Journal of Biomolecular NMR, 5 (1995) 23–32
ESCOM

J-Bio NMR 258

Solution structure of the Oct-1 POU homeodomain determined by NMR and restrained molecular dynamics

M. Cox[a], P.J.A. van Tilborg[a], W. de Laat[a], R. Boelens[a], H.C. van Leeuwen[b], P.C. van der Vliet[b] and R. Kaptein[a,*]

[a]*Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands*
[b]*Laboratory for Physiological Chemistry, Utrecht University, Stratenum, P.O. Box 80042, 3508 TA Utrecht, The Netherlands*

Received 30 September 1994
Accepted 21 December 1994

Keywords: Homeodomains; POU proteins; Structure refinement

Summary

The POU homeodomain (POU_{hd}), a divergent member of the well-studied class of homeodomain proteins, is the C-terminal part of the bipartite POU domain, the conserved DNA-binding domain of the POU proteins. In this paper we present the solution structure of POU_{hd} of the human Oct-1 transcription factor. This fragment was overexpressed in *Escherichia coli* and studied by two- and three-dimensional homo- and heteronuclear NMR techniques, resulting in virtually complete ^{1}H and ^{15}N resonance assignments for residues 2–60. Using distance and dihedral constraints derived from the NMR data, 50 distance geometry structures were calculated, which were refined by means of restrained molecular dynamics. From this set a total of 31 refined structures were selected, having low constraint energy and few constraint violations. The ensemble of 31 structures displays a root-mean-square deviation of the coordinates of 0.59 Å with respect to the average structure, calculated over the backbone atoms of residues 6 to 54. The fold of POU_{hd} is very similar to that of the canonical homeodomains. Interestingly, the recognition helix of the free POU_{hd} ends at residue 53, while in the cocrystal structure of the intact POU domain with the DNA octamer motif [Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. and Pabo, C.O. (1994) *Cell*, **77**, 21–32] this helix in the POU_{hd} subdomain is extended as far as residue 60.

Introduction

The homeodomain is a well-known DNA-binding domain that has been extensively studied over the past decade. It was originally identified in *Drosophila melanogaster* (McGinnis et al., 1984; Scott and Weiner, 1984), as being responsible for the specific DNA recognition of many proteins that act as developmental regulators. For many *Drosophila* homeodomain proteins, a vertebrate homologue has been identified since. It is now well established as a highly conserved domain, occurring in many different organisms. The regulatory action of transcription factors depends on their sequence-specific recognition of DNA, combined with interactions with factors of the basal transcriptional machinery. Specificity of a transcription factor can be increased by increasing the number of specific protein–DNA interactions. Nature can achieve this in different ways. Oligomers of DNA-binding domains can be formed by, for example, the dimeric λ and 434 repressors or the tetrameric lac headpiece. Moreover, heterodimerisation with RXR has been reported for retinoic acid receptor (RAR), thyroid hormone receptor (TR) and vitamin D3 receptor (VD3R) before binding to the target DNA sequence (Yu et al., 1991). Another principle is exemplified by the POU domain, the DNA-binding domain of POU proteins (for reviews see Verrijzer and Van der Vliet, 1993; Wright, 1994). This bipartite domain contains an N-terminal POU-specific domain, connected via a linker to a C-terminal POU homeodomain. Both domains have DNA-binding ability, and since they are covalently linked in the POU domain, the specificity in DNA recognition is very high. As has been

*To whom correspondence should be addressed.

shown by binding-site selection experiments (Verrijzer et al., 1992), the POU domain has a preference for a stretch of nine base pairs as its target DNA sequence. The bipartite character of the POU domain has been confirmed by a crystal structure of the human Oct-1 POU domain bound to its cognate DNA sequence (Klemm et al., 1994). In this structure, direct and DNA-sequence specific protein–DNA contacts were found for nine base pairs.

In order to understand the underlying mechanisms of the regulation of gene expression at a detailed and molecular level, structural studies of high resolution are indispensable. Up until now, the three-dimensional structures of the *Antennapedia* (Billeter et al., 1990) and the *fushi tarazu (ftz)* (Qian et al., 1994) homeodomains have been determined by NMR spectroscopy. Recently, the NMR structure of Oct-2 POU_{hd} has been solved (Sivaraja et al., 1994). Furthermore, the structures of the complexes of *Antennapedia*, *Mat α2* and *engrailed* have been elucidated by NMR (Otting et al., 1990) and X-ray crystallography (Kissinger et al., 1990; Wolberger et al., 1991), respectively. These studies have shown that there is indeed a general mode for homeodomains to recognise their target DNA sequence. DNA binding is mediated by the third helix, which binds in the major groove of the DNA, while the flexible N-terminus inserts into the minor groove, employing arginine residues to contact DNA phosphate groups and bases. These features appear in all homeodomain–DNA complexes, including the complex of the POU domain. In this paper we present the three-dimensional structure of the human Oct-1 POU_{hd} domain by means of NMR spectroscopy, combined with restrained molecular dynamics calculations. This domain has a sequence identity with *Antennapedia* and *fushi tarazu* homeodomains of 33%, while the latter proteins have a mutual sequence identity of 88%, calculated over residues 1 to 58. This demonstrates that, in an evolutionary sense, POU_{hd} is distinct from the classical homeodomains. In this study, a comparison is made of the classical homeodomains and the POU_{hd} domain at a structural level, and the differences are discussed between free POU_{hd} and POU_{hd} in the intact POU domain complexed with its cognate DNA sequence.

