J. Mol. Biol. (1997) **267**, 673–683

JMB

The NMR Solution Structure of the Non-classical Homeodomain from the Rat Liver LFB1/HNF1 Transcription Factor

Oliver Schott, Martin Billeter*, Barbara Leiting, Gerhard Wider and Kurt Wüthrich

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg CH-8093 Zürich, Switzerland

The nuclear magnetic resonance (NMR) solution structure of the non-classical homeodomain from the rat liver LFB1/HNF1 transcription factor was determined with the program DIANA from an input of 1356 nuclear Overhauser enhancement (NOE) upper distance constraints and 228 dihedral angle constraints collected using experiments with the unlabelled, the uniformly ^{15}N-labelled and the uniformly ^{13}C-labelled protein. Out of a group of 50 independently calculated conformers the 20 conformers with the smallest residual DIANA target function values were refined by energy minimization with the program OPAL and are used to represent the NMR structure. The average of the pairwise root-mean-square deviations (r.m.s.d.) of these 20 individual NMR conformers relative to the mean coordinates is 0.73 Å (1 Å = 0.1 nm) for the backbone atoms N, C^{α} and C′ of residues 15 to 82. The chain-terminal polypeptide segments 1–14 and 90–99 are disordered in solution. The globular fold contains three well-defined helices comprising the residues 19 to 29, 37 to 53 and 71 to 81, and the third helix is extended by a less well-ordered fourth helix with residues 82 to 89, which coincides with corresponding observations in ''classical'' homeodomains. Side-chain analysis resulted in 33 ''best-defined'' side-chains, with global displacements smaller than 1.1 Å, and addition of these side-chains to the global superposition of residues 15 to 82 resulted in a r.m.s.d of 0.81 Å. The protein contains two hydrophobic cores, one of which corresponds to the helical packing seen in classical homeodomains, while the other one stabilizes the conformation of the 21-residue insertion between helices II and III. The individual helices and their relative spatial arrangements are stabilized by a variety of structural motifs, which include medium-range and long-range hydrogen bonds and salt bridges. Detailed comparison with the *Antennapedia* homeodomain, and studies of the complex formation with an operator DNA half-site provided initial information on the DNA-binding mode of the LFB1/HNF1 homeodomain.

Keywords: LFB1/HNF1 transcription factor; homeodomains; transcriptional regulation; protein structure; NMR

*Corresponding author

Introduction

The LFB1/HNF1-protein is a major component of the hepatocyte-specific expression apparatus of many genes (DeSimone & Cortese, 1991; Mendel & Crabtree, 1991). It activates the transcription of several liver-specific genes by binding as a dimer to palindromic *cis*-acting elements. A 99-residue recombinant polypeptide containing the sequence of residues 195 to 286 of the wild-type protein in positions 5 to 96 binds as a monomer to a half-palindromic binding site with a stability of

Present address: Barbara Leiting, Merck Research Laboratories, Mail Code RY50-105, P.O. Box 2000, Rahway, NJ 07065, USA

Abbreviations used: LFB1/HNF1-homeodomain, polypeptide fragment 195 to 286 of the LFB1/HNF1 transcription factor from rat liver; *Antp*, *Antennapedia*; *ftz*, *fushi tarazu*; 1D, 2D, 3D, one-, two-, three-dimensional; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; COSY, correlation spectroscopy; TOCSY, total correlation spectroscopy; r.m.s.d., root mean square deviation.

0022–2836/97/130673–11 $25.00/0/mb970905

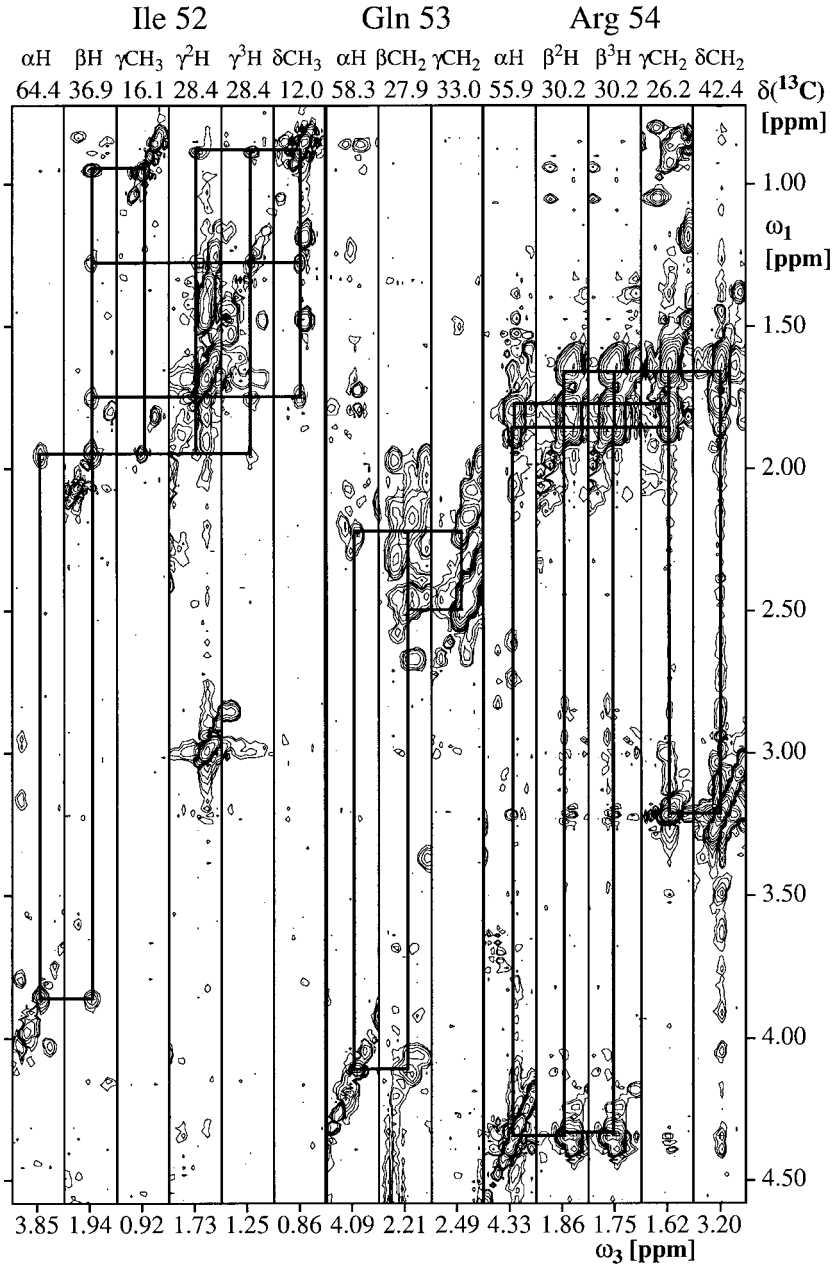

Figure 1. Composite plot of $\omega_1(^1H)-\omega_3(^1H)$ strips taken from a 3D HCCH-COSY spectrum recorded with the uniformly ^{13}C-labelled LFB1/HNF1-homeodomain (protein concentration 3 mM, solvent ^2H_2O, $p^2H = 4.6$, $t = 22°C$, ^1H-frequency = 600 MHz). Individual strips of the residues Ile52, Gln53 and Arg54 were taken from different $\omega_1(^1H)-\omega_3(^1H)$ planes at the corresponding ^{13}C frequencies and ordered according to the side-chain atom positions from α to the periphery. The assignment pathways are indicated with lines connecting the diagonal peaks and cross-peaks of the individual pairs of coupled resonances. At the top and the bottom of the individual strips the ^1H and ^{13}C chemical shifts, respectively, of the diagonal peaks are indicated.

