $40 \text{ kJmol}^{-1} \text{ Å}^{-2}$  for the harmonic distance restraining potential and one of  $110 \text{ kJmol}^{-1}$  for the cosine dihedral restraining potential followed by 10 ps of RMD and REM. A time step of 1 fs was used while the system was coupled to a heating bath with a 0.1 ps coupling constant. A cut-off of 8 Å was used and the non-bonded pair list was updated every 10 steps. Initial velocities were randomly taken from a Maxwellian distribution at 300 K.

#### *Iterative relaxation matrix approach (IRMA)*

Since the experimental NOE matrix is usually incomplete, it is not possible to calculate inter-proton distances corrected for spin diffusion directly by diagonalization of the NOE matrix. If, however, a model structure is available, the missing NOEs can be calculated from the coordinates and spectral densities by using standard matrix techniques (Macura and Ernst, 1980). Mobility effects such as methyl rotation and aromatic ring flips can be described by averaging the corresponding elements of the relaxation matrix (Koning et al., 1990). Fast local motions can be included by using order parameters to describe the spectral densities, following the model free approach of Lipari and Szabo (Lipari and Szabo, 1982; Koning et al., 1991). The complete NOE matrix, made up of both experimental NOEs and calculated model-based NOEs, is transformed back to the corresponding relaxation matrix from which new distances are obtained. These distances are used for RMD simulations, and the improved structure obtained is used as input for a new IRMA cycle (Boelens et al., 1988). This procedure is repeated until convergence is reached. In our study, the ensemble of 29 structures was used to obtain averaged inter-proton distances, and from them theoretical NOEs, which were combined with experimental intensities, following the ‘ensemble IRMA’ protocol (Bonvin et al., in press).

For IRMA, a  $\tau_c$  of 4.5 ns and a leakage rate of  $1.5 \text{ s}^{-1}$  were used. Tyrosine  $\text{H}\epsilon\text{-H}\delta$ , helix  $\text{H}\alpha\text{-NH}$  ( $i, i+3$ ) and  $\text{H}\alpha\text{-H}\beta$  ( $i, i+3$ ), and the  $\text{Ala}^{24}$   $\text{H}\alpha\text{-H}\beta$  NOE intensities were used for scaling of experimental and theoretical matrices. Relaxation matrix elements that differed more than the standard deviation from the average were set to 4.5 Å and included in the restraint set after adding pseudo-atom corrections. The cycle of energy minimization, 10 ps molecular dynamics and

energy minimization was then repeated, using the parameters mentioned above. After two IRMA cycles, the best 15 structures in terms of potential and restraint energy were selected and used for further IRMA calculations. Model-based stereospecific assignments as proposed by Bonvin et al. (in press) were obtained for 17 residues on the basis of NOE intensities in the 100 and 150 ms data sets. For the last two IRMA cycles, lower bound restraints were included in the calculations.

In order to include the effects of fast local motions into the IRMA calculations, order parameters  $S^2$  were included in the fifth IRMA cycle. These order parameters were calculated from the plateau values of the internal correlation functions of proton-proton vectors, which were obtained from a free molecular dynamics simulation of a low-energy DNA-binding domain structure in a box with 2062 waters and 13 chloride ions. Zinc coordination was maintained in the free molecular dynamics simulation by applying distance restraints between cysteine sulphur atoms and zinc ions, using a force constant of  $45 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$ . After position restrained energy minimization with a force constant of  $80 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$ , 10 ps of free molecular dynamics followed with a 2-fs time step and temperature and pressure relaxation times of 0.01 ps and 0.05 ps, respectively. For the next 110 ps, a time step of 1 fs and relaxation times of 0.1 ps and 0.5 ps were used. Structures were saved every 0.05 ps and the last 60 ps of trajectory was used to calculate internal correlation functions for all inter-proton vectors shorter than 6 Å. Order parameters were computed from the average correlation functions between  $t=10$  and  $t=30$  ps, as described by Bonvin et al. (in press).

All DG, DDD, IRMA and RMD calculations were performed on Silicon Graphics 4D/35 and 4D GTX workstations. Free molecular dynamics simulations in water were performed on a Convex 210 computer. Structures were examined on Silicon Graphics IRIS workstations using BIOSYMs Insight II software.

