The EMBO Journal vol.12 no.5 pp.1797-1803, 1993

The three-dimensional NMR-solution structure of the polypeptide fragment 195-286 of the LFB1/HNF1 transcription factor from rat liver comprises a non-classical homeodomain

Barbara Leiting[1], Raffaele De Francesco[2], Licia Tomei[2], Riccardo Cortese[2], Gottfried Otting[1] and Kurt Wüthrich[1]

[1]Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland, and [2]IRBM, Istituto di Ricerche di Biologia Moleculare P.Angeletti, Via Pontina Km 30 600, 00040 Pomezia, Roma, Italy

Communicated by R.Cortese

The three-dimensional backbone fold of a polypeptide fragment from the rat LFB1/HNF1 transcription factor was determined by nuclear magnetic resonance (NMR) spectroscopy in solution. This fragment contains an amino acid sequence that is ~22% homologous to the well known homeodomains, but which contains 81 amino acid residues as compared with 60 residues in 'typical' homeodomains. For the present studies we used a recombinant 99 amino acid polypeptide containing this sequence in positions 10-90, which was uniformly labelled with ^{15}N and also doubly labelled with ^{15}N and ^{13}C. The NMR structure of this polypeptide contains three α-helices comprising the residues 18-29, 36-50 and 71-84, a loop formed by residues 30-35, and a long stretch of non-regular secondary structure linking the second and third helices. The relative location and orientation of the helices is very similar to that in the *Antennapedia* (*Antp*) homeodomain structure, despite the fact that helix II is elongated by about one turn. This confirms a recently advanced hypothesis based on sequence comparisons that this polypeptide segment of LFB1/HNF1 might represent a homeodomain-like structural element. The helix-turn-helix motif, which has been shown to comprise the DNA recognition helix in the *Antp* homeodomain, can readily be recognized in the LFB1/HNF1 homeodomain, in spite of an extensive modification of the primary structure. The two residues of the tight turn in the *Antp* homeodomain are replaced by a 23 residue linker region between the two helices in LFB1/HNF1, which bulges out from the rest of the molecule and thus enables the formation of a non-classical helix-turn-helix motif.

Key words: helix-turn-helix motif/LFB1/HNF1 homeodomain/nuclear magnetic resonance/protein structure

Introduction

The classical homeodomain consists of a 60 amino acid residue polypeptide segment that has been found in many proteins involved in transcriptional regulation in a variety of species ranging from yeast to man (Gehring, 1987). Functional studies showed that the homeodomain is responsible for sequence-specific DNA binding (Müller *et al.*, 1988; Scott *et al.*, 1989; Hayashi and Scott, 1990). The three-dimensional nuclear magnetic resonance (NMR) structure of the *Antennapedia* (*Antp*) homeodomain in aqueous solution (Qian *et al.*, 1989; Billeter *et al.*, 1990) and subsequent crystal structure determinations of highly homologous homeodomains in complexes with DNA (Kissinger *et al.*, 1990; Wolberger *et al.*, 1991) all showed protein folds with three α-helical regions, of which the second and third helices form a helix-turn-helix motif that is nearly identical to that described previously for a variety of prokaryotic repressor proteins (Otting *et al.*, 1988). In the complexes with DNA, the second helix of the helix-turn-helix motif is located in the major groove of the DNA (Kissinger *et al.*, 1990; Otting *et al.*, 1990; Wolberger *et al.*, 1991) and may therefore be referred to as the 'recognition helix'. To investigate further whether the typical homeodomain molecular architecture is preserved also in domains with more extensively diverging amino acid sequences, we have started an NMR structure determination of a polypeptide segment from the LFB1/HNF1 transcription factor from rat liver for which alignments with ~22% sequence homology with the *Antp* homeodomain have been proposed (see below).

The LFB1/HNF1 protein is a major component of the hepatocyte-specific transcription activator apparatus (De Simone and Cortese, 1991; Mendel and Crabtree, 1991) and consists of 628 residues. The native protein binds as a homodimer to a palindromic DNA recognition sequence. Three distinct protein domains are involved in DNA binding, of which the third domain shows significant homology to the 'classical' homeodomain sequence (Frain *et al.*, 1989; Nicosia *et al.*, 1990; Tomei *et al.*, 1992). In particular, the four residues Trp48, Phe49, Asn51 and Arg53 of the *Antp* homeodomain, which are strictly conserved among all presently known homeodomain sequences (Wüthrich and Gehring, 1992), are present in LFB1/HNF1, and six out of 14 additional highly conserved residues are also retained. Furthermore, sequence alignments suggest an insertion of 21 residues (Finney, 1990), which increases the length of the presumed LFB1/HNF1 homeodomain sequence to 81 residues as compared with 60 residues for the *Antp* homeodomain. Recently reported DNA binding studies (Tomei *et al.*, 1992) showed that a polypeptide comprising residues 195-286 of LFB1/HNF1 binds as a monomer to a half site of the palindromic LFB1/HNF1 recognition sequence with a binding constant of $\sim 10^{-9}$ M, which is comparable with the stability of the complex formed between the *Antp* homeodomain and the BS2 DNA recognition site (Müller *et al.*, 1988; Affolter *et al.*, 1990). To investigate structural relations between this region of LFB1/HNF1 and the classical homeodomains, we prepared a 99 amino acid polypeptide containing the residues 195-286 of the rat LFB1/HNF1 protein and four N-terminal and three C-terminal residues from the overexpression system used. This paper describes the determination of the secondary structure and the global backbone fold of this polypeptide by NMR,

and the three-dimensional (3D) backbone fold is compared with that of the *Antp* homeodomain (Qian *et al.*, 1989; Billeter *et al.*, 1990).

Results and discussion

Initial inspection of 1D and 2D ^{1}H NMR spectra of the 99 amino acid residue polypeptide containing the presumed LFB1/HNF1 homeodomain (see Introduction) showed that the individual resonance lines are rather poorly separated when compared with other proteins of similar size. On the basis of the structural information presented in this paper, it is clear that this limited spectral resolution is mainly due to two factors (Wüthrich, 1986): (i) the protein contains α-helices as the only type of regular secondary structure; (ii) the chain ends are flexibly disordered, with concommitant small dispersion of the ^{1}H NMR chemical shifts (Wüthrich, 1976). The practical consequence was that we prepared uniformly ^{15}N-labelled and ^{15}N^{13}C doubly-labelled LFB1/HNF1 to obtain improved spectral resolution for the resonance assignments and the collection of conformational constraints.