$K_d \approx 10^{-9}$ M (Tomei *et al.*, 1992). In this construct the LFB1/HNF1-homeodomain, which differs from classical homeodomains (Gehring *et al.*, 1994) by a 21-residue insertion between the two helices of the helix-turn-helix motif, extends from residues 9 to 89. The global backbone fold of this protein fragment was previously determined by NMR (Leiting *et al.*, 1993), and a crystal structure at 2.8 Å resolution of the same polypeptide is also available (Ceska *et al.*, 1993). Although the crystals used for the structure determination were grown in a solution containing the protein and a DNA duplex with the operator sequence, only the protein was observed in the crystals. It thus appears unlikely that co-crystals containing a DNA complex of this homeodomain can readily be obtained, and NMR may thus be the only technique capable of providing further structural data on this system. As an essential step toward a complete structure determination of the complex, we describe here a high-quality NMR structure determination of the free LFB1/HNF1-homeodomain.

Results and Discussion

Resonance assignments

As is commonly observed in proteins containing exclusively helices as regular secondary structures, the homonuclear 1D and 2D ^1H NMR spectra of the LFB1/HNF1-homeodomain are poorly resolved. Therefore we used almost exclusively heteronuclear-resolved three-dimensional data sets for the

present study. All spectra used for the resonance assignment and the structure determination were recorded at pH 4.6, which differs from the pH value of 3.6 used by Leiting *et al.* (1993), and the temperature was 22°C throughout.

For the sequence-specific resonance assignments we initially followed the standard strategy of establishing sequential connectivities between the ^{1}H spin systems of the individual amino acid residues with ^{1}H–^{1}H NOEs (Wüthrich, 1986), using 3D ^{15}N-resolved ^{1}H–^{1}H COSY and NOESY spectra. The sequence-specific resonance assignments of the polypeptide backbone were subsequently verified using a constant-time 3D HNCA experiment recorded with a ^{13}C,^{15}N-doubly labelled protein sample. Only a small number of residues showed significant chemical shift changes between the presently used pH value of 4.6 and pH 3.6 (Leiting *et al.*, 1993). The previous partial resonance assignments for the amino acid side-chains (Leiting *et al.*, 1993) were completed using 3D HCCH-COSY, and a 3D HCCH-TOCSY experiment with a mixing time of 19.3 ms, both recorded with a uniformly ^{13}C-labelled protein sample. As an illustration, Figure 1 shows the side-chain assignments by HCCH-COSY for the residues 52 to 54.

The ^{1}H, ^{13}C and ^{15}N chemical shifts of the LFB1/HNF1-homeodomain are indirectly referenced (Wishart *et al.*, 1995). Complete assignments were obtained for the polypeptide backbone and for the non-labile protons of the amino acid-side chains, with the sole exceptions of δCH_2 of Arg13, δC of Arg30 and γC of Arg39. The peripheral side-chain NH_n moieties were only partially assigned because of extensive spectral overlap. The hydroxyl protons of Ser, Thr and Tyr were not observed.

Individual proton assignments for 14 βCH_2 groups were obtained with the program HABAS. In addition, 18 out of the 22 methyl groups of valine and leucine were stereospecifically assigned with the use of biosynthetically directed fractional ^{13}C labelling (Senn *et al.*, 1989; Neri *et al.*, 1989).

Collection of conformational constraints and structure calculation

On the basis of the complete ^{1}H resonance assignments, a more extensive set of NOEs has been assigned than in previously reported work (Leiting *et al.*, 1993), using the following three NOESY spectra (see Materials and Methods): 2D [^{1}H,^{1}H]-NOESY in ^2H_2O solution with a mixing time of 40 ms; 3D ^{15}N-resolved [^{1}H,^{1}H]-NOESY in H_2O solution with a mixing time of 50 ms; 3D ^{13}C-resolved [^{1}H,^{1}H]-NOESY in ^2H_2O solution with a mixing time of 60 ms. A total of 2792 cross-peaks were assigned in these three data sets. NOESY cross-peak integration was achieved using an option of the program package XEASY (Bartels *et al.*, 1995). In addition, 84 $^3J_{HN\alpha}$ homonuclear coupling constants were measured by a series of *J*-modulated [^{15}N,^{1}H]-COSY spectra (Billeter *et al.*, 1992; Neri *et al.*, 1990). After the initial processing of these

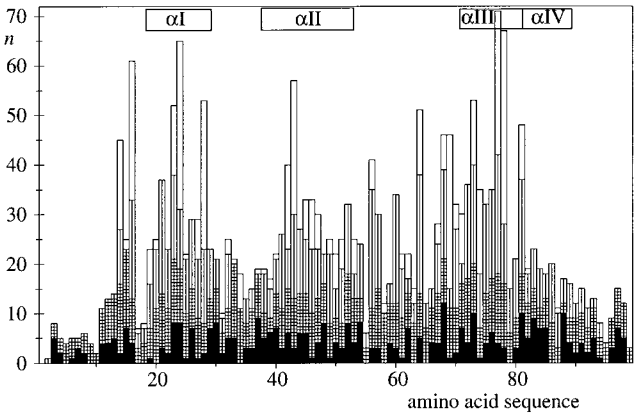

Figure 2. Plot of the number of NOE upper distance constraints per residue, *n*, *versus* the sequence of the LFB1/HNF1-homeodomain. The constraints are specified as follows: filled, intraresidual; cross-hatched, constraints between protons in sequentially neighbouring residues; vertically hatched, medium-range constraints between protons located in residues separated by two to five positions along the sequence; open, all longer-range constraints. At the top the sequence locations of the regular secondary structures are indicated.

data with the programs HABAS (Güntert *et al.*, 1989) and DIANA (Güntert *et al.*, 1991), the input of ''relevant'' constraints for the structure calculations with the program DIANA consisted of 1356 NOE upper limits on proton–proton distances (369 intraresidual, 359 sequential, 400 medium-range, 228 long-range) and 228 constraints on ϕ, ψ and χ^1 angles. The sequence distribution of the NOE distance constraints is shown in Figure 2. High constraint density is found in the α-helical regions. No medium-range or long-range NOE distance constraints can be found either for the N-terminal 13-residue segment or for the C-terminal 10-residue segment, showing that these two chain terminal polypeptide segments are not observably involved in non-bonding interactions with the protein core.

The quality of the NMR structure calculated with the program DIANA (Güntert *et al.*, 1991) and refined by energy-minimization with the program OPAL (Luginbühl *et al.*, 1996) is documented with the Tables 1 and 2, and Figures 3A and 4. Table 1 shows that the NMR structure represented by the 20 best DIANA conformers represents a good fit of the experimental data, with very small residual constraint violations. Table 2 shows that the LFB1/HNF1-homeodomain can be structurally characterized with good precision for the polypeptide backbone segment 15–82 and for 33 ''best-defined'' amino acid side-chains forming the protein core. This is supplemented by the visual display of the displacements of residues 14 to 89 after superposition for minimal r.m.s.d of the backbone atoms N, C^α and C′ of residues 15 to 82 (Figure 3A), which shows that the helices I, II and III have only small local backbone displacements, while increased disorder is found in the loop connecting the helices II

Table 1. Analysis of the 20 DIANA conformers of LFB1/HNF1 used to represent the NMR structure, after restrained energy minimization with the program OPAL

Quantity	Average ± standard deviation (range)[a]
DIANA target function (Å^2)[b]	0.65 ± 0.19 (0.31 . . . 0.96)
AMBER energy (kcal/mol)[c]	−3269.2 ± 190.9 (−3761.5 . . . − 2906.0)
Residual violations of NOE distance constraints	
Sum (Å)	8.42 ± 0.56 (7.17 . . . 9.75)
Maximum (Å)	0.10 ± 0.00 (0.10 . . . 0.11)
Residual violations of dihedral angle constraints	
Sum (degrees)	10.0 ± 2.9 (4.3 . . . 15.1)
Maximum (degrees)	1.8 ± 0.3 (1.1 . . . 2.2)

A total of 50 conformers were calculated and the 20 structures with the smallest residual DIANA target function values were subjected to energy minimization in a water bath with a water shell of minimal thickness 6.0 Å.

[a] For the group of 20 conformers after energy minimization.