## RESULTS AND DISCUSSION

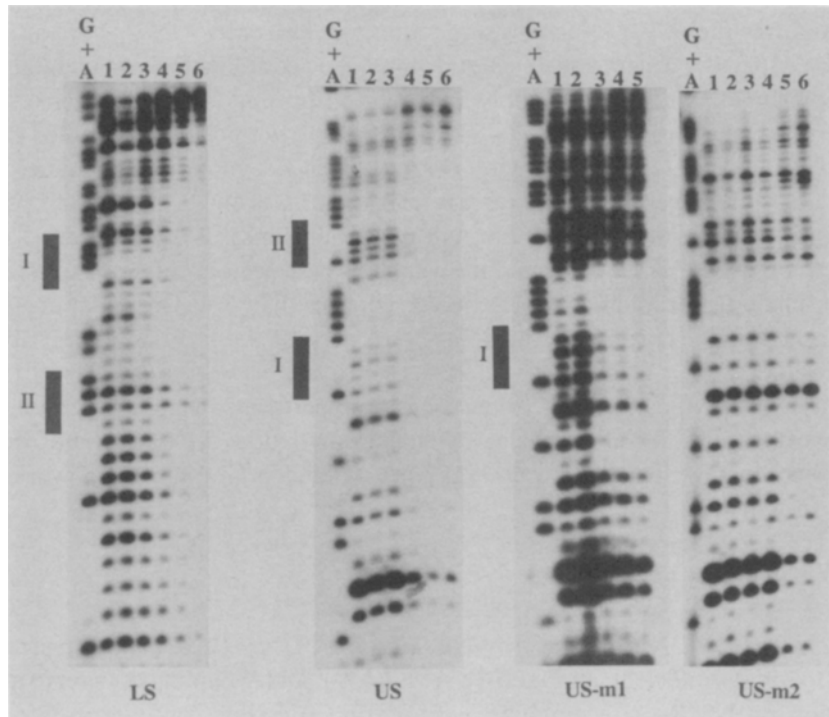
### *Molecular cloning and DNA binding properties of the hRAR- $\beta$ DNA-binding domain*

In order to obtain sufficient amounts of pure hRAR- $\beta$  DNA-binding domain to allow structure determination by NMR, a recombinant vector (pET3) coding for residues 75 to 153 of the hRAR- $\beta$  was constructed. The polypeptide encoded comprised the 79-residue DNA-binding domain C of the hRAR- $\beta$  with an additional N-terminal methionine (see Fig. 1). The polypeptide was overexpressed in *Escherichia coli* using the T7 expression system (Studier and Moffat, 1986) and subsequently isolated to 95% purity, as verified by SDS-PAGE (results not shown).

To demonstrate that the bacterially expressed RAR DNA-binding domain peptide still contained the sequence-specific DNA-binding properties of its parent molecule, purified RAR DNA-binding domain was used in a DNase I protection assay. Incubation of a  $^{32}\text{P}$ -labelled DNA fragment that contains a retinoic acid responsive element from the hRAR $\beta$  gene (De Thé et al., 1990) with the RAR DNA-binding domain peptide resulted in protection of the sequence spanning the retinoic acid response element against digestion by DNase I. The results of such an experiment are shown in Fig. 2A for both the upper and the lower strand of the labelled probe. The protected region included the direct repeat that makes up the retinoic acid response element. To show that this protection was the result of sequence-specific DNA binding, two other probes were used in similar experiments. The first, designated RARE-m1, contains a retinoic acid response element that has been mutated in one of the two half sites (see Fig. 2B for its exact

sequence). This change resulted in a complete disappearance of protection specifically around the mutated half site. Mutation of both half sites completely abolished the protection of the RAR DNA-binding domain against DNase I attack, as shown by the results of the protection assay with the probe RARE-m2. These results clearly show that protection against DNase I digestion is the result of sequence-specific binding of the RAR DNA-binding domain peptide to its cognate sequence.

**A**



**B**

$\beta$ -RARE

```

||| ||||| |||| | ||||| | | |||
AGGTCGAGGGTAGGGTTCACCGAAAGTTCACCTCGTCGACTCT
TCCAGCTCCCATCCCAAGTGGCTTCAAGTGAGCAGCTGAGA
| | | | | | | | | | | | | | | | | |

```

$\beta$ -RARE m1

```

||| ||||| ||| | | ||||| | | |||
AGGTCGAGGGTAGGcTTacCCGAAAGTTCACCTCGTCGACTCT

```

$\beta$ -RARE m2

```

||| ||||| ||| | | ||||| | | |||
AGGTCGAGGGTAGGcTTacCCGAAAcTTacCTCGTCGACTCT