Sequence-specific resonance assignments

Figure 1 affords a survey of the sequential NOE connectivities $d_{\alpha N}$, d_{NN} and $d_{\beta N}$ (Billeter *et al.*, 1982; Wüthrich, 1986) that were identified in the process of obtaining sequence-specific ^{1}H NMR assignments for the LFB1/HNF1 homeodomain. The sequential NOE connectivities were obtained primarily from a (3D) ^{15}N-correlated ^{1}H,^{1}H-NOESY spectrum recorded with the uniformly ^{15}N-labelled protein (Fesik and Zuiderweg, 1988; Marion *et al.*, 1989; Messerle *et al.*, 1989). As an illustration, Figure 2 shows the data used for the assignments from residues 36–54, which were subsequently found to comprise the helix II. The 3D spectrum used can be visualized as consisting of 64 $\omega_1(^1\text{H})-\omega_3(^1\text{H})$ planes separated along $\omega_2(^{15}\text{N})$ by the different ^{15}N chemical shifts of the residues observed in the individual planes. Each $\omega_1(^1\text{H})-\omega_3(^1\text{H})$ plane represents a 2D ^{1}H,^{1}-NOESY spectrum. To obtain a composite strip plot as shown in Figure 2, strips extending over the entire spectrum along $\omega_1(^1\text{H})$ are selected in all 64 planes, with a width of 0.15 p.p.m. along $\omega_3(^1\text{H})$ and centred about the amide proton chemical shifts of the individual residues. Because of the sequence-dependent distribution of the ^{15}N chemical shifts (e.g. Neri *et al.*, 1992), strips belonging to sequentially neighbouring residues may be found in a single plane, or in any possible combination of two out of the total of 64 $\omega_1(^1\text{H})-\omega_3(^1\text{H})$ planes. Therefore, to identify the sequential NOE connectivities (Billeter *et al.*, 1982; Wagner and Wüthrich, 1982; Wüthrich, 1986), all possible combinations of any two strips from all 64 $\omega_1(^1\text{H})-\omega_3(^1\text{H})$ planes must be evaluated (a process that has been largely automated in our laboratory). The presentation of Figure 2 is then obtained by arranging the strips in the order of the amino acid sequence in a single, composite $\omega_1(^1\text{H})-\omega_3(^1\text{H})$ plane (Driscoll *et al.*, 1990; Wüthrich *et al.*, 1991). This composite strip plot then further supports an efficient search for additional sequential NOEs and medium-range NOE connectivities (Figure 2).

With the aforementioned ^{15}N-correlated ^{1}H,^{1}H-NOESY experiment with the uniformly ^{15}N-labelled protein, at least one sequential NOE connectivity was identified for all except

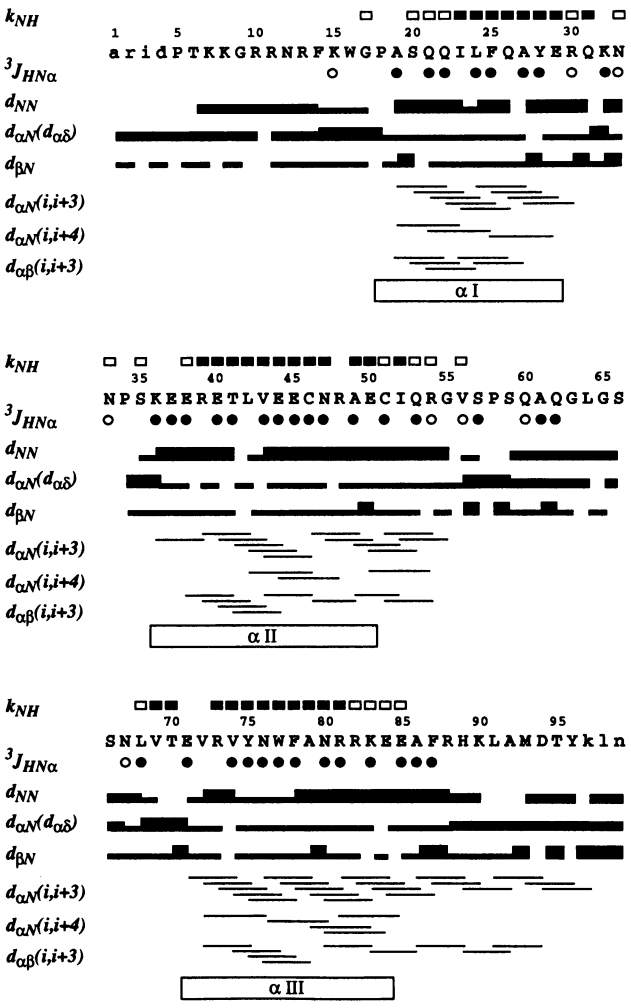

Fig. 1. Amino acid sequence and secondary structure of the 99 amino acid polypeptide containing the residues 195–286 of the rat LFB1/HNF1 protein in positions 5–96, and survey of sequential and medium-range NOE connectivities, amide proton exchange data and $^3J_{HN\alpha}$ coupling constants. Upper case letters represent the residues corresponding to the amino acid sequence of LFB1/HNF1, and lower case letters of both chain ends identify the peptide segments that originate from the overexpression system used and are not part of the LFB1/HNF1 sequence. Filled and open boxes above the sequence identify the residues with exchange rate constants $<10^{-3}$/min, and between 10^{-2} and 10^{-3}/min, respectively, in D_2O at pD 3.6 and 20°C. Below the sequence, filled circles identify residues with $^3J_{HN\alpha}$ coupling constants <6.0 Hz, and open circles those with $^3J_{HN\alpha}$ >8.0 Hz. The following three rows present the sequential connectivities d_{NN}, $d_{\alpha N}$ and $d_{\beta N}$, where the width of the bar reflects the NOE intensity. Medium-range NOE connectivities $d_{\alpha N(i,i+3)}$, $d_{\alpha N(i,i+4)}$ and $d_{\alpha\beta(i,i+3)}$ are represented by lines starting and ending at the positions of the interacting residues. At the bottom, the locations of three α-helices are indicated, which were identified from the data in this figure and the structure calculations with the program DIANA (Figure 3).