[b] The value given for the DIANA target function corresponds to the value before energy minimization (the DIANA target function is not defined after energy minimization, since the conformers no longer have ECEPP standard geometry).

[c] The value given represents the intra-protein interaction energy.

and III, and in the helix IV from residues 82 to 89. The N and C-terminal segments are disordered in solution and are not shown in Figure 3A. The data of Figure 3A are quantified by the plots of the variability along the sequence of the backbone dihedral angles in Figure 4.

The global architecture of the LFB1/HNF1-homeodomain

The global architecture is dominated by the spatial arrangement of the four α-helices (Figure 3A), which are the only regular secondary structures, where helix IV is an extension of the ''recognition helix'' III (see below). α-Helices were identified when a hydrogen bond between the carbonyl oxygen of residue i and the amide proton of residue $i + 4$ was assigned in at least ten conformers. This resulted in the identification of helix I from residues 19 to 29, helix II from residues 37 to 53, helix III from residues 71 to 81, and helix IV from residues 82 to 89. For these helical segments the ϕ and ψ dihedral angles are within a narrow range centred about the standard α-helix values of −57 and −47°, respectively (Figure 4). The helices I and II lie almost antiparallel in a down-and-up arrangement in the orientation of Figure 3A, while the helices III and IV are oriented approximately perpendicular to this two-helix bundle. The helices are held together by a core of ten tightly packed hydrophobic amino acids (Trp16, Ile23, Leu24, Ala27, Leu42, Val43, Val69, Val74, Trp77, Phe78). Seven of these amino acids are either invariant (Trp77, Phe78) or highly conserved (Trp16, Leu24, Leu42, Val69, Val74) in all homeodomains (Scott *et al.*, 1989). The invariant Trp77 and Phe78 and the highly conserved residues 16, 24, 42, 69 and 74 are all buried inside the molecule. As mentioned before, the N and C-terminal segments 1–13 and 90–99 (not shown in Figure 3) are disordered in solution, which is expected from the scarcity of medium-range and long-range distance constraints (Figure 2). This global fold coincides quite closely with a previous NMR investigation (Leiting *et al.*, 1993) and a low resolution X-ray crystal structure (Ceska *et al.*, 1993; see below). Compared to the report by Leiting *et al.* (1993) the entire polypeptide backbone structure is more precisely defined and the description of the fold is supported by precise positioning of a large fraction of the amino acid side-chains (Table 2). The only significant qualitative difference is that the helix II now extends from residues 37 to 53, rather than 36 to 50.

A total of 29 out of the 33 ''best-defined'' side-chains (Table 2) are hydrophobic. Conversely, with the sole exception of Phe25, which sticks out into the solvent, all hydrophobic side-chains belong to the best-defined side-chains, and no charged side-chain is found in this group (Table 2). The best-de-

Table 2. Quantitative characterization of the structure determination of the LFB1/HNF1 homeodomain by NMR in solution, comparison with the X-ray structure of LFB1/HNF1, and comparison with the NMR solution structure of the *Antennapedia* homeodomain

Atoms used for comparison	r.m.s.d. (Å)
⟨NMR⟩ - NMR[a]	
N, C^α, C′ of residues 15 to 82	0.73 ± 0.11
Same with best-defined side-chains[b]	0.81 ± 0.09
Same with all heavy atoms	1.27 ± 0.08
N, C^α, C′ of residues 15 to 89	0.93 ± 0.20
N, C^α, C′ of residues 71 to 89	0.70 ± 0.23
⟨NMR⟩ - X-ray[c]	
N, C^α, C′ of residues 15 to 82	1.13
Same with best-defined side-chains	1.77
⟨NMR⟩(LFB1/HNF1) - ⟨NMR⟩(*Antp*)[d]	
N, C^α, C′ (16–46, 70–84; 8–38, 41–55)[e,f]	1.40
N, C^α, C′ (16–46, 70–89; 8–38, 41–60)[e,g]	1.53
N, C^α, C′ (36–46, 70–81; 28–38, 41–52)[e,h]	0.85

[a] ⟨NMR⟩ denotes the average coordinates of the 20 conformers that represent the solution structure of the LFB1/HNF1-homeodomain, which were calculated after superposition of the conformers 2 to 20 onto conformer 1 for minimal r.m.s.d. of the backbone atoms of the residues 15 to 82; NMR denotes the 20 individual, energy-refined DIANA conformers.

[b] 33 residues with $D^{sc}_{glob} \leqslant 1.1$Å: Trp16, Pro18, Ala19, Ser20, Ile23, Leu24, Gln26, Ala27, Tyr28, Pro34, Thr41, Leu42, Val43, Cys46, Asn47, Ala49, Cys51, Ile52, Val56, Ser57, Pro58, Ser59, Ala61, Leu64, Leu68, Val69, Thr70, Val72, Val74, Tyr75, Trp77, Phe78, Ala79.

[c] The coordinates of the X-ray structure were taken from the Protein Data Bank (PDB-entry 1lfb).

[d] The mean structure of the *Antennapedia* homeodomain was calculated for residues 6–58, using the data from Billeter *et al.* (1990).

[e] The first two segments are from the LFB1/HNF1 homeodomain, the latter two segments contain the corresponding locations in the *Antp* homeodomain (Billeter *et al.*, 1990).

[f] These two segments give rise to optimal superposition of the 3D structures.

[g] These two segments comprise all four helices.

[h] These two segments form the two helices of the helix–turn–helix motif.