Materials and Methods

Protein expression and purification

A gene cassette coding for the Oct-1 POU homeodomain was constructed by employing the polymerase chain reaction (PCR). The PCR product was cut at an upstream NdeI and a downstream BamHI restriction site and subsequently cloned into the expression vector pET15b. Expression of the construct resulted in a protein containing subsequently an N-terminal His_6-tag, a five-residue thrombin cleavage site, allowing for the removal of the His_6-tag after purification of the protein, and the POU_{hd}. The NMR samples contained the five-residue cleavage site, combined with the 62-residue POU_{hd}. The construct was transformed in the *Escherichia coli* strain BL21(DE3) using methods described by Studier et al. (1990). Cells were grown in a 6 l culture, and 4 h after induction they were harvested by centrifugation (4 krpm, 10 min). The pellet was suspended in 120 ml of buffer A, containing 50 mM Tris (pH = 8), 250 mM NaCl, 0.1% Nonidet P-40, 5 mM Na-MBS, 1.0 mM PMSF, 30 mg lysozyme, and 5 mM β-mercaptoethanol. After a freeze-thaw step, the cells were incubated with DNase (50 μg/ml) for 15 min at room temperature, followed by centrifugation (10 krpm, 20 min, 4 °C). The supernatant was centrifuged again (20 krpm, 20 min, 4 °C) and the resulting supernatant was purified by Ni-NTA affinity chromatography (Qiagen). The clear lysate, loaded on the Ni-column, was washed with buffer A (pH = 8). The His_6-tagged POU_{hd} protein was eluted with a 0–500 mM imidazole gradient, pH = 7. The POU_{hd} fractions from the Ni-column were pooled and diluted 1:1 with buffer I, containing 10 mM Tris (pH = 8), 1 mM DTT and 10% glycerol. The protein solution was loaded on a FastFlow-Sepharose column, which was preequilibrated with buffer I, containing 100 mM NaCl. POU_{hd} fractions were collected by using an NaCl gradient (0.1–1.0 M), and subsequently incubated with 20% thrombin for 1 h at room temperature to remove the His_6-tag. The NaCl concentration was adjusted to 125 mM by diluting the protein solution 1:3 with buffer II, containing 30 mM NaAc (pH = 5.5), 10% glycerol and 1 mM DTT. A second FastFlow-Sepharose column was loaded and washed with buffer II, containing 100 mM NaCl, 30 mM NaAc (pH = 5.5), 10% glycerol and 1 mM DTT. Protein fractions were collected by applying an NaCl gradient (0.1–1.5 M). The purity of POU_{hd} was found to be higher than 95%, as checked by SDS polyacrylamide gel electrophoresis. Protein concentrations were determined by measuring the optical density at 280 nm. A uniformly ^{15}N-labelled protein was obtained in a similar way as described above, from cells grown on 10 l of minimal medium with $^{15}NH_4Cl$ as the only nitrogen source. For NMR measurements, protein samples were concentrated to 2–3 mM, typically in 0.3 M NaCl, 3 mM DTT and 95% H_2O/5% D_2O or 100% D_2O, using Amicon dialysing equipment. The pH was adjusted to 4.7 by adding small aliquots of DCl or NaOD. As the sample appeared to be sensitive to temperatures higher than 27 °C, a protein sample was prepared containing 2.0 M d_5-glycine. This prevented denaturation of the sample at elevated temperatures, while the NMR spectrum remained essentially the same.

NMR spectroscopy

^1H NMR spectra were recorded on Bruker AMX-500, AMXT-600 and Varian Unity Plus 750 MHz spectrome-

ters. Unless mentioned otherwise, all experiments were recorded at 17 °C. The proton spectral width was set to 12.5 ppm. Water suppression of experiments recorded in H_2O was achieved by irradiation of the water resonance for 1.0 s. The proton chemical shifts are referred to the water resonance at 4.75 ppm, relative to 3-(trimethylsilyl)-propionate. The nitrogen chemical shifts are calibrated using the $^{15}NH_4Cl$ resonance at 22 ppm. A DQF-COSY spectrum (Bax and Davis, 1985) was recorded in D_2O, and a regular COSY at 24 °C, both at 500 MHz. A 'clean'-TOCSY spectrum (Griesinger et al., 1988) was recorded with a mixing time of 23 ms at 600 MHz. NOESY spectra with mixing times of 75 and 150 ms were recorded in D_2O at 500 MHz and in H_2O at 600 MHz, respectively. For the TOCSY and NOESY spectra, typically 400 t_1 increments were recorded; for COSY spectra 512 t_1 increments were collected. All homonuclear experiments were processed using shifted sine-bell or squared sine-bell window functions in both domains. Additional zero-filling in t_1 was applied to obtain a spectrum of $1K \times 1K$ data points. Processing was carried out on Silicon Graphics IRIS and INDY workstations, using the TRITON software.

The heteronuclear experiments were recorded on a Bruker AMX-500 spectrometer, equipped with a triple-resonance probe. During acquisition of the heteronuclear experiments, ^1H and ^{15}N nuclei were decoupled using a GARP sequence (Shaka et al., 1985). The heteronuclear J-coupling refocusing delay was set to 4 ms. HMQC and HSQC spectra were recorded with the ^{15}N carrier in the middle of the amide region and with a ^{15}N spectral width of 22 ppm. This way, the N^{ε} signals of the arginine residues were folded in. To determine $^3J_{HNH\alpha}$ coupling constants, a high-resolution HMQC spectrum was recorded with 700 t_1 increments. A 3D NOESY-HSQC was recorded (Marion et al., 1989) with a mixing time of 150 ms (180,64,1024 data points). This data set was processed using linear prediction in t_1 and t_2, resulting in a spectrum $(\omega_1,\omega_2,\omega_3) = (512,64,512)$. To identify slowly exchanging amide protons in HMQC spectra, the uniformly ^{15}N-labelled sample had to be quickly transferred from a 95% H_2O solution to a D_2O solution. Since the protein unfolded upon lyophilisation, the Amicon dialysing equipment was used to transfer the protein to a solution containing 95% D_2O, 4 mM DTT and 300 mM NaCl. This sample was immediately used to record four HMQC spectra of 30 min, 90 min, 4.8 h and 12.8 h, respectively. On the POU_{hd} sample with 2.0 M d_5-glycine, additional NOESY spectra were recorded at 750 MHz, with mixing times of 30 and 100 ms.

Distance and dihedral constraints

Medium- and long-range contacts were identified from the NOESY and NOESY-HSQC spectra. These contacts were used as distance constraints by classifying them as being weak, medium weak, medium, medium strong and strong, and values ranging from 2.5 to 5.0 Å were used as upper bounds. Appropriate pseudoatom corrections for methylene protons and aromatic ring protons were applied when necessary (Wüthrich et al., 1983). Constraints involving methyl groups were corrected for the three-proton intensity and a pseudoatom correction of 0.3 Å was applied (Koning et al., 1990). The lower bound of the distance constraints was set to the sum of the van der Waals radii of the involved atoms. A total of 725 distance constraints was obtained, 179 of which were medium-range contacts, and 140 were long-range. The number and type of constraints are shown per residue in Fig. 1. From the splitting of the signals in the high-resolution HMQC spectrum, a value for the $^3J_{HNH\alpha}$ coupling constant of 49 residues could be determined. These were converted to dihedral constraints for the ϕ angle in the amino acids, with a margin of $\pm 30°$. Additionally, dihedral constraints for 10 χ^1 angles were obtained from a combination of NOESY cross-peak intensities of H^{α}-H^{β} and H^N-H^{β} connectivities, and splitting patterns of COSY H^{α}-H^{β} cross