two pairs of neighbouring residues; the exceptions were Arg30–Glu31 and Val69–Thr70. Subsequently, as a by-product of the collection of NOE distance constraints as input for the structure calculations, these sequential connectivities were confirmed and additional sequential NOEs identified using a 3D ^{13}C-correlated ^{1}H,^{1}H-NOESY spectrum. Figure 1 shows that all sequential connectivities could thus be established, and that all three sequential NOEs $d_{\alpha N}$, d_{NN} and $d_{\beta N}$ were identified for the majority of neighbouring residue pairs.

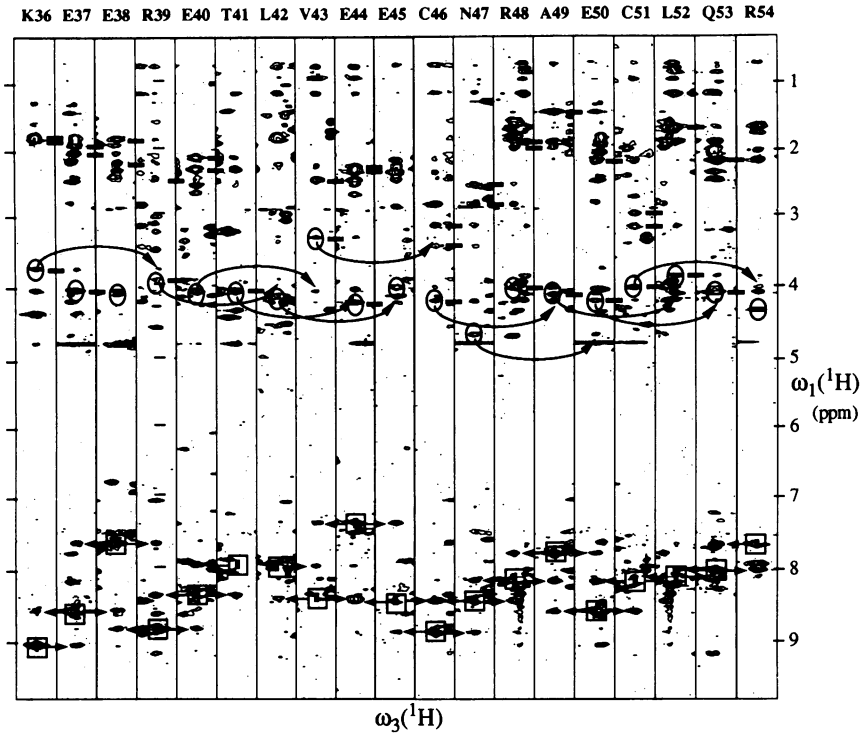

Fig. 2. Composite plot of strips taken from a 3D ^{15}N-correlated ^{1}H,^{1}H-NOESY spectrum of the LFB1/HNF1 homeodomain, which shows sequential connectivities in the polypeptide segment from Lys36 to Arg54, which comprises the entire helix II (protein concentration 4.0 mM, solvent 90% H_2O/10% D_2O, pH = 3.6, T = 20°C, mixing time = 80 ms, ^{1}H frequency = 600 MHz). The individual strips with a width of 0.15 p.p.m. along ω_3 were taken from $\omega_1(^1H)-\omega_3(^1H)$ planes at different $\omega_2(^{15}N)$ chemical shifts and centred about the amide proton chemical shift of the residue indicated at the top. The strips are ordered according to the amino acid sequence. Squares and circles identify the amide proton diagonal peaks and their intraresidual NH−C$^{\alpha}$H cross peaks, respectively. Straight arrows and curved arrows identify the sequential d_{NN} connectivities and the medium-range $d_{\alpha N}$ connectivities, respectively. Horizontal bars are drawn at the ω_1 chemical shifts of the sequential $d_{\alpha N}$ and $d_{\beta N}$ connectivities (see also Figure 1).

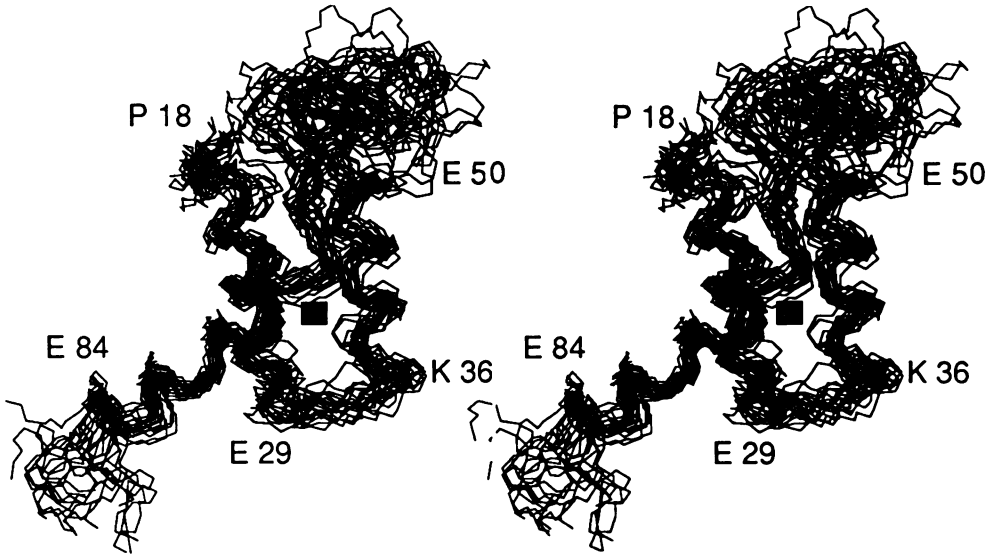

Fig. 3. Stereo view of the 20 energy-minimized DIANA conformers of the LFB1/HNF1 homeodomain used to represent the NMR solution structure of this protein. The residues 19−50 and 68−84 were used for a best fit superposition, and the polypeptide backbone heavy atoms of the residues 14−89 are displayed. The first and last residues of each helix are indicated with the one-letter amino acid symbols and the sequence positions, except that the location of E71 at the start of helix III is identified by a black square. A schematic drawing of the molecule in a similar orientation is shown in Figure 5A.