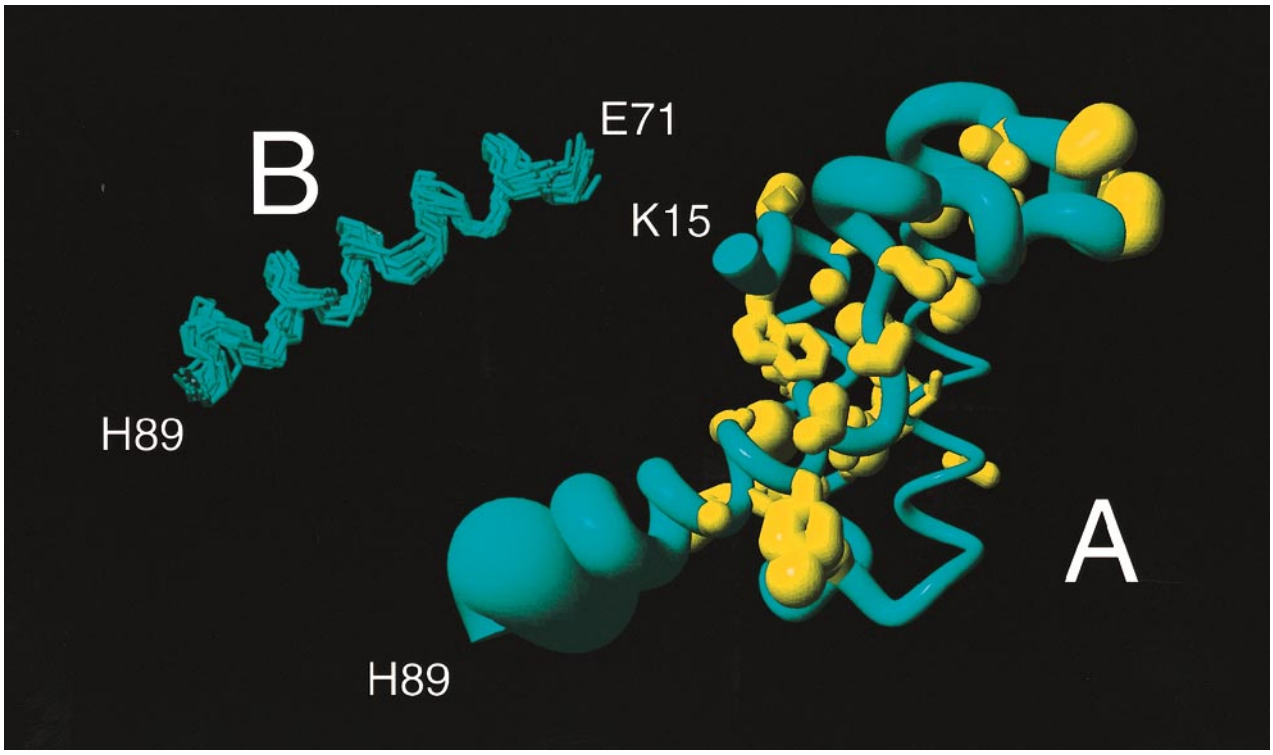

Figure 3. A, Presentation of the NMR structure of the LFB1/HNF1-homeodomain by a cylindrical rod representing, by its variable radius, the displacements of individual heavy atoms among the 20 energy-minimized DIANA conformers used to represent the solution structure. The polypeptide backbone of residues 15 to 89 and the 33 best-defined side-chains (see footnote to Table 2) are shown, with the backbone represented by a spline function through the C^{α}-positions and the side-chains by bonds between the heavy atoms. To compute the displacements the conformers 2 to 20 were superimposed for pairwise minimal r.m.s.d. of the backbone atoms N, C^{α}, C′ of residues 15 to 82 with conformer 1. B, Local superposition of the 20 final, energy-minimized DIANA conformers for best fit of the backbone atoms N, C^{α} and C′ of residues 71 to 89, which form the helices III and IV (see the text).

fined side-chains form two hydrophobic clusters, one of which is in the centre of the protein and corresponds to the core observed in classical homeodomains. A second, smaller cluster is formed by the extended loop between helices II and III, which consists of the insertion segment (Figure 3A). The charge distribution observed in the LFB1/HNF1-homeodomain shows distinct clusters of positive and negative charges. In particular, the helices III and IV, which are known to have contacts with the major groove of the DNA in other homeodomain-DNA complexes, carry the positive charges of Arg73, Arg81, Arg82, Lys83, Lys90 and Lys91.

Special secondary and tertiary structure motifs

A variety of structural motifs stabilize the individual helices (Figure 5). Helix I is N-capped by Gly17 and initiated by Pro18 (Figure 5A). These residue types are typically found at the onset of α-helical secondary structures (Richardson & Richardson, 1988). The N terminus of helix I is further stabilized by a hydrogen bond $H^{\varepsilon 1}$ Gln21–O′ Gly17, and its C terminus is stabilized by an (*i*,*i* + 4) cation-π interaction (Dougherty, 1996) between ζNH_3^+ of Lys32 and the aromatic ring of Tyr28 (Figure 5A). The loop between helices I and II is bridged by a hydrogen bond $H^{\varepsilon 1}$ Gln31–O^{ε} Glu38 (not shown). Helix II is N-capped by a Pro34–Ser35 dipeptide, and the side-chains of Lys36 and Glu40 form an (*i*,*i* + 4) salt bridge between their charged groups (Figure 5B). An additional salt bridge in helix II is observed between the side chains of Glu44 and Arg48, and it is C-capped by Gly55, which has an α_L-conformation, and the C-terminal end is further stabilized by a hydrogen bond O^{ε} Glu50–H^{ε} Gln53. Helix III (not shown) is initiated by Thr70 and Glu71, which is a residue type that is often observed in one of the first three helical positions (Richardson & Richardson, 1988), and there is a hydrogen bond $H^{\delta 1}$ Asn76–O′ Val72. In helix IV there is a salt bridge between the charged groups of Arg81 and Glu85 (Figure 6).

The global fold of the LFB1/HNF1-homeodomain is stabilized by numerous specific interhelical interactions (Figure 6). The packing of helix I against helix II involves a hydrogen bond H^{ε} Gln26–O^{ε} Glu45. Long-range cation-π interactions have been identified between the guanidinium group of Arg73 and the aromatic ring of Phe14, and between Arg81 and Tyr28, and there is a hydrogen bond $H^{\delta 2}$ Asn47–O′ Val69. Helix IV is held against the C-terminal end of helix I by a hydrogen bond O^{ε} Glu85–H^{η} Tyr28. It is worth noting, that the guanidinium group of Arg81 is sandwiched

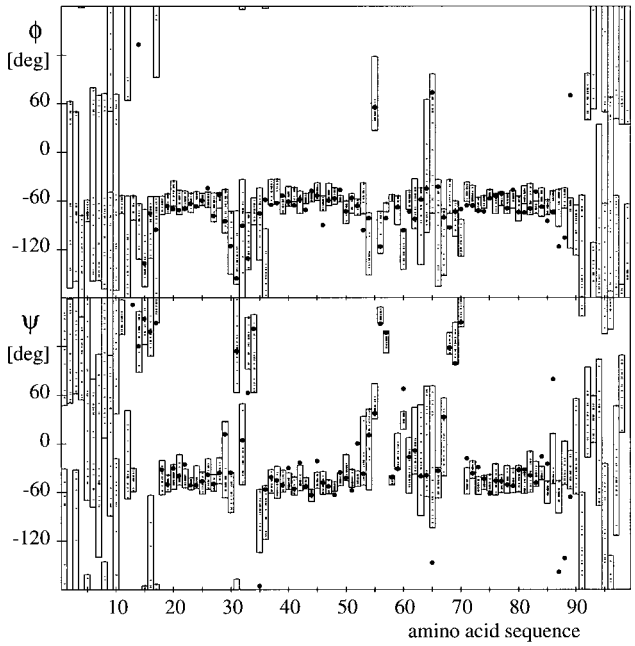

Figure 4. Plot of the dihedral angle values ϕ and ψ *versus* the sequence of the LFB1/HNF1-homeodomain. The small dots correspond to the ϕ and ψ angle values in the 20 energy-minimized DIANA conformers used to represent the solution structure, the large dots represent the value in the crystal structure (the coordinates were reported only for the segment 13–89), and the rectangles indicate the dihedral angle ranges spanned by the 20 NMR conformers.

between the aromatic rings of Tyr28 and Trp77, which gives rise to a 2-fold cation-π interaction (Figure 6).

Comparison of the NMR and X-ray crystal structures of the LFB1/HNF1-homeodomain

The crystal structure of the LFB1/HNF1-homeodomain has been determined at a resolution of 2.8 Å (Ceska *et al.*, 1993). The r.m.s.d. value between the mean coordinates of the NMR structure and the crystal structure of 1.13 Å (Table 2) is based on best-fit superposition for minimal r.m.s.d. of the segment 15–82, which includes the 21-residue insertion in the helix-turn-helix motif. Since the average of the pairwise r.m.s.d. values between the mean NMR coordinates and the 20 individual NMR conformers is only 0.73 Å (Table 2), there must be some significant differences between the NMR and crystal structures. This conclusion is also supported by the displacements between the 33 side-chains which are best-defined in the NMR structure (Table 2), as calculated after superposition of the backbone atoms N, C^{α} and C' of residues 15 to 82. Inspection of Figure 7, which provides a detailed comparison of the NMR and crystal structures for the polypeptide backbone and the heavy atoms of the 33 best-defined side-chains, shows that differences occur at the end of

