JMB

Three-dimensional Structure of the DNA-binding Domain of the Fructose Repressor from *Escherichia coli* by ^{1}H and ^{15}N NMR

François Penin*, Christophe Geourjon, Roland Montserret
Anja Böckmann, Anne Lesage, Yin Shan Yang
Christelle Bonod-Bidaud, Jean-Claude Cortay, Didier Nègre
Alain J. Cozzone and Gilbert Deléage

Institut de Biologie et Chimie des Protéines, Centre National de la Recherche Scientifique 7 passage du Vercors 69367 Lyon, France

FruR is an *Escherichia coli* transcriptional regulator that belongs to the LacI DNA-binding protein family. By using ^{1}H and ^{15}N NMR spectroscopy, we have determined the three-dimensional solution structure of the FruR N-terminal DNA-binding domain consisting of 57 amino acid residues. A total of 809 NMR-derived distances and 54 dihedral angle constraints have been used for molecular modelling with the X-PLOR program. The resulting set of calculated structures presents an average root-mean-square deviation of 0.37 Å at the main-chain level for the first 47 residues. This highly defined N-terminal part of the structure reveals a similar topology for the three α-helices when compared to the 3D structures of LacI and PurR counterparts. The most striking difference lies in the connection between helix II and helix III, in which three additional residues are present in FruR. This connecting segment is well structured and contains a type III turn. Apart from hydrophobic interactions of non-polar residues with the core of the domain, this connecting segment is stabilised by several hydrogen bonds and by the aromatic ring stacking between Tyr19 of helix II and Tyr28 of the turn. The region containing the putative ''hinge helix'' (helix IV), that has been described in PurR-DNA complex to make specific base contacts in the minor groove of DNA, is unfolded. Examination of hydrogen bonds highlights the importance of homologous residues that seem to be conserved for their ability to fulfil helix N and C-capping roles in the LacI repressor family.

© 1997 Academic Press Limited

Keywords: fru repressor; heteronuclear NMR spectroscopy; protein structure; DNA-binding domain; helix-turn-helix motif

*Corresponding author

Introduction

The fructose repressor FruR is a tetrameric protein of 334 amino acid residues per subunit that controls transcription initiation at a number of target promoters of genes or operons involved in carbon and energy metabolism in enteric bacteria (Ramseier *et al.*, 1993; for a review, see Saier *et al.*, 1996). On the basis of sequence similarities with other bacterial regulatory proteins, FruR has been classified in the LacI and GalR superfamily of repressors (Weickert & Adhya, 1992), which contains at least 36 members to date according to the Prosite dictionary of signatures and potential sites (Bairoch, 1993). The FruR monomer has been shown to be organised in two functional domains, as described for the other members of the family: (1) the N-terminal DNA-binding domain (DBD) of approximately 60 residues, which exhibits a putative structural helix-turn-helix (HTH) motif and is responsible for the binding of FruR on its operator; (2) the larger C-terminal portion, which displays inducer-binding properties and subunit interaction (Cortay *et al.*, 1994; Scarabel *et al.*, 1995).