Although no amino acid side chain conformations are discussed in this paper, the calculation of the polypeptide backbone fold (Figure 3) was based on ^{1}H resonance assignments for a large majority of all side chains, which enabled the collection of numerous medium-range and long-rang NOE distance constraints. Side chain assignments were obtained using standard spin system assignment strategies with 2D ^{1}H,^{1}H-NMR spectra (Wider *et al.*, 1982;

Wüthrich, 1986), as well as higher-dimensional experiments recorded with the $^{15}N,^{13}C$ doubly-labelled protein, in particular 3D HCCH-TOCSY in D_2O and 3D ^{13}C-correlated $^{1}H,^{1}H$-NOESY in H_2O (see Materials and methods). The side chain resonance assignments were further supported with experiments using a biosynthetically directed fractionally ^{13}C-labelled sample of the LFB1/HNF1 polypeptide (Senn *et al.*, 1989; Szyperski *et al.*, 1992; Wüthrich *et al.*, 1992), which enabled the straightforward stereospecific assignment of 18 out of the total of 22 methyl groups of valine and leucine. Overall 88 side chains were completely assigned, and partial assignments were obtained for 11 side chains. A list of ^{1}H, ^{13}C and ^{15}N chemical shifts will be published once the determination of the LFB1/HNF1 homeodomain structure has been completed. This data will then also be deposited in the relevant data banks.

The secondary structure of the LFB1/HNF1 polypeptide 195–286

A survey of the sequential and medium-range NOE connectivities observed for the LFB1/HNF1 homeodomain, together with the $^{3}J_{HN\alpha}$ coupling constants and the residue positions with slowly exchanging amide protons (Figure 1) resulted in the identification of three helical regions. Four criteria support the presence of these α-helices: (i) $^{3}J_{HN\alpha}$coupling constants <6.0 Hz (Pardi *et al.*, 1984); (ii) strong sequential d_{NN} connectivities (Billeter *et al.*, 1982); (iii) a significant number of medium-range connectivities $d_{\alpha N(i,i+3)}$,$d_{\alpha N(i,i+4)}$ and $d_{\alpha\beta(i,i+3)}$ characteristic of α-helices (Wüthrich *et al.*, 1984); (iv) slow amide proton exchange rates for most residues in the helical polypeptide segments (Wüthrich, 1986). As usual, the start and the end of the individual helices could not be precisely defined by the empirical pattern recognition approach (Wüthrich *et al.*, 1984). Therefore, the ends of the helices determined from the 3D structure calculations are shown in Figure 1. Turn regions and other irregular secondary structures are indicated for the segments 15–17, 30–35, 54–58 and 68–70, which show no uniform patterns of either small or large $^{3}J_{HN\alpha}$ coupling constants, or continuous trains of either weak d_{NN} and strong $d_{\alpha N}$ connectivities, or strong d_{NN} and weak $d_{\alpha N}$ connectivities. The segments 1–14, 59–67 and 88–99 show simultaneously d_{NN} and $d_{\alpha N}$ connectivities of medium intensity and only few medium-range NOE connectivities, which indicates a flexibly disordered conformation (Wüthrich, 1986).

The three-dimensional polypeptide backbone fold of the LFB1/HNF1 polypeptide 195–286

In addition to the sequential and medium-range NOE distance constraints with backbone hydrogen atoms (Figure 1), the aforementioned ^{1}H NMR assignments for the majority of the amino acid side chains enabled the identification of numerous further local and long-range NOE constraints. An as yet incomplete input of ~800 NOE constraints for a structure calculation was thus obtained, which is beset with some ambiguities relative to possible spin diffusion effects because of the rather long NOESY mixing times used (Wüthrich, 1986) (see Materials and methods). Although the structure calculations with the distance geometry program DIANA (Güntert *et al.*, 1991a; for details see Materials and methods) included the complete polypeptide chain with amino acid side chains, we presently discuss only the backbone fold and defer discussion of the complete 3D structure until after further refinement of the NMR structure.

Figure 3 shows a superposition of the backbone heavy atoms of residues 14–89 of the 20 energy-refined DIANA conformers selected to represent the solution structure of the LFB1/HNF1 polypeptide. The average of the pairwise RMSD values of these 20 conformers relative to the mean coordinates calculated for the backbone heavy atoms of residues 19–50 and 68–84 is 1.24 Å. The well defined parts of the structure include three helices comprising the residues 18–29, 36–50 and 71–84. There is an antiparallel arrangement of the helices I and II, and helix III is located approximately perpendicular to the other two helices. The residues 51–53 and 85–87 at the C-termini of the helices II and III have some helical character (Figure 1) but they are less well defined by the NMR data, which is probably due to increased structural flexibility.

Comparison of the LFB1/HNF1 homeodomain with the Antp homeodomain structure

A sequence alignment of the LFB1/HNF1 homeodomain and the *Antp* homeodomain indicating the degree of sequence homology, the location of helical secondary structures, and an outline of the segments with similar 3D structures is shown in Figure 4. Superimposing residues 16–46 and 70–84 of the LFB1/HNF1 homeodomain on the corresponding residues 8–38 and 41–55 of the *Antp* homeodomain, the average of the pairwise RMSD values between the 20 conformers used to represent the LFB1/HNF1 homeodomain structure and those for the *Antp* homeodomain is 1.51 Å for the backbone heavy atoms. These residues include nearly all of the well defined parts of the *Antp* homeodomain structure (Qian *et al.*, 1989). The main differences between the two homeodomain structures clearly result from having a fragment of 23 amino acids (47–69) in the LFB1/HNF1 homeodomain in the place of the two residues Cys39 and Leu40 in the *Antp* homeodomain. In Figure 5, which shows schematic presentations of the global backbone folds of the two homeodomains, it is seen that the extra residues of the LFB1/HNF1 homeodomain loop out of an *Antp* homeodomain-like conformation, and that the structure formed by helix III and part of helix II in LFB1/HNF1 is closely related to a classical helix–turn–helix motif.