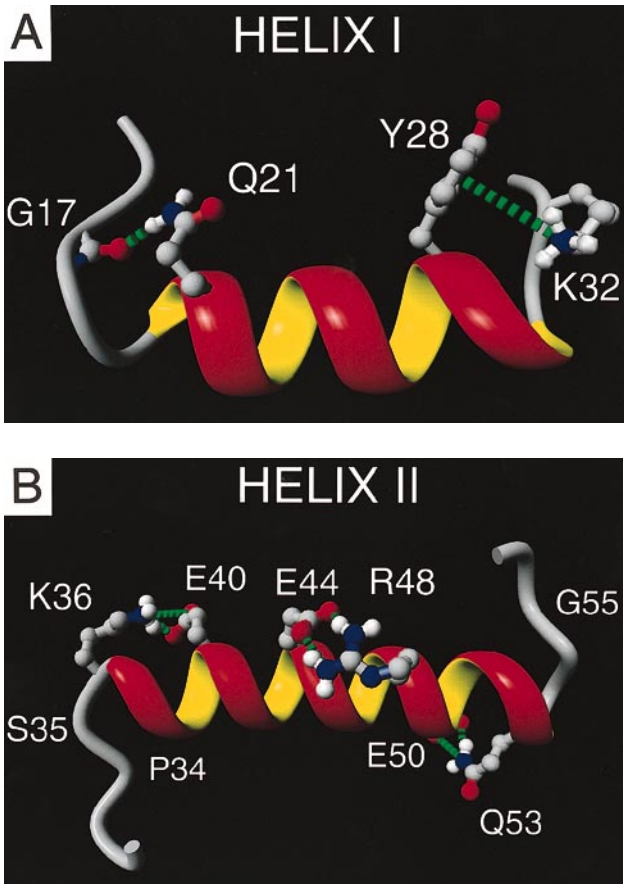

Figure 5. Local structural motifs which stabilize individual helices in the NMR solution structure of the LFB1/HNF1-homeodomain. The polypeptide backbone is shown as a red and yellow ribbon for the helices, and as a grey cylinder otherwise. The side-chains are represented as ball-and-stick models with the following colour code: carbon atoms, grey; nitrogen atoms, blue; oxygen atoms, red; hydrogen atoms, white. Hydrogen atoms are displayed only when part of a specific interaction. Hydrogen bonds, salt bridges and cation-π interactions are indicated by green, broken cylinders. A, Helix I, with the N-capping residue Gly17 forming the hydrogen bond $H^{\varepsilon 1}$ Gln21–O′ Gly17, and the C-terminal cation-π interaction between the positively charged ζNH_3^+-group of Lys32 and the aromatic ring of Tyr28. B Helix II, with the N-terminal dipeptide segment Pro34–Ser35, two salt bridges between charged side-chain groups, Lys36–Glu40 and Glu44–Arg48, the hydrogen bond O^{ε} Glu50–H^{ε} Gln53, and the C-capping residue Gly55, which is in the α_L-conformation.

helix I, in the loop connecting the helices II and III, and at the end of helix IV. The core side-chains can be quite tightly superimposed. These visual impressions are supplemented by comparison of the backbone–backbone hydrogen bonds in the two structures, which shows that the NMR structure contains regular α-helical structure through residue 31, whereas the C-terminal turn of helix I in the crystal structure is distorted and includes a 3_{10}-helix hydrogen bond. Extensive differences in hydrogen bonding are observed in the ''insertion''

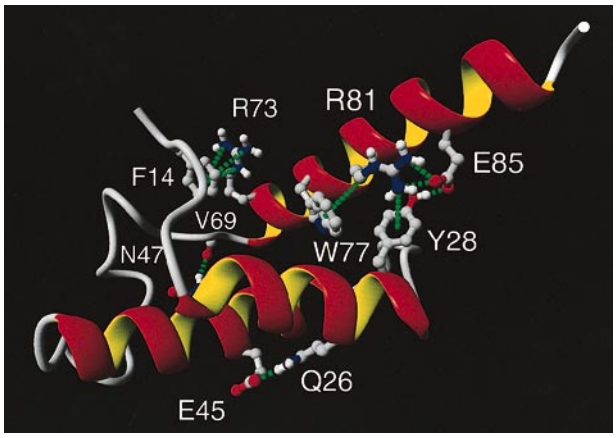

Figure 6. Hydrogen bonding interactions, salt bridges and cation–π interactions that stabilize the global fold of the LFB1/HNF1-homeodomain. Same presentation as in Figure 5. The following specific interactions are shown: the hydrogen bond $H^{\delta 2}$ Asn47–O′ Val69, cation-π interactions between the guanidinium group of Arg73 and the aromatic ring of Phe14, the hydrogen bond H^{ε} Gln26–O^{ε} Glu45, cation–π interactions between the guanidinium group of Arg81 and the aromatic rings of Trp77 and Tyr28, a salt bridge between the side-chains of Arg81 and Glu85, and a hydrogen bond O^{e} Glu85–H^{η} Tyr28.

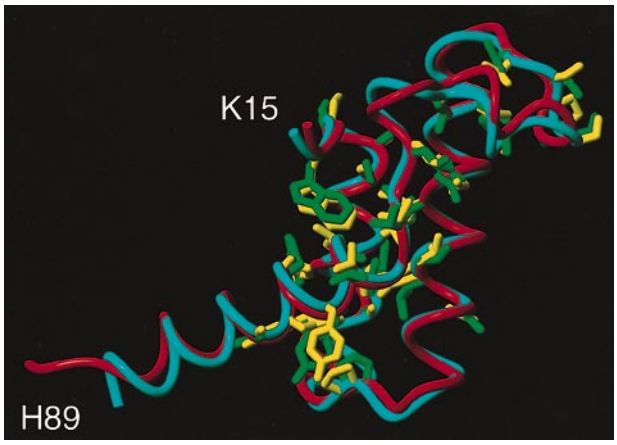

Figure 7. Comparison of the NMR solution structure and the X-ray crystal structure of the LFB1/HNF1-homeodomains. The backbones of the mean solution structure and the crystal structure are shown as cyan and red tubes, respectively. Only residues 15 to 89 are shown, since the N and C-terminal ends 1 to 14 and 90 to 99 are disordered in solution, and have not been reported in the crystal structure. The 33 side-chains which are ''best-defined'' in the NMR structure (see footnote to Table 2) are shown in yellow for the solution structure and green for the crystal structure.

from residues 52 to 73, where the hydrogen bonds in the NMR structure identify two separate turns of 3_{10}-helix, with residues 51 to 55, and 57 to 61, respectively, which are not reported in the crystal structure. Finally, hydrogen bonds in the NMR structure indicate a regular helix for residues 82 to 89, but the corresponding hydrogen bonds are absent in the crystal structure. The structural differences indicated by different hydrogen bonding are also manifested in significant differences of the backbone dihedral angles ϕ and ψ for corresponding residues in the two structures (Figure 4).

Only part of the previously mentioned special structural motifs in the solution structure of the LFB1/HNF1-homeodomain are present also in the crystal structure. For example, in helix I the intramolecular hydrogen bond $H^{\varepsilon 1}$ Gln21–O′ Gly17 is replaced by protein–protein crystal contacts involving an intermolecular hydrogen bond $H^{\varepsilon 2}$ Gln21–$O^{\varepsilon 1}$ Gln21, and the cation–π interactions between the positively charged ξNH_3^+-group of Lys32 and the aromatic ring of Tyr28 cannot be identified either in the crystal structure. In helix II, the absence of all local structural motifs shown in Figure 5B can again be explained by protein–protein contacts in the crystal involving intermolecular hydrogen bonds H^{η} Tyr75–$O^{\varepsilon 1}$ Glu40, H^{N} Glu37–$O^{\varepsilon 1}$ Glu44, H^{N} Glu38–$O^{\varepsilon 2}$ Glu44 and $H^{\varepsilon 22}$ Gln53–O′ Glu48. The intramolecular cation–π interactions of Arg81 with the aromatic rings of Tyr28 and Trp77 are absent in the crystal structure, and the salt bridge between Gln26 and Glu45 (Figure 6) is replaced by intermolecular crystal contacts involving a hydrogen bond between the side-chains of Gln26 and Glu29, and a salt bridge between Glu45 and Arg48. The cation–π interaction of the side-chain of Arg73 with the aromatic ring of Phe14 is one of the few specific interactions in Figures 5 and 6 that can also be observed in the crystal structure. Overall, the NMR solution structure differs from the crystal structure in several aspects that are relevant for protein stability and protein folding. The eight-residue elongation of the C-terminal helix represents an environment-dependent change in local stability, and its real importance concerns the latent helix propensity for residues 82 to 89 that allows the extension of the C-terminal helix in homeodomains upon DNA-binding (Qian *et al.*, 1994; Wolberger 1996). The replacement of helix caps in the crystal structure by interprotein contacts is of general interest with regard to interpretation of structural data relating to protein folding and stability.