The HTH DNA-binding motif has been described in many prokaryotic and eukaryotic regulatory proteins (Brennan & Matthews, 1989; Brennan, 1992; Wintjens & Rooman, 1996). Its high degree of conservation in the LacI family, as well as the significant sequence identity of their 16 to 18 bp operator sites in this family, suggest that FruR binds to its operator in a similar way to that of all members of the family (Schumacher *et al.*, 1994; Nègre *et al.*, 1996). The three-dimensional structures of two LacI family members are now available: the lactose repressor, LacI, and the purine repressor, PurR. The structure of LacI DBD has been obtained from NMR studies, either free (Kaptein *et al.*, 1985; de Vlieg *et al.*, 1988; Slijper *et al.*, 1996) or complexed to its DNA operator (Lamerichs *et al.*, 1989; Chuprina *et al.*, 1993). More recently, the crystal structure of the whole tetrameric Lac repressor complexed with two 21 bp symmetric operator DNA has been solved (Lewis *et al.*, 1996). The PurR dimer structure complexed to its operator has been determined by crystallography (Schumacher *et al.*, 1994), while free PurR DBD structure (residues 1 to 59) was obtained by NMR (Nagadoi *et al.*, 1995). Free or complexed LacI and PurR DBD exhibit a very similar topology for their HTH motif (α-helices I and II) and their α-helix III, which stabilises the HTH structure. For the free LacI domain, the connecting segment between helix II and helix III is slightly disordered and a significant conformational change occurs in this region upon binding on the DNA operator (Slijper *et al.*, 1996). Conversely, the corresponding region in the free PurR DBD is rather well structured, and the complexation with DNA does not induce a large conformational change. For both proteins, specific contacts have been identified between the bases of their operator sites and the amino acid residues of the connecting segment, in addition to those observed in the recognition helix II (Lamerichs *et al.*, 1989, 1990; Chuprina *et al.*, 1993; Schumacher *et al.*, 1994). Finally, the study of PurR complexed to its operator has highlighted specific contacts between bases in the DNA minor groove and residues of a fourth α-helix, namely the "hinge helix" (residues 48 to 56, Schumacher *et al.*, 1994). The insertion of two symetrically related hinge helices into the minor groove kinks the central CG base-pairs of the PurR operator site. In contrast, the hinge helix has been shown to be unfolded in free PurR DBD (Nagadoi *et al.*, 1995) and in free LacI DBD (Spronk *et al.*, 1996). The coil to hinge helix transition results not only from specific DNA binding but requires protein-protein interaction between the two related hinge helices, as shown recently for LacI DBD by Spronk *et al.* (1996).

From these studies, some rules of DNA base recognition by key amino acid residues have been deduced that fit well for several other members of the LacI family (Schumacher *et al.*, 1994), including FruR (Nègre *et al.*, 1996). However, rather large differences in the amino acid sequences and operator specificities are observed among the 36 putative members of the LacI family, and detailed 3D structural analyses of several other members are required to obtain a better knowledge of the DNA binding specificities in the entire family. In this context, FruR is particularly interesting, since its DBD shows a good similarity to the LacI and PurR counterparts, but exhibits differences obviously related to the specificities of the operators (Nègre *et al.*, 1996). Besides the recognition helix II, the main difference lies in the loop between helices II and III in which three additional amino acid residues are present. Thus, the 3D structure analysis of FruR DBD alone or complexed to its operator would help to specify the basic rules of DNA recognition for the LacI family, as well as in the other DNA-binding protein families, and should extend our knowledge of the specific and reversible protein-DNA binding mechanisms. We report here the first step towards this goal with the solution structure of the FruR DNA-binding domain deduced from ^{1}H and ^{15}N NMR experiments. Besides the identification of residues likely to be involved in DNA recognition, a detailed analysis of helices end capping highlights the fact that consensus residues are conserved in the whole LacI family to ensure the stabilisation of helices I, II and III by current N and C-capping motifs. Conversely, no obvious stabilisation by N and/or C-capping occurs for the hinge helix. This is in agreement with the fact that this helix appears to be folded only when LacI family members are oligomerised and complexed with their DNA operator. Finally, concerning the connecting segment between helices II and III that contains three additional residues when compared to LacI and PurR, we propose that its particular conformation may be related to the property of FruR DBD to bind to several natural DNA operators of low palindromy.