```



TABLE 1  
ROTAMERS AND REFERENCE DIHEDRAL ANGLES FOR STEREOSPECIFICALLY ASSIGNED RESIDUES

Residue	Rotamer <sup>a</sup>	H $\beta$ 2 or H $\gamma$ 1	H $\beta$ 3 or H $\gamma$ 2	Reference angle $\chi$ (degrees)
Phe <sup>9</sup>	g <sup>2</sup> t <sup>3</sup>	3.29	3.15	180
Val <sup>10</sup>	g <sup>+</sup>	0.89	1.11	-60
Gln <sup>12</sup>	g <sup>2</sup> t <sup>3</sup>	2.50	2.57	180
Val <sup>22</sup>	t	0.76	0.66	180
Ser <sup>23</sup>	g <sup>2</sup> t <sup>3</sup>	3.13	3.53	180
Cys <sup>25</sup>	g <sup>2</sup> t <sup>3</sup>	2.82	3.51	180
Phe <sup>31</sup>	g <sup>2</sup> t <sup>3</sup>	3.28	3.04	180
Phe <sup>32</sup>	g <sup>2</sup> t <sup>3</sup>	3.42	2.93	180
Gln <sup>37</sup>	g <sup>2</sup> t <sup>3</sup>	2.15	1.97	180
Lys <sup>38</sup>	g <sup>2</sup> g <sup>3</sup>	2.01	1.68	60
Asn <sup>39</sup>	g <sup>2</sup> g <sup>3</sup>	2.99	2.62	60
Met <sup>40</sup>	g <sup>2</sup> g <sup>3</sup>	1.22	1.38	60
Tyr <sup>42</sup>	g <sup>2</sup> g <sup>3</sup>	2.77	3.53	60
Asp <sup>47</sup>	g <sup>2</sup> g <sup>3</sup>	2.82	2.53	60
Asn <sup>49</sup>	g <sup>2</sup> g <sup>3</sup>	2.75	2.65	60
Cys <sup>50</sup>	g <sup>2</sup> t <sup>3</sup>	2.94	2.65	180
Val <sup>51</sup>	g <sup>+</sup>	0.89	1.06	-60
Asn <sup>53</sup>	g <sup>2</sup> g <sup>3</sup>	3.31	2.78	60
Val <sup>55</sup>	g <sup>+</sup>	0.94	1.02	-60
Cys <sup>63</sup>	g <sup>2</sup> g <sup>3</sup>	3.08	2.69	60
Phe <sup>69</sup>	t <sup>2</sup> g <sup>3</sup>	3.23	3.42	-60
Met <sup>73</sup>	t <sup>2</sup> g <sup>3</sup>	0.58	0.76	-60
Val <sup>78</sup>	g <sup>+</sup>	0.89	0.58	-60

<sup>a</sup> Rotamers g<sup>2</sup>g<sup>3</sup>, g<sup>2</sup>t<sup>3</sup> and t<sup>2</sup>g<sup>3</sup> denote *gauche* and *trans* positions of the  $\beta$ 2 and  $\beta$ 3 protons with respect to the H $\alpha$  proton. For valines the rotamers g<sup>-</sup>, t and g<sup>+</sup> denote the *trans* or *gauche* position of the  $\gamma$ 1-methyl group with respect to the H $\alpha$  proton.

Although the region of the non-mutated probe protected against DNase I attack completely spans the sequence of the retinoic acid response element, there were two short stretches on both

←

Fig. 2. Analysis of protection against DNase I attack by RAR DNA-binding domain. (A) Footprint assays using probes containing RAR binding sites. Synthetic dsDNA fragments were used as probes in a DNase I protection assay, radioactively labelled on the upper (US) or lower (LS) strand as indicated for the various experiments shown. Names of probes correspond to the sequences given below in panel B. Lanes marked G+A contain a purine-specific sequence reaction (Maxam and Gilbert, 1980) performed on the same DNA fragment that was used in the protection assay as sequence-specific size markers. Lanes 1 and 2 contain labelled probe in binding mix without added RAR DNA-binding domain protein, digested with 1/2 or 2 units of DNase I, respectively. Lanes 3–6 contain probes, bound to an increasing amount of RAR DNA-binding domain protein, digested with 2 U of enzyme. Lane 3: 50 ng; lane 4: 100 ng; lane 5: 150 ng and lane 6: 200 ng of protein. (B) Mapping of protected areas on the sequences of the probes. The protection found in the footprinting experiments shown in A is mapped on the nucleotide sequences of the probes that were used. The purine residues in the labelled strands are indicated by vertical dashes over the sequence to facilitate easy reference to panel A. The half sites of the retinoic acid response element are underlined; mutations in this sequence that abolish protein binding are indicated by small capital letters instead of capital letters in the sequence lettering. The protected areas are indicated by horizontal lines. Dotted lines are used wherever the exact ending of the protection could not be established.