Comparison of the recognition helix of the LFB1/HNF1-homeodomain with those in classical homeodomains

In the initial structure determination of the *Antp* homeodomain (Qian *et al.*, 1989) the recognition helix revealed two special features: (1) Compared to prokaryotic repressor proteins the helix was elongated by two turns, which was subsequently found to be related to a different DNA-binding mode when compared to prokaryotic repressors (Otting *et al.*, 1990). (2) The recognition helix was clearly divided into a well structured N-terminal part (helix III) and a less precisely defined C-term-

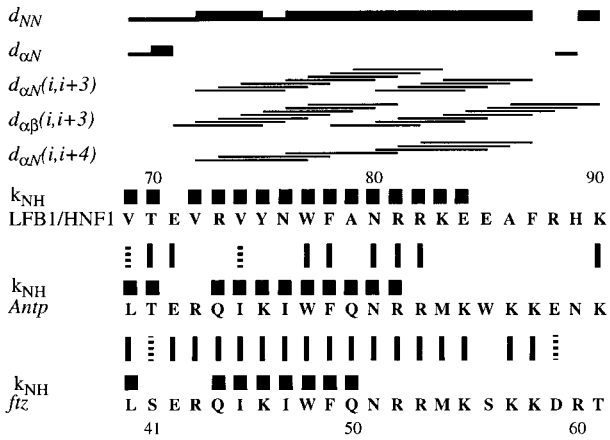

Figure 8. Amino acid sequence alignment for the ''recognition helix'' in the LFB1/HNF1, *Antp* and *ftz* homeodomains (the fragment shown forms the helices III and IV), and survey of the sequential and medium-range NOE connectivities in the segment 69–90 of the LFB1/HNF1-homeodomain. The widths of the bars for the sequential connectivities, d_{NN} and $d_{\alpha N}$, reflect the relative NOE intensities. Medium-range connectivities, $d_{\alpha N}(i,i+3)$, $d_{\alpha\beta}(i,i+3)$ and $d_{\alpha N}(i,i+4)$, are represented by lines starting and ending at the positions of the interacting residues. The numeration in the centre of the figure refers to the LFB1/HNF1-homeodomain, and the numbers at the bottom to the *Antp* and *ftz* homeodomains. In the sequence comparisons identical amino acids are connected by filled vertical bars and conservatively exchanged residues by stippled bars. Black squares above the sequences identify residues with observably slowed amide proton exchange.

inal elongation (helix IV). Subsequently, evidence was presented that helix IV was either absent or only little populated in the solution structures of the *ftz* (Qian *et al.*, 1994) and *NK-2* homeodomains (Tsao *et al.*, 1994, 1995), and a similar observation was made in the mutant *Antp (C39S, W56S)* homeodomain (unpublished). In contrast, helix IV was found to be well characterized in all presently available structures of DNA complexes with homeodomains (Wolberger, 1996; Schott, 1996). In this context, special interest in the present project was focused on the structural features of the recognition helix in the LFB1/HNF1-homeodomain.

A local superposition of the 20 DIANA conformers for the helices III and IV (Table 2, Figure 3B) shows that the entire helical segment 71–89 is locally precisely defined, which is also indicated by direct inspection of the sequential and medium-range NOEs (Figure 8). In contrast, in a global superposition of the LFB1/HNF1-homeodomain for best fit of the residues 15 to 82 the helix IV appears increasingly disordered toward the C terminus (Figure 3A), and a superposition for residues 15 to 89 gives a significantly increased r.m.s.d. value when compared to the polypeptide segment 15 to 82 (Table 2). These observations indicate that the stability of helix IV in the LFB1/HNF1-homeodomain is comparable to that in the *Antp*-homeodomain, and clearly increased stability is noted when compared to the *ftz* and *NK-2* homeodomains.

Figure 8 compares the amino acid sequences of the helices III and IV for the LFB1/HNF1, *Antp* and *ftz* homeodomains, and shows the sequential and medium-range NOE connectivities identified in LFB1/HNF1. The increased stability of the helix IV in the *Antp* homeodomain, when compared to *ftz*, is mainly due to contacts of Trp56 to helix III and the core of the protein (unpublished). In helix IV of the LFB1/HNF1 homeodomain no long-range contacts to the protein core could be identified, but it is stabilized by a salt bridge between the charged side-chains of Arg81 and Glu85 (Figure 6). As helix IV protrudes into the solvent, additional stability may be conferred by the high density of charges in the segment of residues 82 to 85. The positive and negative charges in these positions are located in the front and the back of helix IV, respectively, and may therefore be important in DNA binding. In conclusion, although there is only little sequence homology between the LFB1/HNF1 and *Antp* homeodomains in the region of helix IV, this region shows surprisingly similar structural behaviour in the two proteins. In *Antp* the helical structure is stabilized by the core-contacts of Trp56, while in LFB1/HNF1 the helical structure is maintained due to the presence of a salt bridge and a high abundance of charged residues.

Indications for the DNA-binding mode of the LFB1/HNF1-homeodomain

With regard to DNA-recognition it is of fundamental interest that the LFB1/HNF1-homeodomain contains a surprisingly similar helix–turn–helix motif to that of classical homeodomains, in spite of the 21-residue insertion in the turn. The global fold outside of this insertion is very little affected and can be closely superimposed with classical homeodomain structures. The insertion causes an elongation of the second helix, and the addition of a loop region and an extended chain running parallel to the extension of the second helix (Figure 3A). We performed preliminary experiments on a complex formed by the ^{15}N-labelled LFB1/HNF1-homeodomain and the unlabelled DNA duplex d-CCTTGGTTAATCGC · d-GCGATTAACCAAGG, which contains one LFB1/HNF1 consensus half-site (Frain *et al.*, 1989; Tomei *et al.*, 1992). Observation of the imino proton resonances of the DNA at different protein: DNA ratios during titration with the LFB1/HNF1-homeodomain showed that the intermediate spectra are superpositions of the spectra of the free DNA and the 1:1 complex, indicating that the exchange is slow on the chemical shift time scale. This result coincides with similar studies of the *Antp* system (Otting *et al.*, 1990) and shows that a detailed structure determination of the complex by NMR will be

technically feasible. In a preliminary round of data collection using the complex with the uniformly ^{15}N-labelled LFB1/HNF1-homeodomain (Schott, 1996), nearly complete ^{1}H NMR assignments could be achieved for both the protein and the DNA, and in addition to intramolecular NOEs a small number of intermolecular short distances between the DNA and the protein could be identified. Docking of the protein to the DNA with the use of these constraints indicates that the helices III and IV bind into the major groove of the DNA and that the overall orientation of the protein molecule relative to the DNA is similar to that found in other homeodomain–DNA complexes. The 21-residue insertion between the helices II and III has no direct contacts to the DNA, and is actually at a distance of about 10 Å away from it. Thus the 21-residue insertion appears not to be involved in a direct way in the DNA recognition process, but it is exposed in the structure of the binary DNA–protein complex and could readily undergo ''higher-order'' interactions, for example, with other proteins involved in the transcription complex. Similar conclusions resulted from model considerations based on the X-ray crystal structure of the LFB1/HNF1-homeodomain (Ceska *et al.*, 1993).