Results and Discussion

Overproduction and purification of FruR(1-57)*

We have shown previously that FruR DBD could be overproduced in *E. coli* as a functional domain able to bind strongly and specifically to its operator (FruR(2-57), Scarabel *et al.*, 1995). This domain had been overproduced in fusion with glutathione-S-transferase and obtained pure after cleavage of the fusion protein by thrombin. However, due to the difficulty in removing the protease completely, this preparation was unstable with time, yielding NMR spectra of insufficient quality for an accurate determination of the 3D structure of FruR DBD. To overcome this problem, the first 57 N-terminal residues of FruR were cloned in fusion with a 6xHis tag in pCB4 plasmid that allowed large-scale production of the recombinant protein. The fusion protein, termed FruR(1-57)*, exhibited a C-terminal LQHHHHHH sequence extension that permitted the simple purification of either normal or ^{15}N-labelled DBD on a Ni^{2+} affinity column followed by cation-exchange chromatography (see Materials and Methods). The main features of the final preparation of FruR(1-57)* were the following: (i) purity was over 99% as checked by SDS-PAGE and reversed phase HPLC; (ii) the yield was 25 mg of soluble protein per litre of growth medium; (iii) N-terminal sequencing yielded the sequence NH_2-Met-Lys-Leu-Asp-Glu, i.e., a protein without methionine processing; (iv) electrospray mass spectroscopy analysis gave a molecular mass of 7380.4($\pm$1.5) Da (calculated value of 7379.5 Da). The molecular mass of the ^{15}N-labelled domain was 7481.2($\pm$1.5) Da, indicating that ^{15}N enrichment was around 96%; (v) the circular dichroism spectrum was typical of an α-helix protein; the content in α-helix deduced from CD is in agreement with that calculated from the NMR 3D structure (see below); (vi) specificity for the *ace* operator DNA was qualitatively indistinguishable from that of the complete FruR protein or of the recombinant FruR(2-57), as checked by gel retardation assay and DNA methylation-protection experiments (Scarabel *et al.*, 1995). Moreover, preliminary NMR experiments (not shown) indicated that ^{15}N-labelled FruR(1-57)* forms a tight complex with a synthetic 14 bp DNA that mimics the consensus half-site operator of FruR (Nègre *et al.*, 1996). These results clearly demonstrate that FruR(1-57)* exhibits all the expected structural and functional properties of the FruR DNA-binding domain, and that the C-terminal LQHHHHHH extension does not interfere in the complexation with DNA. On the other hand, the use of pCB4 overproduction plasmid allows a simple, rapid and high-yield preparation of FruR(1-57)*. This protein exhibits a high level of stability and can be stored frozen without losing either its biological activity or the resolution of NMR spectra. The high quality of NMR spectra presented below was largely due to this high degree of purity and homogeneity.

Resonance assignment

FruR residues (1-57)*

The 2D and 3D homo- and heteronuclear NMR experiments carried out with FruR(1-57)* samples at pH 5.9 and 20°C yielded well-resolved spectra, as illustrated by the extract of 2D homonuclear TOCSY (Figure 1). Identification of the amino acid spin systems and sequential assignment were achieved by using mainly the basic strategy described by Wüthrich (1986) with TOCSY, DQF-COSY and NOESY spectra collected at various mixing times and temperatures. Although FruR(1-57)* yielded proton NMR spectra with rather narrow line-widths, the predominantly α-helical nature of the protein resulted in a limited chemical shift dispersion. The use of uniformly ^{15}N-labelled FruR(1-57)* and heteronuclear ^{15}N-^{1}H NMR experiments (HSQC and 3D HMQC-TOCSY and 3D NOESY-HMQC) allowed the confirmation of the sequential assignment and the resolution of ambiguous assignment. The strategy of 3D heteronuclear spectral analysis consisted in locating previously identified spin systems in the slices along the ^{15}N dimension, before proceeding with the usual sequential walk by identifying NHi + 1-(NH, H$^\alpha$, H$^\beta$)i cross-peaks on NOESY-HMQC slices (data not shown). Figure 2 shows the amide region of the ^{15}N-^{1}H HSQC spectrum of uniformly labelled FruR(1-57)*. The dispersion in the ^{15}N dimension is sufficient to resolve most resonances, although some overlapping pairs of ^{15}N-^{1}H cross-peaks are observed: (Thr34, Lys37), (Lys24, Asn50) and (Glu44, Asn46). All of the expected ^{15}N-^{1}H backbone amide correlations for residues 2 to 60 are observed, although Lys2 and Leu3 residues give low-intensity correlations because of rapid exchange of their amide protons with water, as shown by Böckmann *et al.* (1996). For the same reason, ^{15}N-^{1}H correlations for histidine residues 61 to 65 at 500 MHz were detected only weakly in HSQC and localised in the dotted area in Figure 2(b) (see below). On the contrary, the five pairs of side-chain amide correlations of Asn22, Gln27, Asn46, Asn50 and Gln59 are clearly observable. Upfield from the amide protons are five cross-peaks for the NH$^\varepsilon$ protons of Arg8, Arg14, Arg29, Arg43 and Arg57 (Figure 2(a)), together with a broad cross-peak of the N$^\eta$ protons of these residues (not shown).