Conclusions

The backbone fold of the solution structure of the LFB1/HNF1 homeodomain (Figures 3 and 5A) clearly demonstrates that despite the limited sequence homology, this polypeptide belongs to the family of homeodomains. Three α-helical regions known from other homeodomain structures (Qian *et al.*, 1989; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991) also occur in the LFB1/HNF1 homeodomain, and they are nearly identically oriented with respect to each other. This homeodomain includes a non-classical helix–turn–helix motif of which the classical two residue turn is replaced by 23 residues looping out into an elongated second helix and a poorly defined loop region. This result confirms the previous hypothesis by Finney (1990), based on considerations of sequence homology, that the extra residues would be located between the second and third helices.

Further research with the LFB1/HNF1 system is directed

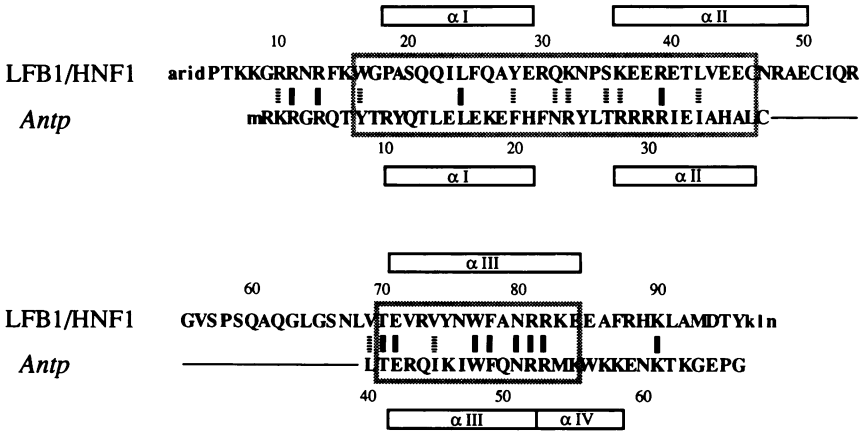

Fig. 4. Amino acid sequence alignment and secondary structure comparison of the presently studied LFB1/HNF1 homeodomain polypeptide and an *Antp* homeodomain polypeptide (Qian *et al.*, 1989). Identical amino acid residues are connected by solid vertical bars, and conservatively exchanged residues by stippled bars. The numeration is for the LFB1/HNF1 polypeptide used for the present studies, a thin line in the *Antp* sequence indicates the region 48–68 in LFB1/HNF1, where there is an insertion of 21 residues relative to the *Antp* homeodomain. Lower case letters represent those amino acid residues that come from the expression constructs and do not correspond to the sequences of the wild-type proteins. Rectangles above and below the sequences identify the helical regions in the LFB1/HNF1 and the *Antp* homeodomains, respectively. The boxed parts of the sequences identify those segments of the polypeptides which superimpose best in the 3D structures, with a RMSD value of 1.51 Å calculated for the backbone N, C^{α} and C′ atoms.

at answering the question as to possible functional roles of the extra residues in the LFB1/HNF1 homeodomain. In the current structure determination, the extra loop appears to be flexibly disordered. It will therefore be of special interest to see whether the structure of this polypeptide segment will change upon complexation with DNA.

Materials and methods

Protein expression and purification

A polypeptide containing the sequence 195–286 of the rat LFB1/HNF1 protein was expressed from the *Escherichia coli* strain BL21(DE3)/pT7.7XL-HOM and purified as described elsewhere (Tomei *et al.*, 1992). 200 pmol of this polypeptide were applied to a protein micro-sequencer (Knauer Model 810, Dr Ing H.Knauer GmbH, Berlin, Germany) equipped with an isocratic HPLC-system (Frank, 1989). The analysis of the N-terminal 20 residues showed that the LFB1/HNF1 homeodomain polypeptide used subsequently for the NMR studies does not contain an N-terminal Met, and starts with an Ala residue. For the NMR measurements the protein was further purified by reversed phase chromatography (C_8 column material, Pharmacia) using a linear gradient of 0–60% acetonitrile in H_2O in the presence of 0.1% trifluoroacetic acid. The fractions containing the protein were pooled and lyophilized. For the expression of the uniformly ^{15}N-labeled LFB1/HNF1 polypeptide, the cells were grown in 10 l of minimal medium (10 mM $(^{15}NH_4)_2SO_4$, 96 mM potassium phosphate buffer at pH 6.8, 0.5 mM Mg_2SO_4, 13 μM $FeSO_4$, 5 μM (+)-biotin, 7 μM vitamin B_1 and 30 mM D(+)-glucose). The growth temperature was 37°C, and the induction time was 4 h. 3.2 μmol of the polypeptide (50 mg) were obtained from 10 l of the cell culture. The $^{13}C,^{15}N$ doubly labelled LFB1/HNF1 polypeptide was produced using 4 l of the aforementioned minimal medium with 5 mM $[^{13}C_6]$-D(+)-glucose (1.3 g/l of fermenter grade material from Martek) in place of 30 mM of unlabelled glucose. Optimal protein yield was obtained by inducing the cells for 22 h at 26°C, including rapid exchange of the medium prior to induction, when the culture reached a cell density of OD_{600} = 0.8. The exchange of the medium was achieved by pelleting the cells for 10 min at 4500 *g* and room temperature, and resuspension of the cell pellet immediately afterwards in fresh, pre-warmed minimal medium followed by 10 min recovery at 26°C. The protein expression was induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). 2.5 μmol of $^{13}C,^{15}N$ doubly labelled LFB1/HNF1 polypeptide (40 mg) were obtained from 4 l of cell culture, with >98% uniform labelling for both isotopes.