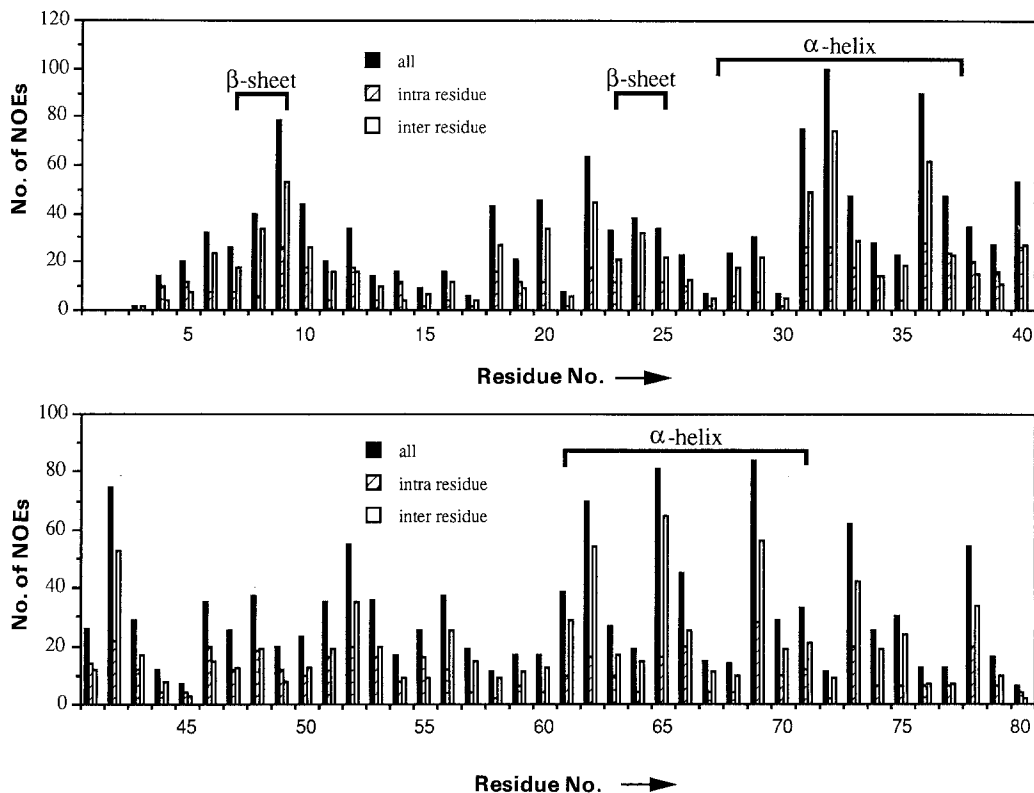


Fig. 3. The number of NOEs as a function of residue number. Besides the total number of NOEs, the numbers of intra- and inter-residue NOEs are also shown. Inter-residue NOEs were counted for both residues involved.

strands within this region that were not protected or only partly protected. These are indicated in Fig. 2B, where the protected area is mapped on the probe's sequence. On the map of the m1 probe

TABLE 2  
ENERGIES OF THE 15 BEST STRUCTURES OF THE hRAR- $\beta$  DNA-BINDING DOMAIN IN SUCCESSIVE STAGES OF THE REFINEMENT

Structures	$E_{\text{pot}}^a$ (kJ mol $^{-1}$ )	$E_{\text{res}}^a$ (kJ mol $^{-1}$ )	$E_{\text{dih}}^a$ (kJ mol $^{-1}$ )
DG/DDD/RMD	-3599	3500	267
IRMA 1	-3921	3140	288
IRMA 2	-3954	2970	279
IRMA 3	-3930	2320	292
IRMA 4	-4010	2060	244
IRMA 5	-4007	1825	257

<sup>a</sup> GROMOS energies in kJ mol $^{-1}$ . The potential energy ( $E_{\text{pot}}$ ) does not include restraint energies. Energies for distance restraints ( $E_{\text{res}}$ ) and dihedral angles ( $E_{\text{dih}}$ ) were calculated with the final restraint set for all cycles in order to obtain comparable numbers.

a corresponding gap can be seen in the protection of the non-mutated half site. This feature of the DNase I protection assays can be understood if it is realized that DNase I interacts with the minor groove of DNA (Suck and Oefner, 1986). In the case of a single half site or two separated half sites, the DNase I molecule could attack the DNA minor groove opposite to the binding site of the RAR DNA-binding domain.

As the purified hRAR- $\beta$  DNA-binding domain proved to be functionally intact, we proceeded to record NMR spectra. Sequence-specific assignments of nitrogen-15 and proton resonances for all but the first three N-terminal residues were obtained (Katahira et al., 1992). The proton resonance assignment allowed us to assign medium- and long-range NOEs, which are essential for tertiary structure determination

#### *Stereospecific assignments and experimental restraints*

In total 23 residues could be assigned stereospecifically on the basis of their  $^3J_{\alpha\beta}$  coupling constants and NH- $\beta/\alpha$ - $\beta$  NOEs. The  $\chi_1$ -angles of these residues were restrained to the appropriate values and are listed in Table 1. Of 30 residues the  $\phi$ -angle could be restrained to  $-120^\circ$  (residues