The combined analysis of homo- and heteronuclear spectra allowed the complete ^{1}H and ^{15}N assignment for residues 1 to 60 of FruR(1-57)*. Assignments of amino acid spin systems are shown in Figure 1. Stereospecific assignments were achieved for several β-methylene and γ-methyl protons (see below).

Concerning the C-terminal LQHHHHHH extension of FruR(1-57)*, the presence of sequential NOEs allowed the assignment of Leu58, Gln59 and His60 unambiguously (Figure 1). For the 61 to 65 histidine residues, most of the NH-H$^\alpha$ and NH-

Figure 1. NH(horizontal axis)-$C_{aliphatic}$H(vertical axis) regions of the 500 MHz TOCSY spectrum for FruR(1-57)* 3 mM in 10 mM sodium phosphate buffer (pH 5.9) containing 50 mM NaCl, 70 mM NH_4Cl, 5% ^2H_2O and 0.05% sodium azide at 20°C (mixing times 80 ms). The cross-peaks between NH and its own H^α and side-chain protons are indicated by the corresponding residue number. Cross-peaks of unassigned His residues (61 to 65) are labelled H*.

$H^{\beta\beta'}$ correlations gave low-intensity cross-peaks on TOCSY spectra (labelled H* in Figure 1). Moreover, only very weak ^{15}N-^1H correlations were observed for these residues in HSQC spectra. This is due to rapid exchange of backbone amide protons of His61 to His65 with water protons in the millisecond time-scale, as estimated recently by using the water selective NOESY-HSQC spectrum of ^{15}N-labelled FruR(1-57)* (Böckmann *et al.*, 1996). While His61 to His65 amide protons gave weak signals, AMX spin systems of individual His residues were detected. However, sequential assignment was not possible because of the absence of any observable NOE correlation. Moreover, the H^α and $H^{\beta\beta'}$ chemical shifts of His61 to His65 were typical of the random coil configuration (Wishart *et al.*, 1991). In addition, the H^α-$H^{\beta\beta'}$ correlations gave much thinner cross-peaks than the other residues of FruR(1-57)*, indicating a higher relaxation time for the His61 to His65 residues and therefore a high local mobility. All these data clearly indicate that the sequence encompassing the His61 to His65 residues does not exhibit any stable conformation. Besides, the absence of any observable NOE between the LQHHHHHH extension and the rest of the domain indicates that no interaction occurs between them, and that it does not induce any confor-

mational change in the DBD domain of FruR. This latter conclusion is supported by the fact that FruR(2-57) exhibited a NOESY spectrum almost superimposable with that of FruR(1-57)*, showing that the DBD domain of FruR presents the same conformation in both cases. Hence, implication of the LQHHHHHH extension was not further considered in the determination of the structure of the FruR DBD.