A sample of biosynthetically directed fractionally ^{13}C-labelled (Szyperski *et al.*, 1992) LFB1/HNF1 polypeptide was prepared, using as a nutrient the $^{13}C,^{15}N$ doubly labelled medium from the aforementioned preparation, which had been separated from the cells before induction. To this medium we added 3.6 mg of unlabelled D(+)-glucose, and a sufficient amount of pelleted cells grown in an unlabelled minimal medium to reestablish a cell

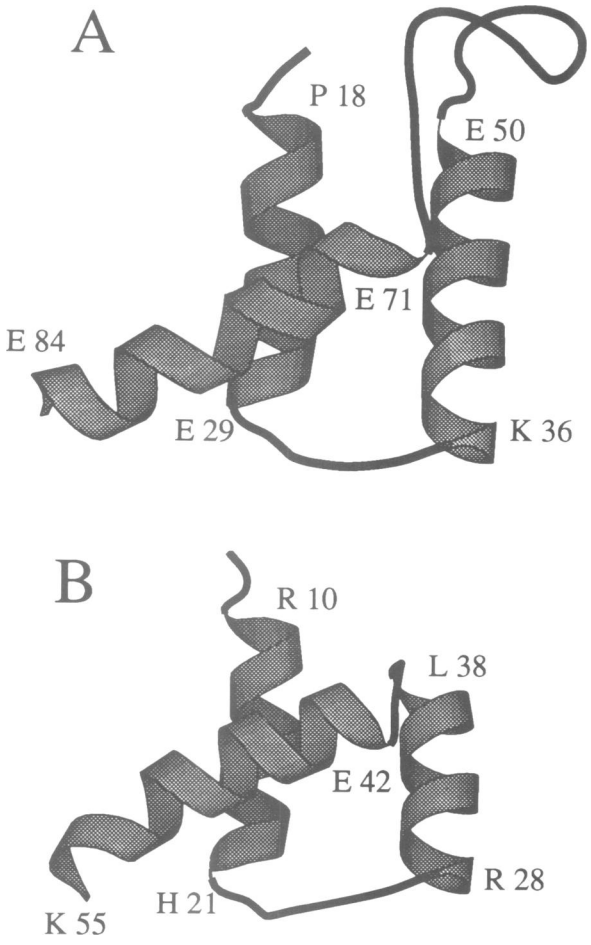

Fig. 5. (**A**) Ribbon drawing of the mean NMR solution structure of the LFB1/HNF1 homeodomain, which was obtained by averaging of the coordinates of the 20 conformers shown in Figure 3. (**B**) Same presentation of the *Antp* homeodomain (Qian *et al.*, 1989). The orientation of the two molecules was taken from a fit of the corresponding helical regions for minimal RMSD. The positions of selected residues are identified.

density of OD_{600} = 0.8. The final ratio of [$^{13}C_6$]glucose to [$^{12}C_6$]glucose was 0.3. The cell density was measured to be OD_{600} = 0.8. After 10 min recovery these cells were induced with 1 mM IPTG for 8 h at 37°C. We obtained 1.6 μmol of the fractionally labelled LFB1/HNF1 polypeptide (25 mg). The degree of ^{13}C labelling was determined by NMR to be 30%. The sample was also enriched with ^{15}N to the extent of ~80%.

NMR sample preparation

The solution conditions for the structure determination of the LFB1/HNF1 polypeptide were chosen on the basis of the following exploratory experiments. Circular dichroism spectroscopy at 100 μM protein concentration showed that the native protein conformation was significantly stabilized by the presence of 100 mM NaCl, where the native state could be maintained over the pH range 3.0–5.0 and over the temperature range from 15 to 20°C. The protein was found to precipitate at pH values between 5.5 and 8.0. The denaturation of the protein by heat or low pH was found to be reversible only under reducing conditions, when the two cysteinyl residues in the polypeptide (Figure 1) were protected from oxidation. NMR samples were prepared by dissolving the lyophilized protein in the argon-purged solvent (either H_2O/D_2O 9:1, or D_2O) containing 100 mM NaCl; the pH was adjusted to 3.6 by adding minute amounts of NaOH or NaOD, respectively. For measurements in D_2O, complete solvent exchange was achieved by incubation at 30°C for 1 h, lypohilization and redissolving in D_2O. The final protein concentrations in the individual samples were 2–4 mM. At concentrations above 2 mM and a temperature of 20°C, aggregation occurred after several days, accompanied by an increase of the viscosity of the solution. Therefore, the protein solutions had to be rechromatographed after ~2 weeks of measuring time.