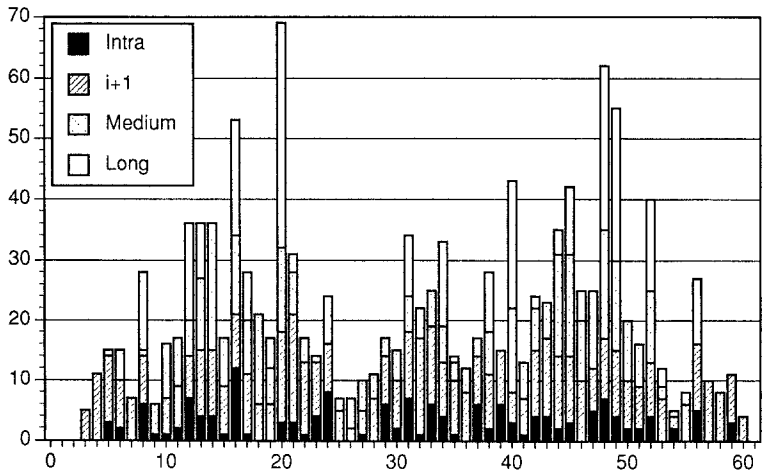

Fig. 1. Number of constraints as a function of the residue number of POU_{hd}. Four different types of constraints are indicated.

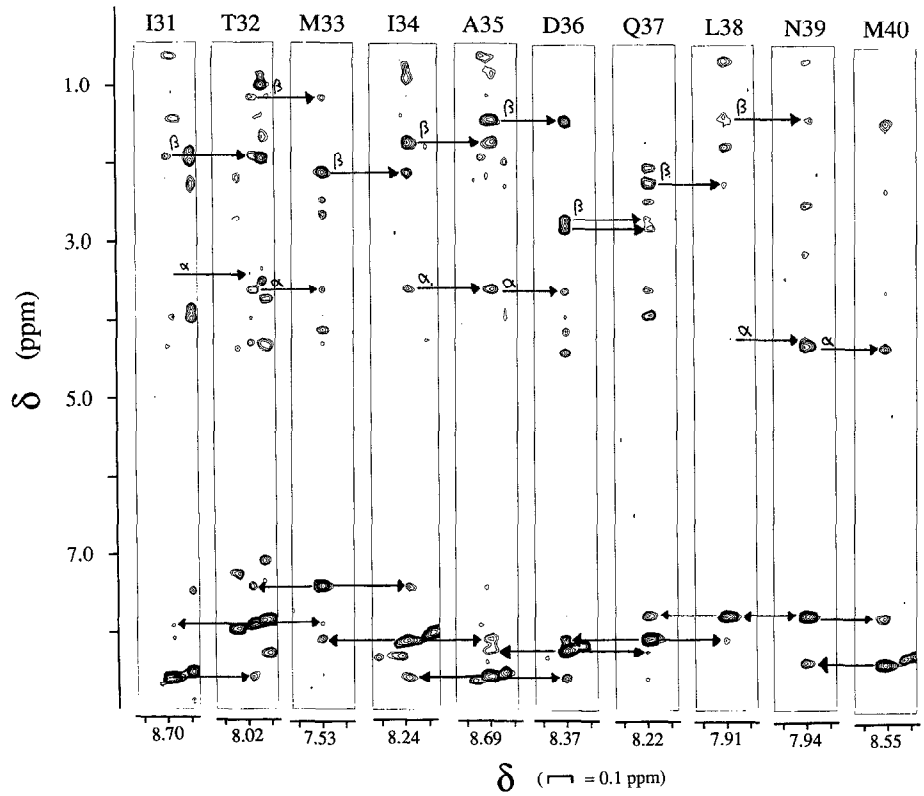

Fig. 2. Sequential assignment in the 3D NOESY-HSQC spectrum of POU_{hd}. Depicted is a composite spectrum consisting of strips taken from different ^{15}N planes. Sequential $d_{\alpha N}$, $d_{\beta N}$ and d_{NN} contacts are indicated by arrows for the stretch Ile^{31} to Met^{40}.

peaks. For eight residues, two of the three favourable rotamers could be ruled out. For the other two residues, only one rotamer could be excluded, resulting in margins of $\pm 30°$ and $\pm 90°$, respectively.

Structure calculations

For the generation of structures from the experimental data, the program DG-II (Biosym Technologies, Inc.) was used. A total of 50 structures was calculated by embedding in four dimensions, after which the coordinates were projected into three dimensions. Subsequently, the structures were optimised by subjecting them to a minimisation with a simple error function. On all 50 conformations a structure refinement was performed using the Consistent Valence Force Field implemented in the Discover software, developed by Biosym. The bond lengths were accounted for by regular harmonic potentials. The protocol consisted of a short restrained energy minimisation (REM) of 150 steps, followed by 10 ps restrained molecular dynamics (RMD) in vacuo at a constant temperature of 27 °C, with a time step of 1 fs. Finally, a REM of 600 steps was carried out. Charges were not included during the calculations. A cutoff distance of 12 Å was used for nonbonded interactions. During the distance geometry calculations, hydrogen bonds in the helices were implemented by distance constraints between

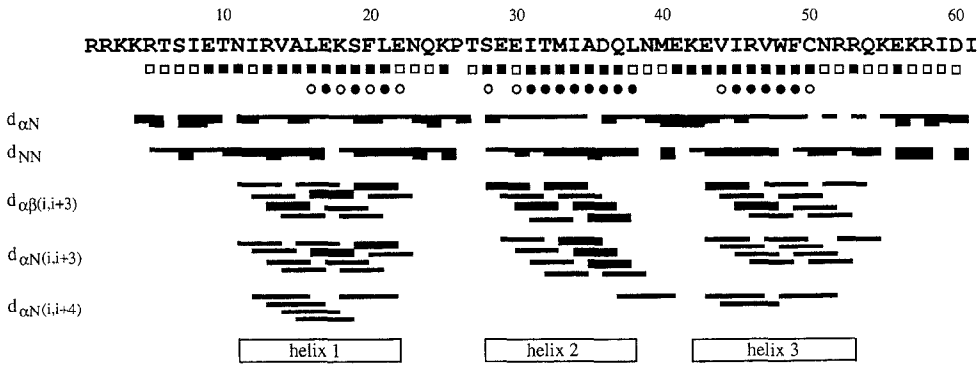

Fig. 3. Secondary structure of POU_{hd}. Sequential and medium-range contacts identified in the NMR spectra are indicated by bars. The thickness of the bars is a measure of the intensity of the signal. Open circles indicate slowly exchanging amide protons; filled circles indicate very slowly exchanging amide protons. Open squares denote $^3J_{NH_C\alpha}$ coupling constants < 6 Hz; closed squares denote $^3J_{NH_C\alpha}$ coupling constants > 6 Hz.