Materials and Methods

Protein expression and purification, and preparation of the NMR samples

Unlabelled, uniformly ^{15}N-labelled, uniformly 13C-labelled, uniformly ^{13}C,^{15}N-labelled, and fractionally (10%) ^{13}C-labelled preparations of the LFB1/HNF1-homeodomain were used in this study. The *E. coli* strain BL21(DE3)/PT7.7XL-HOM was used to express the protein, as described by Tomei *et al.* (1992). The protein was purified using reversed-phase chromatography with a linear gradient of 0% to 60% acetonitrile in H_2O. The fractions containing the protein were pooled and lyophilized. For the preparation of the ^{15}N-labelled protein, the cells were grown in ten litres of minimal medium (10 mM $(^{15}NH_4)_2SO_4$, 96 mM potassium phosphate buffer (pH 6.8), 0.5 mM Mg_2SO_4, 13 µM $FeSO_4$, 5 µM (+)-biotin, 7 µM vitamin B_1, 30 mM D(+)-glucose) at 37°C. The induction time was four hours. To prepare the ^{13}C-labelled and the double labelled protein, the cells were grown on four litres of the above minimal medium containing 5 mM [$^{13}C_6$]-D(+)-glucose in the place of the unlabelled glucose to an OA_{600} of 0.6, when the medium was exchanged against new, identical medium and the cells were induced for 22 hours at 26°C using 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). The cells were harvested and fractured in a French press at 1500 psi at 4°C. The cell fractions were removed by centrifugation, and the DNA fractions and the cell proteins were separated by polyethylene-imine precipitation and ammonium-sulphate precipitation. For the final purification step of the LFB1/HNF1 polypeptide we used a cation exchange column (heparin Sepharose CL-6B or SP-Sepharose, Pharmacia) under reducing conditions (20 mM Tris-HCl, (pH 8.8), 1 mM EDTA, 1 mM DTT, 50 mM NaCl). Fractions containing the protein were pooled, titrated to pH 4.0 and extensively dialysed against water. The pH adjustment prior to the dialysis was needed to keep the protein in the reduced state during the dialysis and to prevent precipitation. The dialysed protein was sterile filtered and lyophilized.

NMR samples were prepared by dissolving the lyophilized protein in 50 mM NaCl in either 90% H_2O/10% ^2H_2O or 100% ^2H_2O at pH 4.6. The final protein concentrations in the different NMR samples were between 3 and 5 mM. These samples were stable for several days at room temperature, but some protein precipitation was noted for all samples when recording lengthy experiments. In particular, the ^{13}C,^{15}N-labelled protein was unstable and only limited use was possible.

NMR spectroscopy

NMR measurements were performed on a Bruker AMX-600 spectrometer at 22°C. Quadrature detection was achieved using States–TPPI (Marion *et al.*, 1989). The following experiments were recorded: 3D ^{15}N-resolved-[^{1}H,^{1}H]-NOESY in H_2O solution of the ^{15}N-labelled protein (Fesik & Zuiderweg, 1988; Messerle *et al.*, 1989) ($\tau_m = 50$ ms, $t_{1max} = 28.8$ ms, $t_{2max} = 11.6$ ms, $t_{3max} = 131$ ms, time domain data size $200 \times 12 \times 1024$ complex points, total recording time 102 hours); 3D HNCA (Grzesiek & Bax, 1992) in H_2O solution of the ^{13}C,^{15}N-labelled protein ($t_{1max} = 9.6$ ms, $t_{2max} = 21.3$ ms, $t_{3max} = 65.5$ ms, time domain data size $46 \times 22 \times 512$ complex points, total recording time 59 hours); 3D HCCH-COSY (Bax *et al.*, 1990b) in ^2H_2O solution of the ^{13}C-labelled protein ($t_{1max} = 27.2$ ms, $t_{2max} = 9.6$ ms, $t_{3max} = 65.5$ ms, time domain data size $160 \times 30 \times 512$ complex data points, total recording time 84 hours); 3D HCCH-TOCSY (Bax *et al.*, 1990a) in ^2H_2O solution of the ^{13}C-labelled protein (mixing time = 19.3 ms, $t_{1max} = 35.3$ ms, $t_{2max} = 12.2$ ms, $t_{3max} = 73.8$ ms, time domain data size $128 \times 38 \times 512$ complex data points, total recording time 95 hours); 3D ^{13}C-resolved NOESY (Zuiderweg *et al.*, 1990) in ^2H_2O solution of the ^{13}C-labelled protein ($\tau_m = 60$ ms, $t_{1max} = 37.4$ ms, $t_{2max} = 9.9$ ms, $t_{3max} = 65.5$ ms, time domain data size $220 \times 31 \times 512$ complex data points, total recording time 120 hours). Prior to Fourier transformation the data sets were multiplied with phase-shifted sine-bell functions (De Marco & Wüthrich, 1976) or with Lorentz–Gauss functions (Ernst *et al.*, 1987). Linear prediction using the singular value decomposition algorithm and Fourier transformation were performed with the program PROSA (Güntert *et al.*, 1992). Coupling constants $^3J_{HN\alpha}$ were measured with a series of *J*-modulated [^{15}N,^{1}H]-COSY spectra (Neri *et al.*, 1990; Billeter *et al.*, 1992) with delays of 12.0, 26.8, 41.6, 56.4, 71.2, 86.0, 100.8 and 115.6 ms. The spectra were analysed and the NOESY cross-peaks integrated with the program XEASY (Bartels *et al.*, 1995).

Structure calculations and structure comparisons

The input for the distance geometry calculations consisted of upper distance limits derived from NOESY cross-peak intensities with the program CALIBA (Güntert *et al.*, 1991), and of dihedral angle constraints derived from combined analysis of $^3J_{HN\alpha}$ coupling constants and intraresidual and sequential NOEs using the program HABAS (Güntert *et al.*, 1989). HABAS also provided a number of stereospecific assignments of β-methylene protons. No explicit hydrogen bond constraints were added to the input data. Nearly complete assignment of the NOESY cross-peaks was achieved in

several rounds of spectral analysis and structure calculation, using the program ASNO (Güntert *et al*., 1993).

Distance geometry calculations were carried out using the program DIANA (Güntert *et al*., 1991) with the REDAC strategy as described by Güntert & Wüthrich (1991). Two REDAC cycles were used, followed by a final minimization at maximum target level employing only the constraints derived directly from the NOEs and the spin–spin coupling constants. In the final round of calculations, 50 conformers were calculated from a set of 50 randomly chosen starting conformers. Restrained energy minimization was applied to the 20 conformers with the lowest DIANA target function values, using the AMBER force field (Weiner *et al*., 1986) as implemented in the program OPAL (Luginbühl *et al*., 1996). The pseudo-potentials for NMR constraints were adjusted such that violations of 0.1 Å for distance constraints and 2.5° for dihedral angle constraints corresponded to $k_BT/2$ at room temperature. The ω angles of all peptide bonds were constrained to 180(± 20)°. The energy minimization was carried out in a shell of water molecules with a minimal thickness of 6.0 Å. A total of 1500 steps of conjugate gradient minimization were performed for each conformer. The resulting 20 energy-minimized conformers are used to represent the solution structure of the LFB1/HNF1 homeodomain.

Superposition of multiple conformers was performed with the program XAM (Xia, 1992). For visual inspection of the structures and for the preparation of colour views we used the program MOLMOL (Koradi *et al*., 1996). Global superposition and pairwise r.m.s.d. values for various subsets of atoms were computed as described by McLachlan (1979), and displacements were calculated according to Billeter *et al*. (1989).

The coordinates have been deposited with the Brookhaven Protein Data Bank; accession number 2lfb.

Acknowledgements

We thank R. Cortese, R. de Francesco and L. Tomei for valuable advice on the biochemical aspects of the project. Financial support was obtained from the Schweizerischer Nationalfonds (project 31.32033.91). We thank Mrs E. Ulrich for the careful processing of the text.

References

Bartels, C., Xia, T., Billeter, M., Güntert, P. & Wüthrich, K. (1995). The Program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR,* **6**, 1–10.

Bax, A., Clore, G. M. & Gronenborn, A. M. (1990a). ^{1}H–^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *J. Magn. Reson.* **88**, 425–431.

Bax, A., Clore, M. G., Driscoll, P. C., Gronenborn, A. M., Ikura, M. & Kay, L. (1990b). Practical aspects of proton–carbon–carbon–proton three-dimensional correlation spectroscopy of ^{13}C-labelled proteins. *J. Magn. Reson.* **87**, 620–627.