NMR measurements

NMR measurements were performed using Bruker AM500 and AMX600 spectrometers. Quadrature detection was used in all dimensions, using either time-proportional phase incrementation (TPPI, Marion and Wüthrich, 1983) or States/TPPI (Marion *et al.*, 1989). Two-dimensional experiments included 2QF-$^1H,^1H$-COSY in D_2O (Rance *et al.*, 1985), clean-TOCSY in D_2O (Griesinger *et al.*, 1988) with a mixing time of 80 ms, $^1H,^1H$-NOESY in D_2O (Anil-Kumar *et al.*, 1980) with a mixing time of 80 ms, $^1H,^{15}N$-COSY and $^1H,^{13}C$-COSY (Bodenhausen and Ruben, 1980; Otting and Wüthrich, 1988). The following 3D spectra were recorded: 3D ^{15}N-correlated $^1H,^1H$-TOCSY (Fesik and Zuiderweg, 1988; Marion *et al.*, 1989), with a mixing time of 60 ms, t_{1max} = 28.8 ms, t_{2max} = 19.4 ms, t_{3max} = 131 ms, time domain data size 400 × 40 × 2048, total recording time ~80 h; 3D ^{15}N-correlated $^1H,^1H$-NOESY (Fesik and Zuiderweg, 1988; Marion *et al.*, 1989; Messerle *et al.*, 1989) with a mixing time of 80 ms, t_{1max} = 28.8 ms, t_{2max} = 19.4 ms, t_{3max} = 131 ms, time domain data size 400 × 40 × 2048, total recording time ~80 h; 3D HCCH-TOCSY (Bax *et al.*, 1990) with a mixing time of 15.5 ms, t_{1max} = 18.4 ms, t_{2max} = 9.0 ms, t_{3max} = 65.5 ms, time domain data size 256 × 56 × 1024, total recording time ~60 h; 3D ^{13}C-correlated $^1H,^1H$-NOESY in D_2O using the experimental scheme described by Messerle *et al.* (1989) with a selective π pulse at the ^{13}C frequency of the carbonyls in the middle of t_2, using a mixing time of 40 ms, t_{1max} = 29.4 ms, t_{2max} = 9.0 ms, t_{3max} = 65.5 ms, time domain data size 320 × 56 × 1024, total recording time ~ 96 h; 3D ^{13}C-correlated $^1H,^1H$-NOESY in H_2O using a novel experimental scheme derived from that of Messerle *et al.* (1989) by permutation of the time domains so that the $^{13}C-^1H$ moieties were frequency labelled by their ^{13}C- and ^1H-chemical shifts during t_1 and t_2, respectively, before the NOESY mixing period, which ensured that all αH–NH cross peaks could be observed at the amide proton chemical shift in ω_3 without interference from the t_1 noise (mixing time = 80 ms, t_{1max} = 9.0 ms, t_{2max} = 20.9 ms, t_{3max} = 65.5 ms, time domain data size 56 × 290 × 1024, total recording time ~144 h). The data sets were multiplied in all dimensions with phase shifted sine-ball functions (DeMarco and Wüthrich, 1976) prior to Fourier transformation. The time domain data of the ^{13}C-correlated 3D spectra were extended in the ^{13}C dimension by linear prediction from 28 to 46 data points, zero-filled to 64 points, and multiplied with a shifted sine-bell function before Fourier transformation using the program PROSA (Güntert *et al.*, 1992).

Amide proton exchange rates were measured by dissolving the lyophilized ^{15}N-labelled protein in D_2O and monitoring the time course of the amide proton signals in a series of $^1N,^{15}N$-COSY spectra (Otting and Wüthrich, 1988) with a recording time of ~40 min per spectrum.

Coupling constants $^3J_{HN\alpha}$ were measured with a series of J-modulated $^1H,^{15}N$-COSY spectra (Neri *et al.*, 1990; Billeter *et al.*, 1992). Seven experiments with J modulation periods of 12.1, 30.1, 45.1, 60.1, 75.1, 90.1 and 105 ms were recorded. Each individual recording took ~8.5 h. The program EASY (Eccles *et al.*, 1991) was used for peak picking and peak volume integration. The rate constants for the amide proton exchange were obtained using a non-linear fit routine.

Determination of the three-dimensional structure

The NOESY spectra were analysed using a 3D version of the program package EASY (Eccles *et al.*, 1991). The 2D and 3D cross peak volumes were converted to upper distance limits with the program CALIBA (Güntert *et al.*, 1989). Dihedral angle constraints based on local NOE distance constraints and experimental data on the scalar coupling constants $^3J_{NH\alpha}$ were generated with the program HABAS (Güntert *et al.*, 1989). Structure calculations were done using the program DIANA (Güntert *et al.*, 1991a), following a similar strategy to that used with the *Antp(G39S)* homeodomain (Güntert *et al.*, 1991b), except that RE-DAC (Güntert and Wüthrich, 1991) was also employed. Finally, the 20 DIANA conformers with the lowest residual target function values were subjected to restrained energy minimization using the program FANTOM (Schaumann *et al.*, 1990). The ensemble of the resulting 20 energy-minimized DIANA conformers is used to represent the solution structure of the LFB1/HNF1 homeodomain.

Structure comparison

For visual comparison of different structures, stereo views were produced either with Midas Plus (Ferrin *et al.*, 1988), the program Molscript (Kraulis, 1991), or with the structure analysis program XAM (Xia, 1992). Global superpositions and RMSD values for various subsets of atoms were computed as usual (McLachlan, 1979).

Acknowledgements

We thank C.Bartels for the use of a modified form of the program EASY, P.Güntert for help in performing non-linear fits to decay constant measurements, Dr M.Billeter for help with the structure calculations, Dr Y.Q.Qian for helpful discussions and Dr G.Frank for the N-terminal protein sequencing. B.Leiting was supported by the Deutsche Forschungsgemeinschaft (Le 704/1-1). Financial support by the Schweizerischer Nationalfonds (project 31.32033.91) is gratefully acknowledged. This paper was presented at the XV International Conference on Magnetic Resonance in Biological Systems, Jerusalem, Israel, August 16–22, 1992.

References

Affolter,M., Percival-Smith,A., Müller,M., Leupin,W. and Gehring,W.J. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4093–4097.

Anil-Kumar, Ernst,R.R. and Wüthrich,K. (1980) *Biochem. Biophys. Res. Commun.*, **95**, 1–6.

Bax,A., Clore,G.M., Driscoll,P.C., Gronenborn,A.M., Ikura,M. and Kay,L.E. (1990) *J. Magn. Reson.*, **87**, 620–626.

Billeter,M., Braun,W. and Wüthrich,K. (1982) *J. Mol. Biol.*, **155**, 321–346.

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

Billeter,M., Neri,D., Otting,G., Qian,Y.Q. and Wüthrich,K. (1992) *J. Biomol. NMR*, **2**, 257–274.

Bodenhausen,G. and Ruben,D. (1980) *Chem. Phys. Lett.*, **69**, 185–188.

DeMarco,A. and Wüthrich,K. (1976) *J. Magn. Reson.*, **24**, 201–204.

DeSimone,V. and Cortese,R. (1991) *Curr. Opin. Cell Biol.*, **3**, 960–965.

Driscoll,P.C., Gronenborn,A.M., Wingfield,P.T. and Clore,G.M. (1990) *Biochemistry*, **29**, 4668–4682.

Eccles,C., Güntert,P., Billeter,M. and Wüthrich,K. (1991) *J. Biomol. NMR*, **1**, 111–130.