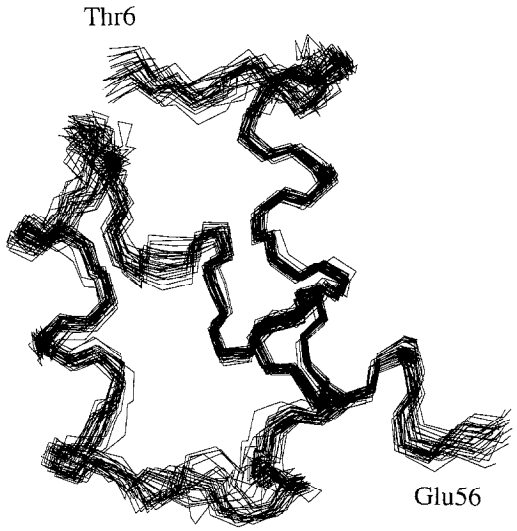

Fig. 4. Superposition of the 31 selected refined structures of POU_{hd}.

the carbonyl oxygen and the amide proton of the i+4th residue. In the RMD and REM calculations, these hydrogen bonds were omitted.

The structure refinement was an iterative procedure. With an initial set of constraints, a number of structures were calculated. Inspection of these structures helped in overcoming ambiguity in the assignment of more NOESY cross peaks. With the additional constraints a new set of structures was calculated, until a final set of 725 distance constraints was obtained. During the several rounds of structure calculation, a number of hydrophobic side chains in the core of the protein (Phe^{20}, Phe^{49}, Leu^{16}) were found to adopt a unique conformation. This allowed exclusion of the rather large pseudoatom corrections of 2.0 Å and 2.4 Å for aromatic rings and the methyl groups of the isopropyl moieties of leucine residues, respectively, resulting in tighter upper bounds for distance constraints involving these side chains.

The ^1H and ^{15}N resonance assignments (Supplementary Material) and the coordinates of the ensemble of structures have been deposited in the Brookhaven Protein Data Bank.

Results

The sequential assignment of the protein signals was performed according to standard procedures, as described by Wüthrich (1986). Initially, spin systems of the individual amino acids were assigned using COSY and TOCSY spectra, followed by identification of sequential connectivities using NOESY spectra. The sequential assignment was confirmed by analysis of the 3D NOESY-HSQC spectrum (Fig. 2), which solved almost all ambiguities present in the homonuclear spectra. Only the assignment of three N- and C-terminal residues was seriously hampered, due to overlap of the resonances, notably in the amide proton region. Over 95% of the ^1H resonances could be assigned for POU_{hd}.

A first inspection of short- and medium-range distance constraints reveals the location of secondary structure elements. The presence of (i,i+3) and (i,i+4) contacts, in combination with slowly exchanging amide protons and small $^3J_{H^NH^\alpha}$ coupling constants, is indicative of the presence of α-helices. Figure 3 demonstrates that there are three regions with these characteristics, which coincide well with the α-helices observed in Oct-3 POU_{hd} (Morita et al., 1994).

From the 50 calculated structures, a total of 31 were selected, having a low distance and dihedral constraint energy, and a low number of constraint violations. The ensemble of 31 structures is presented in Fig. 4. For the five N-terminal and the five C-terminal residues, no medium- and long-range constraints were determined. Therefore, the position of these residues is not well defined and, consequently, not shown. The rmsd of residues 6 to 54 with respect to the average structure is 0.59 Å for the backbone atoms and 1.30 Å for all atoms. For the best defined part, residues 10 to 24 and 28 to 54, these values are 0.49 Å for the backbone and 1.17 Å for all atoms. These residues have an angular order parameter for the ψ and ϕ angles larger than 0.9 (Fig. 5), and include all three helices and the loop connecting helices 2 and 3.

The 31 structures fit the experimental NMR data quite well. None of the structures display violations of the distance constraints larger than 0.55 Å. On average, each structure has less than 14 distance constraint violations larger than 0.2 Å (Fig. 6). The stereochemical quality of the structures was analysed with the software package PROCHECK (Morris et al., 1992). The results are summarised in Table 1. It can be seen that the 31 refined POU_{hd} structures have good stereochemical properties. Most of the structures have dihedral angles close to the ideal values, corresponding to classes 2 and 2 for the ϕ,ψ

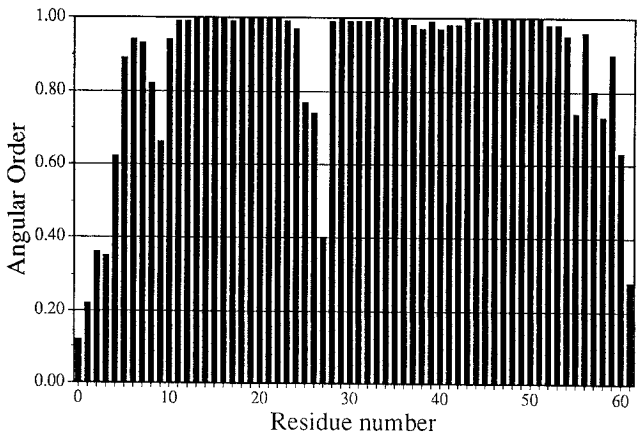

Fig. 5. Backbone angular order parameters for the 31 structures of POU_{hd}. The averages of the ϕ and ψ angle order parameters were taken.

angle distribution and χ^1 angle standard deviation, respectively. In five cases the classification is 1,3 and in seven cases it is 1,2. Only one structure has the classification 2,3. Moreover, the deviations of the bond lengths and angles from ideal geometry are only small, i.e., 0.02 Å for bond lengths and 1.6° for bond angles. The Ramachandran plot of all 31 structures (Fig. 7) displays only few residues in energetically unfavourable regions. These cases are mainly residues in the unstructured N- and C-termini.

The tertiary structure of POU_{hd} is composed of three α-helices (residues 11–22, 28–38 and 42–53) which have extensive interactions with each other. A representation of the average structure is given in Fig. 8. The inner part of the protein is formed by a number of aromatic rings and long aliphatic side chains, originating from Ile^{12}, Leu^{16}, Ile^{31}, Ile^{34}, Met^{40}, Ile^{45}, Trp^{48} and Phe^{49}, indicated in green. The residues in this hydrophobic pocket have a very well defined conformation. The heavy atoms in the side chains of the eight residues superimpose with an rmsd of 0.54 Å with respect to the average structure. Also, the three α-helices have a rather well-defined backbone, and the angles in the main chain differ only slightly from the ideal α-helical values, resulting in regular helices. The presence of slowly exchanging amide protons is in good agreement with the helical regions in the protein. Moreover, this correspondence extends to the occurrence of hydrogen bonds in the calculated structures. Only in the N-terminus of the second helix (residues Met^{33} and Thr^{32}) the hydrogen bonds occur in less than 30% of the structures, while the amide protons of these residues exhibit slow exchange behaviour. In the remaining part of the helical regions, the occurrence of hydrogen bonds in the calculated structures is over 60%.