Billeter, M., Kline, A. D., Braun, W., Huber, R. & Wüthrich, K. (1989). Comparisons of the high-resolution structure of the α-amylase inhibitor Tendamistat determined by nuclear magnetic resonance in solution and by X-ray diffraction in single crystals. *J. Mol. Biol.* **206**, 677–687.

Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. & Wüthrich, K. (1990). Determination of the three-dimensional structure of the *Antennapedia* homeodomain from *Drosophila* in solution by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **214**, 183–197.

Billeter, M., Neri, D., Otting, G., Qian, Y. Q. & Wüthrich, K. (1992). Precise vicinal coupling constants in proteins from non-linear fits of J-modulated [^{15}N,^{1}H]-COSY experiments. *J. Biomol. NMR,* **2**, 257–274.

Ceska, T. A., Lamers, M., Monaci, P., Nicosia, A., Cortese, R. & Suck, D. (1993). The X-ray structure of an atypical homeodomain present in the rat liver transcription factor LFB1/HNF1 and implications for DNA binding. *EMBO J.* **12**, 1805–1810.

De Marco, A. & Wüthrich, K. (1976). Digital filtering with a sinusoidal window function: an alternative technique for resolution enhancement in FT NMR. *J. Magn. Reson.* **24**, 201–204.

DeSimone, V. & Cortese, R. (1991). Transcriptional regulation of liver-specific gene expression. *Curr. Opin. Cell Biol.* **3**, 960–965.

Dougherty, D. A. (1996). Cation–π interactions in chemistry and biology: a new view of benzene, Phe, Tyr and Trp. *Science,* **271**, 163–168.

Ernst, R. R., Bodenhausen, G. & Wokaun, A. (1987). *The Principles of Nuclear Magnetic Resonance in One and Two Dimensions,* Clarendon, Oxford.

Fesik, S. W. & Zuiderweg, E. R. P. (1988). Heteronuclear three-dimensional NMR spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra. *J. Magn. Reson.* **78**, 588–593.

Frain, M., Swart, G., Monaci, P., Nicosia, A., Stämpfli, S., Frank, R. & Cortese, R. (1989). The liver-specific transcription factor LF-B1 contains a highly diverged homeobox DNA binding domain. *Cell,* **59**, 145–157.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K, Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G. & Wüthrich, K. (1994). Homeodomain-DNA recognition. *Cell,* **78**, 211–223.

Grzesiek, S. & Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96**, 432–440.

Güntert, P. & Wüthrich, K. (1991). Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *J. Biomol. NMR,* **1**, 447–456.

Güntert, P., Braun, W., Billeter, M. & Wüthrich, K. (1989). Automated stereospecific ^{1}H NMR assignments and their impact on the precision of protein structure determinations in solution. *J. Am. Chem. Soc.* **111**, 3997–4004.

Güntert, P., Braun, W. & Wüthrich, K. (1991). Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217**, 517–530.

Güntert, P., Dötsch, V., Wider, G. & Wüthrich, K. (1992). Processing of multi-dimensional NMR data with the new software PROSA. *J. Biomol. NMR,* **2**, 619–629.

Güntert, P., Berndt, K. D. & Wüthrich, K. (1993). The program ASNO for computer-supported collection of NOE upper distance constraints as input for pro-

tein structure determination. *J. Biomol. NMR*, **3**, 601–606.
Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.
Leiting, B., De Francesco, R., Tomei, L., Cortese, R., Otting, G. & Wüthrich, K. (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. *EMBO J.* **12**, 1797–1803.
Luginbühl, P., Güntert, P., Billeter, M. & Wüthrich, K. (1996). The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. *J. Biomol. NMR*, **8**, 136–146.
Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989). Rapid recording of 2D NMR spectra without phase cycling: application to the study of hydrogen exchange in proteins. *J. Magn. Reson.* **85**, 393–399.
McLachlan, A. O. (1979). Gene duplication in the structural evolution of chymotrypsin. *J. Mol. Biol.* **128**, 48–79.
Mendel, D. B. & Crabtree, G. R. (1991). HNF-1, a member of a novel class of dimerizing homeodomain proteins. *J. Biol. Chem.* **266 (2)**, 677–680.
Messerle, B. A., Wider, G., Otting, G., Weber, C. & Wüthrich, K. (1989). Solvent suppression using a spin lock in 2D and 3D NMR spectroscopy with H_2.O solutions. *J. Magn. Res.* **85**, 608–613.
Neri, D., Szyperski, T., Otting, G., Senn, H. & Wüthrich, K. (1989). Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically-directed fractional ^{13}C labeling. *Biochemistry*, **28**, 7510–7516.
Neri, D., Otting, G. & Wüthrich, K. (1990). New nuclear magnetic resonance experiment for measurements of the vicinal coupling constants $^3J_{HN\alpha}$ in proteins. *J. Am. Chem. Soc.* **112**, 3663–3665.
Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. & Wüthrich, K. (1990). Protein–DNA contacts in the structure of a homeodomain–DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* **9**, 3085–3092.
Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, J. W. & Wüthrich, K. (1989). The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell*, **59**, 573–580.
Qian, Y. Q., Furukubo-Tokunaga, K., Resendez-Perez, D., Müller, M., Gehring, W. J. & Wüthrich, K. (1994). Nuclear magnetic resonance solution structure of the *fushi tarazu* homeodomain from *Drosophila* and comparison with the *Antennapedia* homeodomain. *J. Mol. Biol.* **238**, 333–345.
Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of a helices. *Science*, **240**, 1648–1652.
Schott, O. (1996). Ermittlung der NMR Lösungsstrukturen der LF-B1 und *fushi tarazu* Homöodomänen sowie deren Homöodomänen-DNS Komplexe. Ph.D. thesis Nr. 11803. ETH, Zürich, Switzerland.
Scott, M. P., Tamkun, J. W. & Hartzell III, G. W. (1989). The structure and function of the homeodomain. *BBA Rev. Cancer*, **989**, 25–49.
Senn, H., Werner, B., Messerle, B. A., Weber, C., Traber, R. & Wüthrich, K. (1989). Stereospecific assignment of the methyl ^1H NMR lines of valine and leucine in polypeptides by non-random ^{13}C labelling. *FEBS Letters*, **249**, 113–118.
Tomei, L., Cortese, R. & De Francesco, R. (1992). A *POU-A* related region dictates the specificity of *LFB1/HNF1* by orienting the two XL-homeodomains in the dimer. *EMBO J.* **11**, 4119–4129.
Tsao, D. H. H., Gruschus, J. M., Wang, L.-H., Nirenberg, M. & Ferretti, J. A. (1994). Elongation of helix III of the *NK-2* homeodomain upon binding to DNA: a secondary structure study by NMR. *Biochemistry*, **33**, 15053–15060.
Tsao, D. H. H., Gruschus, J. M., Wang, L., Nirenberg, M. & Ferretti, J. A. (1995). The three-dimensional solution structure of the *NK-2* homeodomain from *Drosophila*. *J. Mol. Biol.* **251**, 297–307.
Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986). An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* **7**, 230–252.
Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, J. H., Oldfield, E., Markley, J. L. & Sykes, B. D. (1995). ^1H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR*, **6**, 135–140.
Wolberger, C. (1996). Homeodomain interactions. *Curr. Opin. Struct. Biol.* **6**, 62–68.
Wüthrich, K. (1986). In *NMR of Proteins and Nucleic Acids*, Wiley, New York.
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.
Zuiderweg, E. R. P., McIntosh, L. P., Dahlquist, F. W. & Fesik, W. S. (1990). Three-dimensional ^{13}C-resolved proton NOE spectroscopy of uniformly ^{13}C-labelled proteins for the *NMR* assignment and structure determination of larger molecules. *J. Magn. Reson.* **86**, 210–216.

Edited by P. E. Wright

(Received 30 October 1996; received in revised form 30 December 1996; accepted 7 January 1997)

http://www.hbuk.co.uk/jmb

Supplementary material comprising three tables, is available from JMB Online.