Ferrin,T.E., Huang,C.C., Jarvis,L.E. and Langridge,R. (1988) *J. Mol. Graph.*, **6**, 13–27.

Fesik,S.W. and Zuiderweg,E.R.P. (1988) *J. Magn. Reson.*, **78**, 588–593.

Finney,M. (1990) *Cell*, **60**, 5–6.

Frain,M., Swart,G., Monaci,P., Nicosia,A., Stämpfli,S., Frank,R. and Cortese,R. (1989) *Cell*, **59**, 145–157.

Frank,G. (1989) In Wittmann-Liebold,B. (ed.), *Methods in Protein Sequence Analysis*. Proc. 7th International Conference, Springer Verlag, Berlin, Germany, pp. 116–121.

Gehring,W.J. (1987) *Science*, **236**, 1245–1252.

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

Güntert,P. and Wüthrich,K. (1991) *J. Biomol. NMR*, **1**, 447–456.

Güntert,P., Braun,W., Billeter,M. and Wüthrich,K. (1989) *J. Am. Chem. Soc.*, **111**, 3997–4004.

Güntert,P., Braun,W. and Wüthrich,K. (1991a) *J. Mol. Biol.*, **217**, 517–530.

Güntert,P., Qian,Y.Q., Otting,G., Müller,M., Gehring,W. and Wüthrich,K. (1991b) *J. Mol. Biol.*, **217**, 531–540.
Güntert,P., Dötsch,V., Wider,G. and Wüthrich,K. (1992) *J. Biomol. NMR*, **2**, 619–629.
Hayashi,S. and Scott,M.P. (1990) *Cell*, **68**, 283–302.
Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579–590.
Kraulis,P. (1991) *J. Appl. Cryst.*, **24**, 946–950.
Marion,D. and Wüthrich,K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.
Marion,D., Kay,L.E., Sparks,S.W., Torchia,D.A. and Bax,A. (1989) *J. Am. Chem. Soc.*, **111**, 1515–1517.
McLachlan,A.D. (1979) *J. Mol. Biol.*, **128**, 49–79.
Mendel,D.B. and Crabtree,G.R. (1991) *J. Biol. Chem.*, **266**, 677–680.
Messerle,B.A., Wider,G., Otting,G., Weber,C. and Wüthrich,K. (1989) *J. Magn. Reson.*, **85**, 608–6013.
Müller,M., Affolter,M., Leupin,W., Otting,G., Wüthrich,K. and Gehring,W.J. (1988) *EMBO J.*, **7**, 4299–4304.
Neri,D., Otting,G. and Wüthrich,K. (1990) *J. Am. Chem. Soc.*, **112**, 3663–3665.
Neri,D., Wider,G. and Wüthrich,K. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 4397–4401.
Nicosia,A., Monaci,P., Tomei,L., DeFrancesco,R., Nuzzo,M., Stunnenberg,H. and Cortese,R. (1990) *Cell*, **61**, 1225–1236.
Otting,G. and Wüthrich,K. (1988) *J. Magn. Reson.*, **76**, 569–574.
Otting,G., Qian,Y.Q., Müller,M., Affolter,M., Gehring,W.J. and Wüthrich,K. (1988) *EMBO J.*, **7**, 4305–4309.
Otting,G., Qian,Y.Q., Billeter,M., Müller,M., Affolter,M., Gehring,W.J. and Wüthrich,K. (1990) *EMBO J.*, **9**, 3085–3092.
Pardi,A., Billeter,M. and Wüthrich,K. (1984) *J. Mol. Biol.*, **180**, 741–751.
Qian,Y.Q., Billeter,M., Otting,G., Müller,M., Gehring,W.J. and Wüthrich,K. (1989) *Cell*, **59**, 573–580.
Rance,M., Bodenhausen,G., Wagner,G., Wüthrich,K. and Ernst,R.R. (1985) *J. Magn. Reson.*, **62**, 497–510.
Schaumann,T., Braun,W. and Wüthrich,K. (1990) *Biopolymers*, **29**, 679–706.
Scott,M.P., Tamkun,J.W. and Hartzell,G.W. (1989) *Biochim. Biophys. Acta*, **989**, 25–49.
Senn,H., Werner,G., Messerle,B.A., Weber,C., Traber,R. and Wüthrich,K. (1989) *FEBS Lett.*, **249**, 113–118.
Szyperski,T., Neri,D., Leiting,B., Otting,G. and Wüthrich,K. (1992) *J. Biomol. NMR*, **2**, 323–334.
Tomei,L., Cortese,R. and De Francesco,R. (1992) *EMBO J.*, **11**, 4119–4129.
Wagner,G. and Wüthrich,K. (1982) *J. Mol. Biol.*, **155**, 347–366.
Wider,G., Lee,K.H. and Wüthrich,K. (1982) *J. Mol. Biol.*, **155**, 367–388.
Wolberger,C., Vershon,A.K., Liu,B., Johnson,A.D. and Pabo,C.O. (1991) *Cell*, **67**, 517–528.
Wüthrich,K. (1976) *NMR in Biological Research: Peptides and Proteins*. Elsevier, Amsterdam, The Netherlands.
Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.
Wüthrich,K. and Gehring,W. (1992) In McKnight,S.L. and Yamamoto,K. (eds), *Transcriptional Regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, in press.
Wüthrich,K., Billeter,M. and Braun,W. (1984) *J. Mol. Biol.*, **180**, 715–740.
Wüthrich,K., Spitzfaden,C., Memmert,K., Widmer,H. and Wider,G. (1991) *FEBS Lett.*, **2**, 237–247.
Wüthrich,K., Szyperski,T., Leiting,B. and Otting,G. (1992) In Takai,K. (ed.), *Proc. 1st Biennial International Conference on Amino Acid Research*. Elsevier, Amsterdam, The Netherlands, pp. 41–48.
Xia,T. (1992) *Software for Determination and Visual Display of NMR Structures of Proteins: the Distance Geometry Program DGPLAY and the Computer Graphics Programs CONFOR and XAM*. Ph.D. thesis Nr. 9831, ETH Zürich, Switzerland.

Received on November 16, 1992; revised on January 14, 1993