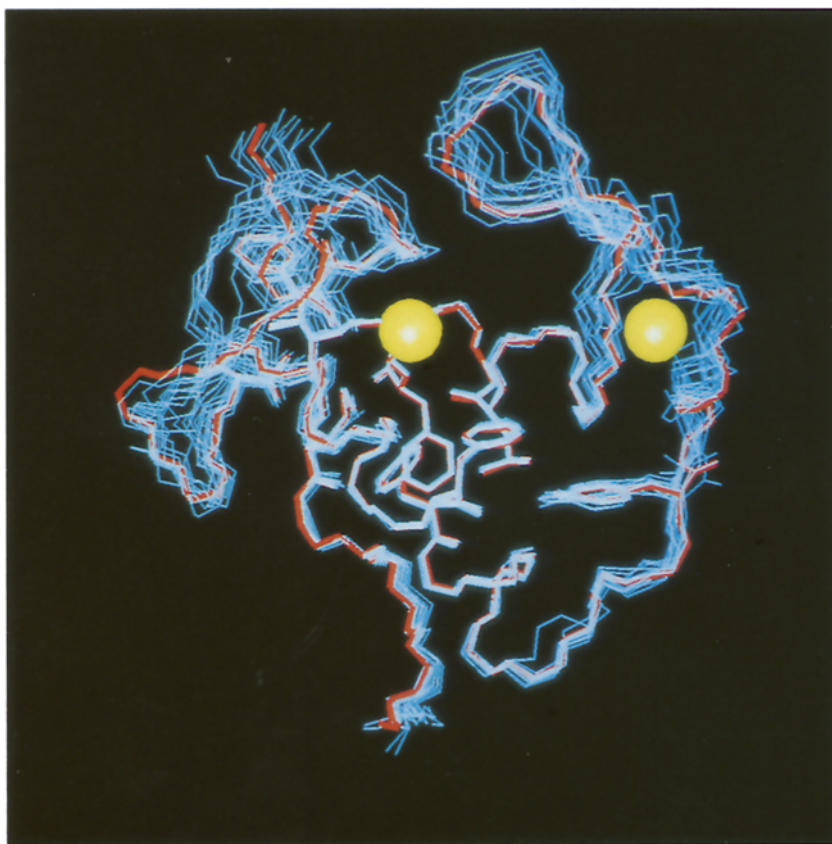


Fig. 4. Polypeptide backbone traces of the 15 lowest energy structures of the RAR DNA-binding domain. Side chains are shown for core residues 9, 22, 31, 32, 42, 65, 69, 71 and 73. Superpositioning was performed only on the helices. The structure with the lowest energy is drawn in red; zinc atoms are in yellow.

5, 10–12, 15, 22, 24, 25, 38–41, 46–49, 51, 53–56, 59, 71, 74–80) on basis of their  $^3J_{\text{H}\alpha}$  coupling constant.

Integrated NOE cross peaks obtained from 2D NOE spectra recorded in H<sub>2</sub>O and D<sub>2</sub>O with mixing times of 50, 100, 150 and 200 ms were translated into distance restraints by determining the initial rates of NOE intensity build-up. A total of 1244 NOE-derived distances were used, of which 448 were intra-residue and 796 inter-residue. The restraint set contained 273 sequential ( $i-j=1$ ), 190 medium-range ( $2 \leq i-j \leq 4$ ) and 333 long-range ( $i-j > 4$ ) restraints. For 18 NOEs, good integrals could not be obtained because of overlap or noise, and they were added as 4.5 Å distances to the restraint list. Figure 3 shows the distribution of NOEs per residue for the hRAR-β DNA-binding domain. Lower bounds were set to 2.0 Å. As shown in Fig. 3, most long-range NOEs were found between hydrophobic residues in, or close to, the two α-helices and the β-sheet, whereas both zinc fingers showed relatively few long-range NOEs. They contacted each other, however, through residues near Cys<sup>11</sup> and Ile<sup>52</sup>. A full list of the restraints used in the calculations is available on request as supplementary material.

#### *The three-dimensional structure of the h-RARβ DNA-binding domain*

A set of 30 structures was generated using a metric matrix distance geometry algorithm and optimized using a conjugate gradient method against distances and chiralities. One structure was discarded because of wrong chiralities on the C-atoms; the other 29 were used as input for a DDD calculation (Scheek and Kaptein, 1988) to obtain a better sampling of the allowed conformational space. The structures obtained after DDD were used as input for subsequent cycles of REM, 10-ps restrained RMD and REM again. For the first cycle, initial rate distances were used and IRMA-derived restraints were applied in the following four cycles. The 15 best structures with model-based stereospecific assignments were used in the third IRMA cycle and additional lower bound restraints were applied after the fourth and fifth IRMA cycle. Table 2 gives an overview of some relevant energies during refinement. Since the number and magnitude of the distance restraints changed during refinement, because of the inclusion of model-based stereospecific assignments and lower bound restraints, the restraint energies were calculated by using the final distance restraint set in order to obtain mutually comparable values. The final average potential energy is  $-4007 \pm 173$  kJ. The final distance restraint energy is  $1825 \pm 86$  kJ, corresponding to an average violation of 0.06 Å and a total violation of 14.9 Å for 2606 upper and lower bound restraints. Most violations were found in the two fingers, which might be due to dynamic processes. There were 12 violations larger than 0.5 Å, the largest being 0.7 Å for a restraint between the two fingers. The final dihedral restraint energy was 257 kJ. The geometrical ideal for the bonds was 0.002 Å and for bond angles 2.8°. Figure 4 shows the backbone trace of these 15 structures and the side chains of the core residues 9, 22, 31, 32, 42, 65, 69, 71 and 73, after superposition on the helices. The α-helical core of the protein is well determined, with a backbone r.m.s.d. with respect to the average of 0.18 Å and 0.37 Å when the side chains of residues 31, 32, 36, 61, 62, 65 and 69 are included. The r.m.s.d. with respect to the average for the whole backbone, excluding residues 1 through 4 for which no long-range NOEs were observed, is 0.76 Å. For the first finger (residues 8–28), the r.m.s.d. of the backbone is 0.79 Å. For the second finger (residues 44–62), the r.m.s.d. is 0.64 Å. That the second finger was determined more accurately is due to the relatively large number of φ-angles that could be restrained in this region and the lower number of NOEs per residue in the first finger (see Fig. 3) causing a large variability around residue 14. The finger