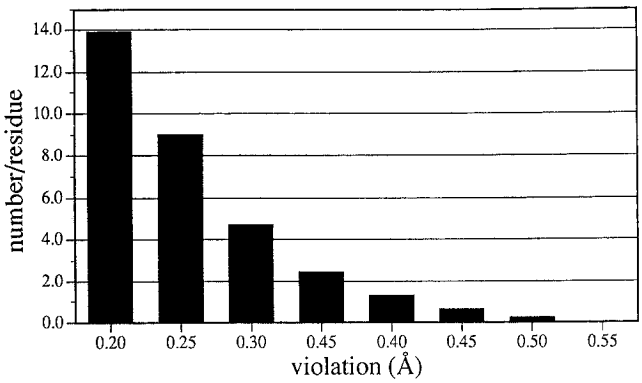

Fig. 6. The average number of distance constraint violations per structure, indicated by the height of the bars. No violations larger than 0.55 Å are observed in any of the 31 refined structures.

The two loops connecting the helices show more conformational freedom. Especially the loop connecting the first and second helix displays considerably more disorder, as reflected by the drastic drop in the angular order parameters of residues 25 to 27 (Fig. 5). Parts of the N- and C-termini fold back on the protein. Residues Thr^6 and Ile^8 have NOEs with residues Val^{44} and Trp^{48}, respectively, in the first half of the third helix. Moreover, residue Met^{40} in the second loop has NOEs with residues Arg^5, Thr^6 and Ile^8, while the calculated structures show a hydrogen bridge between the amide proton of Arg^5 and the backbone carbonyl of Glu^{41}. This region of the N-terminus, roughly ranging from Arg^5 to Glu^9, adopts an extended conformation and runs over the surface of the

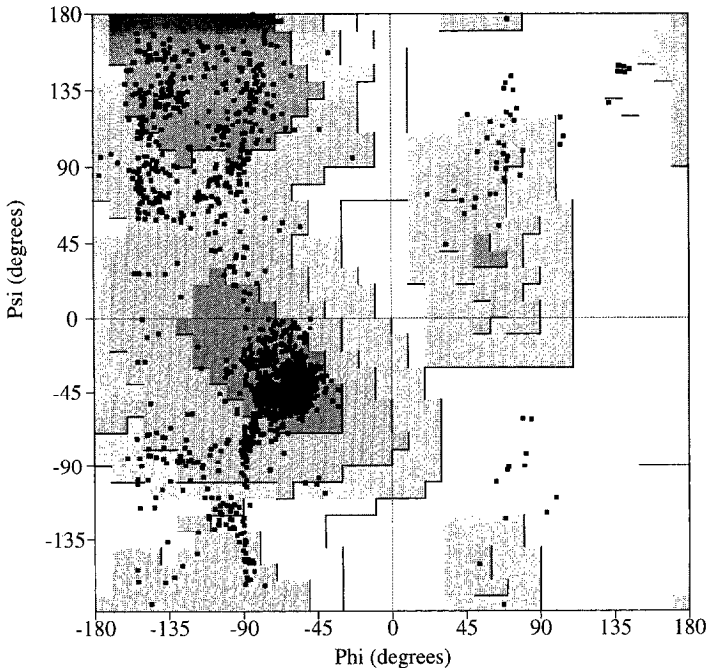

Fig. 7. Ramachandran plot of the 31 refined structures of POU_{hd}. Favourable regions are shaded.

TABLE 1
STEREOCHEMICAL QUALITY OF THE REFINED POU_{hd} STRUCTURES

Parameter	Value in POU_{hd} (°)	Ideal value[a] (°)
χ^1 g(−)	64.0 ± 16.0	64.1 ± 15.7
χ^1 t	193.4 ± 20.5	183.6 ± 16.8
χ^1 g(+)	−66.9 ± 17.0	−66.7 ± 15.0
χ^2	178.0 ± 14.8	177.4 ± 18.5
ϕ helix	−62.2 ± 11.9	−65.3 ± 11.9
ψ helix	−41.8 ± 13.9	−39.4 ± 11.3
ω	176.1 ± 6.5	180.0 ± 5.8
Chirality C^α	31.8 ± 3.9	33.9 ± 3.5

[a] Ideal values according to PROCHECK (Morris et al., 1992).

second loop and the N-terminus of the third helix. In the C-terminus, the aliphatic parts of residues Arg^{52} and Glu^{56} approach the ring of Phe^{20} from two sides, thus fixing part of the C-terminus to the first helix. This results in a rather globular structure for the protein, where the three helices enclose a hydrophobic pocket, while parts of the N- and C-termini are also part of the folded structure.

Discussion

Comparison of POU_{hd} with classical homeodomains

It has been shown recently that the solution structures of the *Antennapedia* and *fushi tarazu* homeodomains are very similar (Qian et al., 1994). This could be expected, since in the well-structured part of the two proteins, from residue 4 to 58, the sequences differ by as few as six residues, and in four cases these are functionally homologous residues. A superposition of the backbone atoms of these parts of the average structures of *Antennapedia* and *ftz* results in an rmsd of 0.77 Å. The structure, as well as the precision of the structure determination, are very similar. Large deviations occur only at the C-terminal part of the last helix. Whereas *ftz* has a third helix, ranging from residue 42 to 52, the helix of *Antennapedia* extends from residue 42 up to 59, with a kink around residue 53. Clearly, POU_{hd} is quite distinct from the classical homeodomains with regard to sequence homology; for instance, POU_{hd} has a sequence identity with *Antennapedia* and *ftz* of only 33%. It is therefore interesting to notice that, with a low amino acid conservation, the overall fold of POU_{hd} is still very similar to that of *ftz* and *Antennapedia*. This can be visualised in the superposition of the backbone atoms of POU_{hd} on some classical homeodomains in Fig. 9. As can be seen in Table 2, the superposition of the C^α atoms of the four homeodomains demonstrates that, although POU_{hd} deviates the most of the four homeodomains, the similarity of the global fold is quite striking.