Secondary structures

Figure 3 shows the sequence of the first 60 residues of FruR(1-57)* and an overview of short-range and medium-range NOEs, the ^3J NHH$^\alpha$ coupling constants, and the chemical shift analysis for H$^\alpha$ protons. Long stretches of $d_{\alpha N}$ (i, $i + 3$), $d_{\alpha\beta}$ (i, $i + 3$) and $d_{\alpha N}$ (i, $i + 4$) connectivities characteristic of α-helices (Wüthrich, 1986) are observed throughout the 2 to 47 region. The α-helical regions were further characterised by the presence of (i) strong d_{NN} and weak $d_{\alpha N}$ connectivities, (ii) weak coupling constants ^3J NHH$^\alpha$ (<6 Hz), and (iii) the negative chemical shift index (CSI) for the H$^\alpha$ proton, which is typical of an α-helix (Wishart *et al.*, 1992). It should be pointed out that, because of chemical shift degeneracies of C$^\alpha$ protons in the he-

Figure 2. Extract of the ^{1}H-^{15}N gHSQC spectrum at 500 MHz of uniformly ^{15}N-labelled 2 mM FruR(1-57)* at 20°C and pH 5.9 (medium as Figure 1). (a) NH$^\varepsilon$ of Arg residues; (b) amide protons. The cross-peaks are labelled according to the residue type and number. Horizontal lines connect cross-peaks for the NH$_2$ groups of Asn and Gln. The grey box indicates the region of NH-H$^\alpha$ cross-peaks of unassigned His residues (61 to 65). These weak cross-peaks were detectable at a lower contour level but clearly visible at 600 MHz (see Böckmann *et al.*, 1996).

lical regions, the absence of a particular connectivity is often due to the fact that the corresponding NOE could not be unambiguously identified. A first α-helix is clearly identified in the 3 to 10 region. A second helix can be detected in the 14 to 22 region, although not all α-helical medium-range NOEs are found. The 24 to 29 region presents most criteria of helical fold, but residues 28 to 30 exhibit large ^3J NHH$^\alpha$ coupling constants and numerous d_{NN} ($i, i + 2$) NOEs. These data provide support for a structure that could be either a short irregular helix or a turn. A third α-helix can be detected in the 32 to 45 region, although several α-helical medium-range NOEs are missing, essentially due to the lack of unambiguous attribution of H$^\alpha$ NOE connectivities for Lys37, Ala40 and Arg43 (H$^\alpha$ protons of these residues resonate at the same frequency). Of interest, the amide proton of residues preceding the three α-helices (i.e., Lys2 for helix I, Arg14 for helix II and Ser31-Asp32 for helix III) appear downfield shifted when compared to other amide proton chemical shifts (see Figure 1). This downfield shift can be attributed to helix dipole effects

(Wishart *et al.*, 1991). Finally, the distribution of secondary structure elements fits quite well with the presence of slowly exchanging amide protons, as reported elsewhere (Böckmann *et al.*, 1996), indicating that these amide protons are involved in hydrogen bonds that stabilise the regular secondary structures.

In contrast to the highly structured 1 to 47 region, the 48 to 59 region appears to be flexibly disordered, since neither medium-range nor long-range NOEs were observed (Figures 3 and 4). In addition, the corresponding residues give strong TOCSY cross-peaks typical of an unfolded segment (Figure 1) and their amide protons are in rather rapid exchange with water protons (Böckmann *et al.*, 1996), indicating the absence of any stable structure. There is thus no indication supporting the formation of the putative helix IV described for PurR when complexed with its operator (the "hinge helix", Schumacher *et al.*, 1994) in this region. One could wonder whether the presence of the LQHHHHHH extension of FruR(1-57)* prevented the hinge helix formation. In fact, none of the other overproduced FruR DBD fragments without LQHHHHHH extension (i.e. FruR(2-57), and FruR(1-63)), presented any indication of the hinge helix formation. This result is in agreement with the data of Nagadoi *et al.* (1995) and Spronk *et al.* (1996), who reported that the hinge helix is unfolded in free PurR and LacI DBD, respectively.