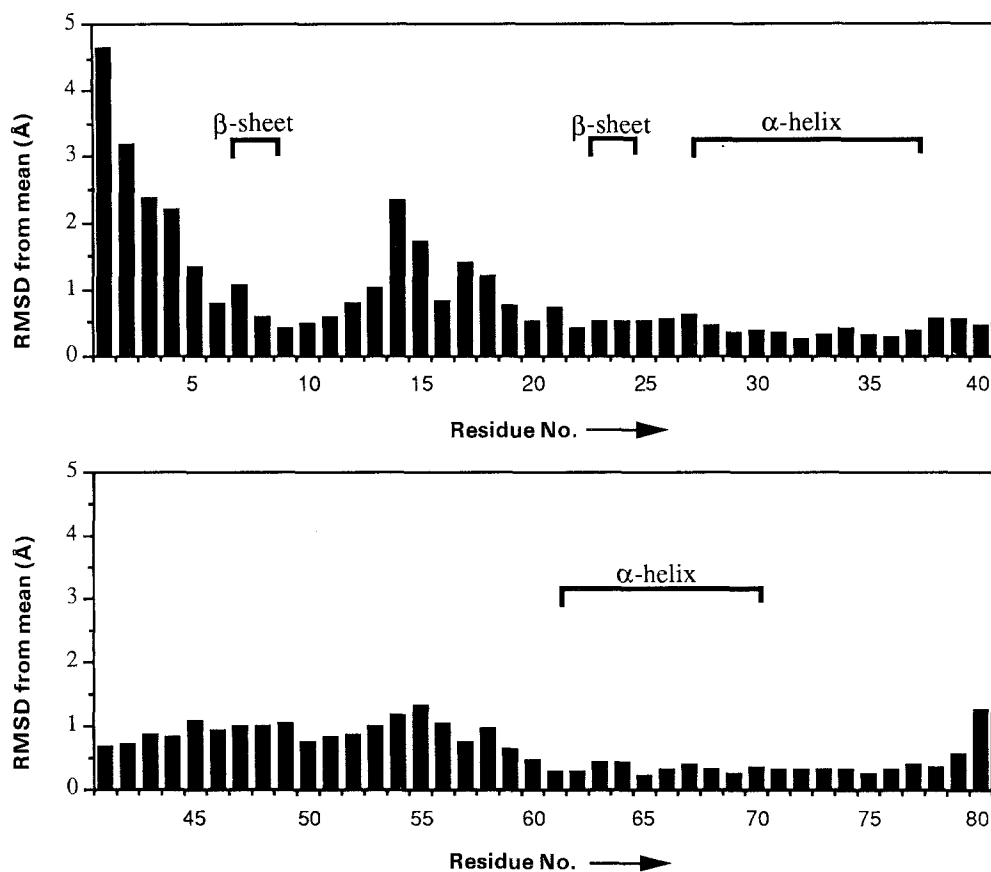


Fig. 5. Root-mean-square deviation of  $C_{\alpha}$  atom positions from the mean. Averaging was done over the 15 lowest energy structures after superposition on the backbone.

loops appeared to have a relatively well determined structure, but the loop as a whole had a higher mobility due to domain motions with respect to the core. A total of four non-glycine residues (13, 26, 57 and 58) were systematically found to lie in the forbidden region of the Ramachandran plot. The conformation of Glu<sup>26</sup> might be necessary to make the connection between the  $\beta$ -sheet (running up to residue 25) and the first  $\alpha$ -helix (starting from residue 27). Coordinates of the structures and NOE restraints will be deposited in the Brookhaven Protein Database.

Figure 5 shows the deviation from the mean of the  $C_{\alpha}$  atom position for the 15 best structures. It can be seen that the helical core was well determined by the restraint set, while the two fingers and the N-terminus showed a higher conformational variability. The higher variability is probably due to an increased flexibility, which was also observed in the  $S^2$  values of the free molecular dynamics calculations. As expected, there is a good correspondence between a low r.m.s.d. and the presence of elements of secondary structure.

Figure 6 shows a more detailed view of the human RAR- $\beta$  DNA-binding domain. The two  $\alpha$ -helices are perpendicular to each other, exposing hydrophilic residues to the solvent. They are connected to each other through a hydrophobic core made up of the aromatic and non-polar

residues located in the two helices. Near the short  $\beta$ -sheet region, NOE contacts between hydrophobic residues involve residues at opposite sides of the  $\beta$ -sheet and directly following the second helix. The two zinc atoms in the N- and C-terminal fingers have S and R chirality, respectively, which has also been found for the glucocorticoid receptor in solution (Hård et al., 1990) and in crystal form (Luisi et al., 1991).

The availability of the tertiary structure of the RAR DNA-binding domain makes it possible for us to explain some striking features of the NMR spectra. The NH proton of Ser<sup>23</sup> was found to be shifted to high field in the NMR spectra. This can be explained on the basis of the structure in which the Ser<sup>23</sup> amide proton is found above the plane of the aromatic ring of Phe<sup>9</sup> at a distance of roughly 2–3 Å. The aromatic ring of Tyr<sup>18</sup> is close, and its plane is oriented towards Ser<sup>23</sup>. The orientation of the Phe<sup>9</sup> side chain could well induce a ring current shift, which would cause the unusual shift of 5.28 ppm.

A slowly exchanging hydroxyl proton was observed in the 2D NOE spectra recorded in <sup>1</sup>H<sub>2</sub>O, which could be identified as the Thr<sup>56</sup>  $\gamma$ -hydroxyl group via a HOHAHA connectivity to the



Fig. 6. Ribbon model of the RAR DNA-binding domain. Non-polar side chains that form the hydrophobic core are shown with their van der Waals surfaces coloured yellow. The two zinc finger subdomains are coloured red and green, respectively.