The positions and lengths of the helices in POU_{hd} coincide almost exactly with the helices in *ftz*. Also, the last helix ends around residue 53, in contrast to *Antennapedia*. It has been argued that the extra helical region in the C-terminus of *Antennapedia* is stabilised by the anchoring function of Trp^{56} to the hydrophobic core (Qian et al., 1994). A tryptophan at this position is unique to *Antennapedia*. The *ftz* homeodomain has a serine at position 56 which does not interact with the core, since its side chain is incapable of making extensive hydrophobic interactions. On the other hand, POU_{hd} has a glutamic acid at the corresponding position. As mentioned above, the structures of POU_{hd} presented in this study show that this Glu^{56}, together with Arg^{52}, forms a small hydrophobic pocket with the ring of Phe^{20}, adjacent to the inner core of the protein. Thus, in the case of POU_{hd}, anchoring of the residue at position 56 against the core does not result in the formation of a stable α-helix. This is further evidenced by the observation of large $^3J_{HNH\alpha}$ values and the absence of slowly exchanging amide protons and of (i,i+3) and (i,i+4) contacts for the residues in the C-terminus. In fact, the observation of medium and strong $d_{\alpha N}$ contacts, in combination with medium and strong d_{NN} contacts in the C-terminus (Fig. 3), may indicate the presence of multiple conformations, possibly including a nascent helix in this region (Waltho et al., 1993). If both the α-helical region and the β-sheet region of the (ϕ,ψ) space are populated, be it averaged over time or over different molecules, the $d_{\alpha N}$ and d_{NN} contact intensities will be high. This is indeed the case for residues 56 to 61 of POU_{hd}, so this part of the protein does not have a well-defined structure. In summary, we find no evidence that the character of the C-terminus of POU_{hd} is α-helical, in correspondence with the *ftz* homeodomain.

TABLE 2
SUPERPOSITION[a] OF POU_{hd} DOMAINS ON RESIDUES 10–54 OF CLASSICAL HOMEODOMAINS

	$POU_{hd}N$	$POU_{hd}X$	Ant	ftz	en	Mat
$POU_{hd}N$	0.00					
$POU_{hd}X$	1.11	0.00				
Ant	1.55	1.02	0.00			
ftz	1.83	1.26	0.93	0.00		
en	1.49	0.81	0.81	0.95	0.00	
Mat	1.99	1.63	1.56	1.69	1.28	0.00

[a] Superpositions are performed on the C^α atoms of the free POU_{hd} ($POU_{hd}N$), the POU_{hd} of the X-ray structure of the POU domain in complex with DNA ($POU_{hd}X$), *Antennapedia* (Ant), *fushi tarazu* (ftz), *engrailed* (en) and *Mat α2* (Mat).

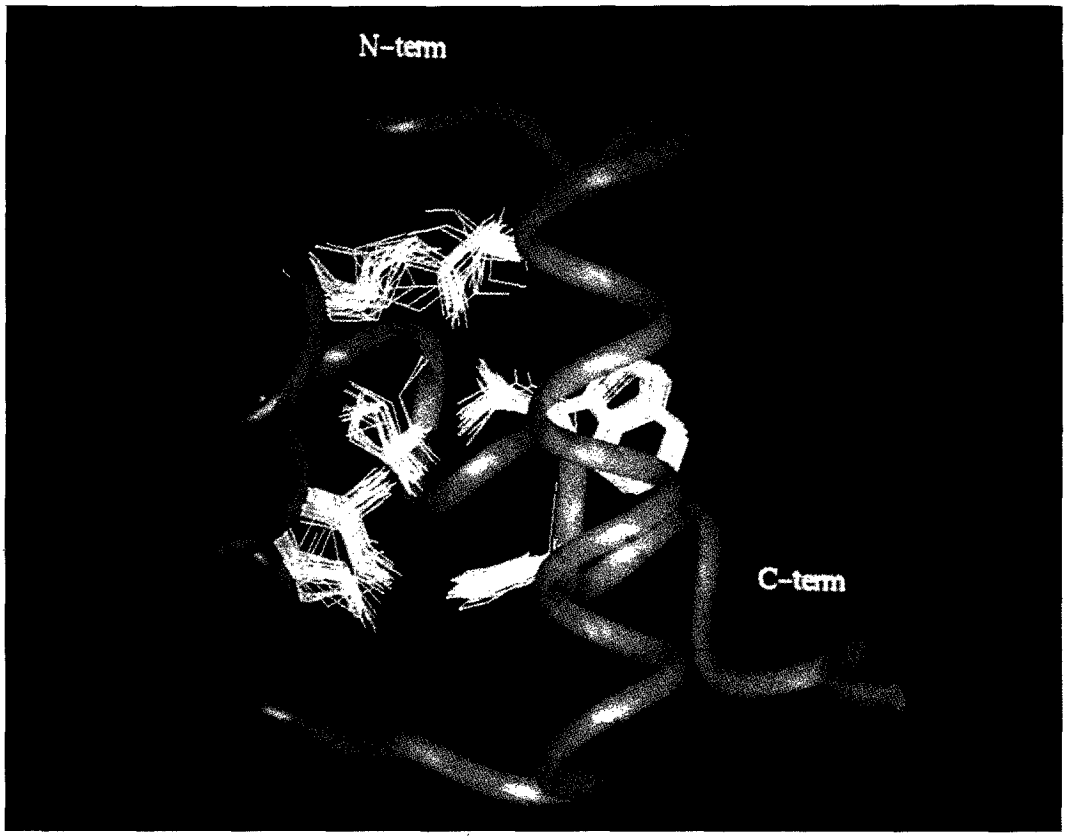

Fig. 8. Schematic drawing of the average structure of POU_{hd}. The heavy atoms of the side chains of the hydrophobic core, comprising residues 12, 16, 31, 34, 40, 45, 48 and 49, are indicated in green. The N-terminus and the C-terminus are labelled.