The 3D structure calculation

In a first approach, only unambiguous NOE constraints were used for structure calculation. An improvement of the constraints set was achieved through a round of NOE back-calculation that allowed the validation of all NOE-derived constraints and the unambiguous identification of several constraints. Finally, a total set of 809 interproton constraints, including 148 long-range, 187 medium-range, 242 sequential and 232 intraresidue interproton distance constraints was used for molecular modelling (Table 1). Figure 4(a) gives the distribution of NOE-derived interproton distance constraints for each residue introduced as input in the structure calculation. In addition, 54 dihedral angles were used, including 39 φ angles deduced from the ^3J NHH$^\alpha$ coupling constants (Figure 3) and 15 χ$_1$ angles deduced from stereospecific assignments obtained for five β-methylene proton pairs (Asp4, Tyr19, Asn22, Tyr28 and His45) for the two isoleucine and the three threonine residues, and for five valine residues (residues 20, 35, 38, 41 and 42). No additional hydrogen bond constraint was introduced into the calculation. Structure calculations were performed with X-PLOR (Brünger, 1992). All the structures generated were accepted, indicating no violation of NOE distances (no deviation >0.5 Å) and of dihedral constraints (no dihedral deviation >5°). In other words, all structures fully satisfied the experimental NMR data and fell into only one family of structure (see
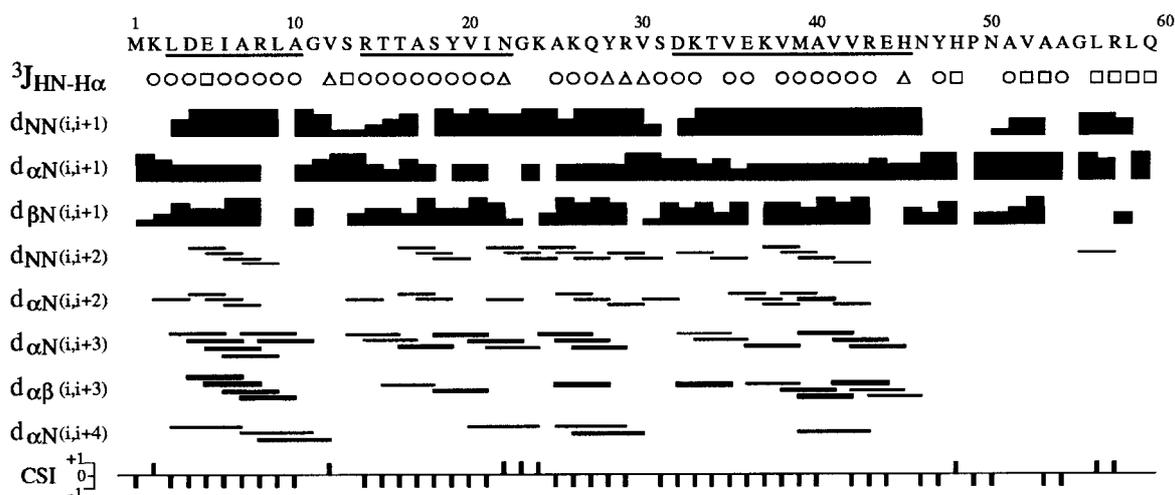
Figure 3. Summary of short and medium-range NOEs observed along the first 60 residues of FruR(1-57)*. The thickness of the connecting bars corresponds to the intensity of the NOEs. Coupling constants ^3J NHH$^\alpha$ were obtained from J-HMQC and HNHA experiments (Kay & Bax, 1990; Vuister & Bax, 1993, respectively): circles, ^3J NHH$^\alpha$ < 6 Hz; squares, $(6 \leqslant ^3J$ NHH$^\alpha \leqslant 9)$Hz; triangles, ^3J NHH$^\alpha$ > 9 Hz. CSI, chemical shift index for ^{1}H$^\alpha$ (Wishart et al., 1992). The α-helical regions are indicated by a line under the amino acid sequence.

below). A final set of 34 best structures from 50 was selected by restricting the cut-off for NOE distances violation to <0.4 Å.