$\gamma$ -methyl group of the same residue. Either the Asn<sup>53</sup> backbone carbonyl oxygen or the Asn<sup>53</sup> side-chain carbonyl oxygen could act as proton acceptors.

Comparison of the RAR DNA-binding domain to the solution and crystal structures of the glucocorticoid receptor DNA-binding domain (Hård et al., 1990, Luisi et al., 1991) showed that there are some differences. In the RAR solution structure, the second helix, running from residue Gln<sup>61</sup> to Glu<sup>70</sup>, is two residues shorter than in the glucocorticoid receptor (Hård et al., 1990; Katahira et al., 1992). In contrast to the glucocorticoid receptor, the <sup>1</sup>H resonances of the C-terminal extended chain directly following the second helix could be assigned completely in RAR. Together with the observation of long-range NOEs from Val<sup>78</sup> to residues in the first helix, this suggests that the C-terminus is folded against the protein surface and not directed into solution. This folding of the C-terminus of the RAR could account for the second  $\alpha$ -helix being two residues shorter than that observed in the glucocorticoid receptor. Furthermore, in the Thr<sup>43</sup>–Arg<sup>49</sup> segment of the RAR solution structure, two turns were observed which were approximately perpendicular. A similar folding of this region, which is responsible for protein–protein interaction in the complex, is also observed in the glucocorticoid receptor crystal structure (residues C476–C482) but in the glucocorticoid receptor solution structure it is less structured, with both turns oriented differently with respect to each other.

Overall, however, the solution structures of the RAR and glucocorticoid receptor DNA-binding domains are very similar. When the solution structures of the glucocorticoid receptor and RAR are superimposed on the helices, the pairwise r.m.s.d. of the N, C $\alpha$ , C backbones of the helices is 1.3 Å. When the RAR compared with the X-ray structure of the glucocorticoid receptor, the helix backbone r.m.s.d. is 1.1 Å. The largest differences in conformation are observed in the second finger that contacts the DNA in the glucocorticoid receptor crystal structure. Here, the overall fold of the finger in the RAR seems to be more similar to that in the glucocorticoid receptor crystal structure than to that in the glucocorticoid receptor solution structure, but this region is not accurately determined in the NMR structures because of a lack of NOEs. A short distorted  $\alpha$ -helix was found in the second finger in the X-ray structure of the glucocorticoid receptor DNA-binding domain. In the RAR DNA-binding domain there is a small segment running from about Thr<sup>56</sup> up to Arg<sup>59</sup> which shows one distorted right-hand turn, but it is clearly not a short  $\alpha$ -helical segment as found in the X-ray structure. It could be that the distorted  $\alpha$ -helix is stabilized by contacts with the DNA and exists only in a rudimentary form in the free protein. Although the RAR DNA-binding domain is structurally very similar to the DNA-binding domain of the glucocorticoid receptor, it binds to retinoic acid response elements consisting of direct repeats of the half site (with a 5-bp spacer) in contrast to the inverted repeat glucocorticoid response element with a 3-bp spacer in between half sites. Since in case of the retinoic acid response element the DNA-binding domains will be further apart and orientated differently with respect to each other, it is hard to imagine the D box playing a similar role in dimerization in the DNA-binding domain of RAR as was found for the DNA-binding domain of the glucocorticoid receptor.

## ACKNOWLEDGEMENTS

We thank Dr. P. Chambon (Strassbourg) for the hRAR- $\beta$  cDNA, Dr. H.G. Stunnenberg (EMBL, Heidelberg) for making his manuscript available to us before publication, and Dr. M.L.

Ganadu for performing the atomic absorption experiments. This research was supported by the Netherlands Organization for Chemical Research (SON) and the Netherlands Organization of Scientific Research (NWO). The 600-MHz  $^1\text{H}$  NMR spectra were recorded at the National Dutch HF-NMR facility in Nijmegen with the assistance of Dr. S. Wijminga and Mr. J. Joordens, M.K. was supported by a grant from the Human Frontier Science Program for Long-Term Fellowships. J.G.S. was supported by a fellowship of the Dutch Cancer Society.