Comparison of free POU_{hd} with DNA-bound POU_{hd}

The structure of POU_{hd} presented in this study is that of the free protein. In order to check if the structure changes upon binding to its target DNA sequence, it is interesting to compare it with the crystal structure of the POU domain (Klemm et al., 1994). This structure comprises both POU_s and POU_{hd}, connected with the linker, bound to a 14-base pair DNA sequence containing the octamer motif. Although the resolution of the crystal structure is not very high, overall differences between the crystal structure and the NMR structure can be identified. A superposition of the C^{α} atoms of residues 10 to 54 of POU_{hd} results in an rmsd value of 1.11 Å, indicating a similar overall fold for the two structures. Figure 10 reveals, however, that the length of the recognition helix is extended as far as residue 60 in the structure of the complex, while in the free protein it runs up to residue 53. It was already mentioned that the structure of the free protein in the C-terminus is not totally disordered. Long-range contacts up to residue Glu^{56} are found. The dynamic behaviour of the C-terminus can be monitored to some extent by inspection of the T_2 relaxation times of the protons. A fast way to measure these is by determining the line widths of the signals, e.g. the amide protons. Furthermore, protons in flexible regions in the protein will display more cross peaks in TOCSY spectra. In the free POU_{hd}, residues 2–5 and 57–60 all have small line widths and many cross peaks in the TOCSY spectrum. Therefore they might be flexible, which is consistent with the structure of the free protein. Also, residues 54–56 have a tendency for small line widths. On the other hand, the residues located in the helices tend to exhibit rather large line widths. Hence, we can conclude that the C-terminal part of the free protein, starting at residue 54, is not completely disordered, but does not have a regular secondary structure either. On complex formation, the C-terminus adopts a helical conformation, to continue the third helix. A similar effect has also been observed in the heterodimers of the homeodomain proteins α2 and a1. The free α2 protein has three helices, but on dimerization a fourth eight-residue helix is induced at the C-terminus of α2 (Philips et al., 1994). Furthermore, this behaviour is reminiscent of the complex formation of the *Antennapedia* homeodomain. The free protein has an elongated recognition helix, running as far as residue 59. On complexing the DNA, this elongation, or fourth helix, is stabilised due to formation of DNA contacts, as monitored by amide exchange rates (Qian et al., 1993).

The actual mechanism for the induction of secondary structure in POU_{hd} is not totally clear. It can be argued that POU_{hd}–DNA contacts are responsible for the stabilising effect. The crystal structure shows that the side chains of Lys^{55} and Lys^{57} make phosphate contacts with the DNA, although the proper orientation of the side

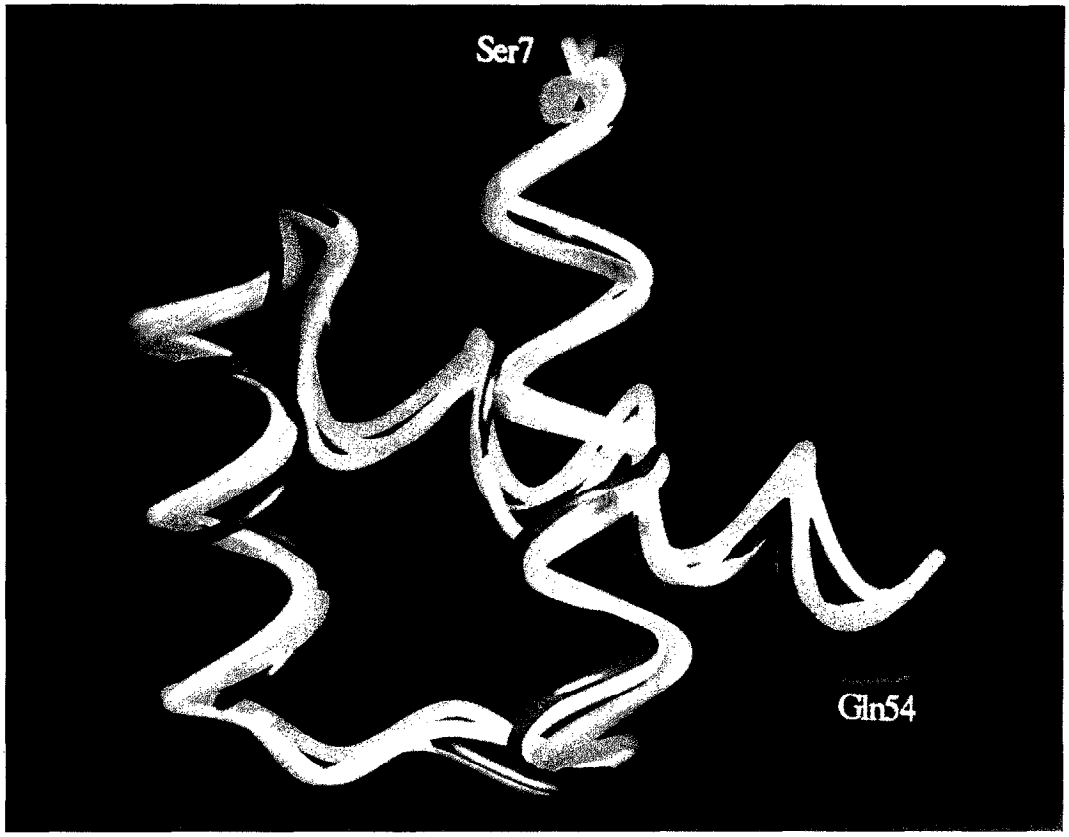

Fig. 9. Superposition of the backbones of four homeodomain proteins. The structures of *Antennapedia*, *engrailed* and *fushi tarazu* are in blue; POU_{hd} is shown in red.