Structure analysis of FruR(1-57)*

The stereoview of the final set of 34 fitted structures, without the C-terminal region of FruR(1-57)*, is presented in Figure 5. Indeed, the 49 to 60 region is highly disordered, due to the deficiency of NOE constraints in this segment for which only sequential correlations were observed (Figures 3 and 4(a)). In contrast, Figure 5 shows that the N-terminal part (residues 1 to 48) is very well defined, reflecting the rather high number of medium and long-range NOE constraints (Figure 4(a)). The r.m.s. deviation between the 34 final structures from residues 1 to 47 is 0.37 Å for the backbone atoms and 1.24 Å when all atoms are considered (Table 1). These data indicate the high resolution of the backbone structure in the 1 to 47 region, and particularly for the regular secondary structured segments that clearly correspond to minima in the r.m.s.d., whereas random segments mostly correspond to maxima in the plot of global and local r.m.s.d. *versus* residue number presented in Figure 4(b) and (c). In particular, a r.m.s.d. as low as 0.3 Å was calculated for the three α-helix elements (Table 1). These results show the good convergence of the calculated structures and the presence of only one family of structures. The stereochemical properties of the backbone dihedral angles are provided in Figure 6 as (ϕ/ψ) Ramachandran plots for the 1 to 47 region of the 34 final structures. About 95% of residues are located in the most favoured or allowed regions, while only 0.5% of Φ/Ψ angles are in disallowed regions (Table 1). The right-handed α-helix area is the most populated region, while Gly11 and 23 (filled triangles) and Asn46 (filled

squares) exhibit positive Φ angles (Figure 6). This is related to the fact that these residues occupy the C' position of the C-terminal ends of helix I, II and III, respectively (see helix capping, below). The validity of the FruR(1-57)* structure was confirmed by the rather low energy found for the calculated molecules (−228 kcal mol^{-1}) and the few deviations from ideal covalent geometry (Table 1). Finally, the removal of NMR constraints in an additional minimisation procedure does not lead to important structural change (not shown). This clearly indicates that the final structures are stable even without distance and dihedral angle constraints.

As expected from sequence homology, the FruR DBD folds in a helix-turn-helix motif (residues 3 to 22) linked to a third helix through a long connecting segment spanning from residue 23 to 31 (Figure 5). According to PROCHECK analysis (Laskowski et al., 1993), the three α-helices extend over residues 3 to 10 (helix I), 14 to 22 (helix II) and 32 to 45 (helix III), and a turn is formed by residues 25 to 28. The structure of the three helices is maintained by a hydrophobic core formed by residues Leu3, Ile6, Ala7, Ala10, Val12, Ala17, Val20, Ile21, Ala25, Val30, Val38, Val41 and Val 42. In addition, the structure is stabilised by a network of hydrogen bonds. The characteristic helical pattern of regular $i,i+4$ hydrogen bonds is observed for most structures. Helices I and II are linked by a left-handed turn (residues 10 to 13) typical of the HTH motif. The connecting segment between helices II and III contains a hydrogen-bonded turn region extending from residues 24 to 31. It is highly structured and stabilised by three hydrogen bonds found in all structures, as illustrated in Figure 7: (i) one of them is found between Lys24-NH of helices II and Tyr19-O' of the connecting segment and participates in the C-capping of helix II (see below); (ii) the two others are found between central

Table 1. Statistics of the 34 final simulated annealing structures of FruR (1-57)*

A. *Constraints used*

Distance restraints

Intra-residue	232
Sequential ($|i - j| = 1$)	242
Medium range ($|i - j| \leqslant 4$)	187
Long range ($|i - j| > 4$)	148
Total distance restraints	809