## REFERENCES

- Bax, A., Griffey, R.H. and Hawkins, B.L. (1983) *J. Am. Chem. Soc.*, **105**, 7188–7190.
- Beato, M. (1989) *Cell*, **56**, 335–344.
- Benbrook, D., Lernhardt, E. and Pfahl, M. (1988) *Nature*, **333**, 669–672.
- Berg, J.M. (1989) *Cell*, **57**, 1065–1069.
- Boelens, R., Koning, T.M.G. and Kaptein, R. (1988) *J. Mol. Struct.*, **173**, 299–311.
- Bonvin, A.M.J.J., Rullmann, J.A.C., Lamerichs, R.M.J.N., Boelens, R. and Kaptein, R. *Proteins: structure, function and genetics*, in press.
- Brand, N.J., Petkovich, M., Krust, A., Chambon, P., de Thé, H., Marchio, A., Tiollais, P. and Dejean, A. (1988) *Nature*, **332**, 850–853.
- Danielsen, M., Hinck, L. and Ringold, G.M. (1989) *Cell*, **57**, 1131–1138.
- De Thé, H., Marchio, A., Tiollais, P. and Dejean, A. (1987) *Nature*, **330**, 667–670.
- De Thé, H., del Mar Vivanco-Ruiz, M., Tiollais, P., Stunnenberg, H. and Dejean, A. (1990) *Nature*, **343**, 177–180.
- Evans, R.M. (1988) *Science*, **240**, 889–895.
- Freedman, L.P., Luisi, B.F., Korszun, Z.R., Basavappa, R., Sigler, P.B. and Yamamoto, K.R. (1988) *Nature*, **334**, 543–546.
- Giguere, V., Ong, E.S., Sequi, P. and Evans, R.M. (1987) *Nature*, **330**, 624–627.
- Green, L.M. and Berg, J.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 4047–4051.
- Härd, T., Kellenbach, E., Boelens, R., Maler, B.A., Dahlman, K., Freedman, L.P., Carlstedt-Duke, J., Yamamoto, K.R., Gustafsson, J. and Kaptein, R. (1990) *Science*, **249**, 157–160.
- Havel, T.F., Kuntz, I.D. and Crippen, G.M. (1983) *Bull. Math. Biol.*, **45**, 665–720.
- Havel, T.F. and Wüthrich, K. (1985) *J. Mol. Biol.*, **182**, 281–294.
- Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1982) *J. Chem. Phys.*, **71**, 4546–4553.
- Katahira, M., Knegtel, R.M.A., Boelens, R., Eib, D., Schilthuis, J.G., van der Saag, P.T. and Kaptein, R. (1992) *Biochemistry*, **31**, 6474–6480.
- Kliwer, S.A., Umesono, K., Mangelsdorf, D.J. and Evans, R.M. (1992) *Nature*, **355**, 446–449.
- Koning, T.M.G., Boelens, R. and Kaptein, R. (1990) *J. Magn. Reson.*, **90**, 111–123.
- Koning, T.M.G., Boelens, R., van der Marel, G.A., van Boom, J.H. and Kaptein, R. (1991) *Biochemistry*, **30**, 3787–3797.
- Krust, A., Kastner, P., Petkovich, M., Zelent, A. and Chambon, P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5310–5314.
- Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) *Meth. Enzymol.*, **154**, 367–382.
- Leid, M., Kastner, P., Lyons, R., Nakshatri, H., Saunders, M., Zacharewski, T., Chen, J.-Y., Staub, A., Garnier, J.-M., Mader, S. and Chambon, P. (1992) *Cell*, **68**, 377–395.
- Lipari, G. and Szabo, A. (1982) *J. Am. Chem. Soc.*, **104**, 4546–4570.
- Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R. and Sigler, P.B. (1991) *Nature*, **352**, 497–505.
- Macura, S. and Ernst, R.R. (1980) *Molec. Phys.*, **41**, 95–117.
- Mader, S., Kumar, V., de Verneuil, H. and Chambon, P. (1989) *Nature*, **338**, 271–274.
- Marion, D. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.
- Maxam, A. and Gilbert, W. (1980) *Meth. Enzymol.*, **65**, 499–560.
- Miller, J., McLachlan, A.D. and Klug, A. (1985) *EMBO J.*, **4**, 1609–1614.
- Müller, L. (1987) *J. Magn. Reson.*, **72**, 191–196.
- Petkovich, M., Brand, N.J., Krust, A. and Chambon, P. (1987) *Nature*, **330**, 444–450.
- Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 479–485.



- Scheek, R.M. and Kaptein, R. (1988) In *Methods in Enzymology*, Vol. 117 (Eds, Oppenheimer, N.J. and James, J.L.) Academic Press, New York pp. 204–208.
- Schwabe, J.W.R., Neuhaus, D. and Rhodes, D. (1990) *Nature*, **348**, 458–461.
- States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) *J. Magn. Reson.*, **48**, 286–297.
- Studier, F.W. and Moffat, B.A. (1986) *J. Mol. Biol.*, **189**, 113–130.
- Suck, D. and Oefner, C. (1986) *Nature*, **321**, 620–625.
- Thaller, C. and Eichele, G. (1987) *Nature*, **327**, 625–628.
- Umesono, K. and Evans, R.M. (1989) *Cell*, **57**, 1131–1146.
- Umesono, K., Murakami, K.K., Thompson, C.C. and Evans, R.M. (1991) *Cell*, **65**, 1255–1266
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
- Yu, V.C., Delsert, C., Andersen, B., Holloway, J.M., Devarey, O.V., Näär, A.M., Kim, S.Y., Boutin, J.-M., Glass, C.K. and Rosenfeld M.G. (1991) *Cell*, **67**, 1251–1266.
- Zhang, X., Hoffmann, B., Tran, P.B.-V., Graupner, G. and Pfahl, M. (1992) *Nature*, **355**, 441–446.
- Zuiderweg, E.R.P., Boelens, R. and Kaptein, R. (1985) *Biopolymers*, **24**, 601–611.