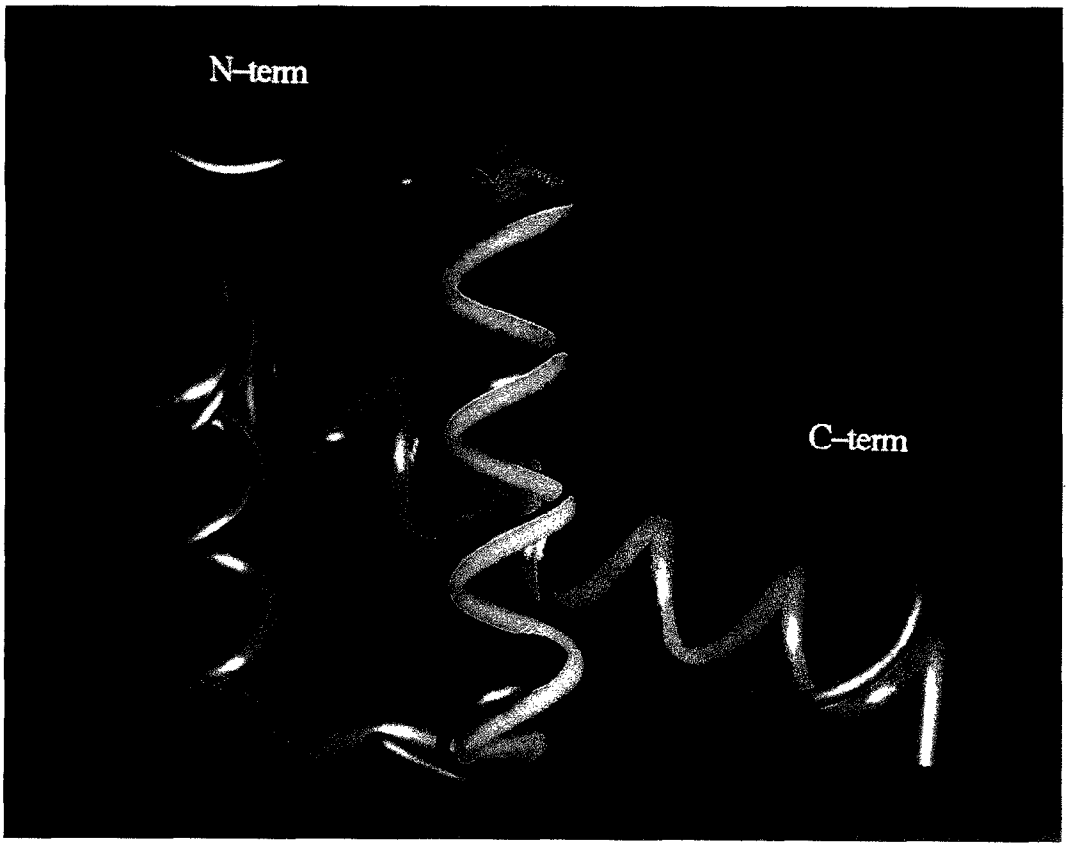

Fig. 10. Superposition of the average solution structure of POU_{hd} as determined by NMR (red) and POU_{hd} of the crystal structure of the POU-domain, complexed with its cognate DNA sequence (green). N- and C-termini are labelled. Note the difference of the C-terminal six amino acids.

chains is not well defined in this region. On the other hand, the presence of POU_s may affect the structure of the recognition helix of POU_{hd}. However, residues Asn^{51} to Lys^{57} have identical chemical shifts in the NMR spectra of free POU_{hd} and the free intact POU domain, consisting of POU_s, the linker and POU_{hd} (M. Cox, unpublished results). This demonstrates that the structure of the C-terminus of POU_{hd} is independent of the presence of POU_s. This is in agreement with the absence of interactions between POU_{hd} and POU_s in the crystal structure of the complex.

The second difference between the free and the DNA-bound protein is the orientation of the N-terminus. In the free protein, the residues at positions 5, 6 and 8 are all in close contact with the third helix, which positions the N-terminus in this region firmly at the surface of the protein. The preceding residues, 1–4, are disordered. In the complex, the N-terminus is well defined. This can be rationalised by the fact that the side chains of residues 3, 5 and 6 and the amide proton of residue 6 are involved in phosphate backbone contacts with the DNA, thus orienting the N-terminus into the minor groove of the DNA.

Conclusions

The results presented in this paper can be summarised as follows. Residues 5 to 54 of the human Oct-1 POU_{hd} adopt a globular fold, which is very similar to the structures of the classical homeodomains *Antennapedia*, *engrailed*, *ftz* and *Mat* α*2*, although the sequence homology between POU_{hd} and the other homeodomains is not as high as among the classical homeodomains themselves. The structure of the free protein is basically similar to the structure of the DNA-bound protein, but with two differences. The N-terminus of POU_{hd} in the complex interacts with the phosphate backbone of the DNA. This causes the N-terminus to adopt a well-defined orientation. The structure of the free protein exhibits disorder in the first four residues and becomes structured only at position 5. Furthermore, evidence has been provided for the absence of an α-helical conformation in the C-terminus. Only upon binding to DNA the recognition helix is extended from residue 54 to 60.

Acknowledgements

We thank M.J.J. Strating for purifying NMR samples. The investigations were supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organisation for Scientific Research (NWO).

References

Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.*, **61**, 306–320.

Billeter, M., Qian, Y., Otting, G., Müller, M., Gehring, W.J. and Wüthrich, K. (1990) *J. Mol. Biol.*, **214**, 183–197.

Griesinger, C., Otting, G., Wüthrich, K. and Ernst, R.R. (1988) *J. Am. Chem. Soc.*, **110**, 7870–7872.

Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990) *Cell*, **63**, 579–590.

Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. and Pabo, C.O. (1994) *Cell*, **77**, 21–32.

Koning, M.M.G., Boelens, R. and Kaptein, R. (1990) *J. Magn. Reson.*, **90**, 111–123.

Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1989) *J. Am. Chem. Soc.*, **111**, 1515–1517.

McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A. and Gehring, W.J. (1984) *Cell*, **37**, 403–408.

Morita, E.H., Shirakawa, M., Hayashi, F., Imagawa, M. and Kyogoku, Y. (1994) *FEBS Lett.*, **321**, 107–110.

Morris, A.L., McArthur, M.W., Hutchinson, E.G. and Thornton, J.M. (1992) *Protein Struct. Funct. Genet.*, **12**, 345–364.

Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affolter, M., Gehring, W.J. and Wüthrich, K. (1990) *EMBO J.*, **9**, 3085–3092.

Philips, C.L., Stark, M.R., Johnson, A.D. and Dahlquist, F.W. (1994) *Biochemistry*, **33**, 9294–9302.

Qian, Y.Q., Furukubo-Tokunaga, K., Resendez-Perez, D., Müller, M., Gehring, W.J. and Wüthrich, K. (1994) *J. Mol. Biol.*, **238**, 333–345.

Qian, Y.Q., Otting, G., Billeter, M., Müller, M., Gehring, W. and Wüthrich, K. (1993) *J. Mol. Biol.*, **234**, 1070–1083.

Scott, M.P. and Weiner, A.J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4115–4119.

Shaka, A.J., Barker, P.B. and Freeman, R.J. (1985) *J. Magn. Reson.*, **64**, 547–552.

Sivaraja, M., Botfield, M.C., Müller, M., Jancso, A. and Weiss, M.A. (1994) *Biochemistry*, **33**, 9845–9855.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorf, J.W. (1990) *Methods Enzymol.*, **185**, 60–89.

Verrijzer, C.P., Alkema, M.J., Van Weperen, W.W., Van Leeuwen, H.C., Strating, M.J.J. and Van der Vliet, P.C. (1992) *EMBO J.*, **11**, 4993–5003.

Verrijzer, C.P. and Van der Vliet, P.C. (1993) *Biochim. Biophys. Acta*, **1173**, 1–21.

Waltho, J.P., Feher, V.A., Merutka, G., Dyson, H.J. and Wright, P.E. (1993) *Biochemistry*, **32**, 6337–6347.

Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D. and Pabo, C.O. (1991) *Cell*, **67**, 517–528.

Wright, P.E. (1994) *Curr. Biol.*, **4**, 22–27.

Wüthrich, K., Billeter, M. and Braun, W. (1983) *J. Mol. Biol.*, **169**, 949–961.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.

Yu, V.C., Delsert, C., Andersen, B., Holloway, J.M., Devary, O.V., Näär, A.M., Kim, S.Y., Boutin, J.-M., Glass, C.K. and Rosenfeld, M.G. (1991) *Cell*, **67**, 1251–1266.