Dihedral angle constraints

ϕ angles	39
χ_1 angles	15

B. *Statistics for 34 final X-PLOR structures*

X-PLOR energy (kcal mol^{-1})	-228 ± 11

NOE violations

Number >0.4 Å	none
R.m.s. deviation (Å)	0.11 ± 0.004

Dihedral violations

Number >5°	none
R.m.s. deviation (deg.)	0.8 ± 0.26

Deviation from idealized covalent geometry

Angles (deg.)	1.42 ± 0.05
Impropers (deg.)	0.23 ± 0.01
Bonds (Å)	0.005 ± 0.0002

R.m.s. deviation (Å)

Backbone (C′, C$^\alpha$, N)	residues 1-47	0.37
	helices	0.30
All heavy atoms	residues 1-47	1.24
	helices	1.18

Ramachandran data[a]

Residues in most-favoured regions (%)	77.0 ± 3.4
Residues in allowed regions (%)	17.9 ± 4.1
Residues in generously allowed regions (%)	4.6 ± 2.6
Residues in disallowed regions (%)	0.5 ± 0.8

[a] From PROCHECK (Laskowski *et al.*, 1993), excluding glycine residues.

Figure 4. Structural characterisation of FruR(1-57)*: number of NOEs and backbone r.m.s. deviations as a function of residue number. (a) Histogram showing the number of sequential $i,i+1$ constraints (black boxes), medium-range constraints with $1 < |i - j| \leqslant 4$ (white boxes), and long-range constraints defined as $|i - j| > 4$ (grey boxes). Each inter-residue NOE appeared twice, once for each of the two interacting residues. (b) Histogram of the atomic r.m.s.d. for backbone heavy-atoms (N, C$^\alpha$,C′) for each residue in the final set of 34 structures. (c) Histogram of three-residue averaged r.m.s.d. for backbone heavy-atoms from the final set of 34 structures.

residues of the connecting segment: Tyr28-NH → Lys24-O′ and Val30-NH → Ala25-O′. Main-chain dihedral angles of residues 25 to 28 fit the description of a type III turn (Chou & Fasman, 1977) and was named turn II (turn I being the turn of the HTH motif). The structure of the connecting segment (Figure 7) is also stabilised by the aromatic ring stacking between Tyr19 and Tyr28. It is worth mentioning that a similar tyrosine stacking has been described in LacI DBD (Tyr7 and Tyr17) as a stabilising element between helices I and II (Kaptein *et al.*, 1985). Finally, inspection of the 3D structure in this region indicates a close proximity between the side-chain amino group of Lys26 and the side-chain carboxyl group of Asp32, suggesting the presence of a salt bridge. This bridge may also participate in the stabilisation of the connecting fragment.

Helix capping

The three helices of FruR DBD are surrounded at their N and C-terminal ends by residues compatible with the formation of helix caps that have been shown to stabilise α-helix and inhibit fraying by forming one or two additional hydrogen bonds

(Richardson & Richardson, 1988; Serrano & Fersht, 1989; Fersht & Serrano, 1993; Harper & Rose, 1993). The convention for specifying amino acid positions in and around helices is (... -N″-N′-Ncap-N1-N2- ... -C3-C2-C1-Ccap-C′-C″- ...) where Ncap and Ccap are the boundary residues that belong to both the helix and the adjacent sequence; each Ncap and Ccap residue makes one additional intrahelical hydrogen bond but departs from helical values of ϕ, ψ angles; the C′ residue generally adopts a left-handed conformation and is fulfilled most easily by Gly and Asn residues. According to these definitions, Lys2, Ser13 and Asn32 are the Ncap residues of helix I, II and III, respectively, while Ala10, Asn22 and His45 are the respective Ccap residues (these residues are underlined in Figure 8). Several hydrogen bonds located
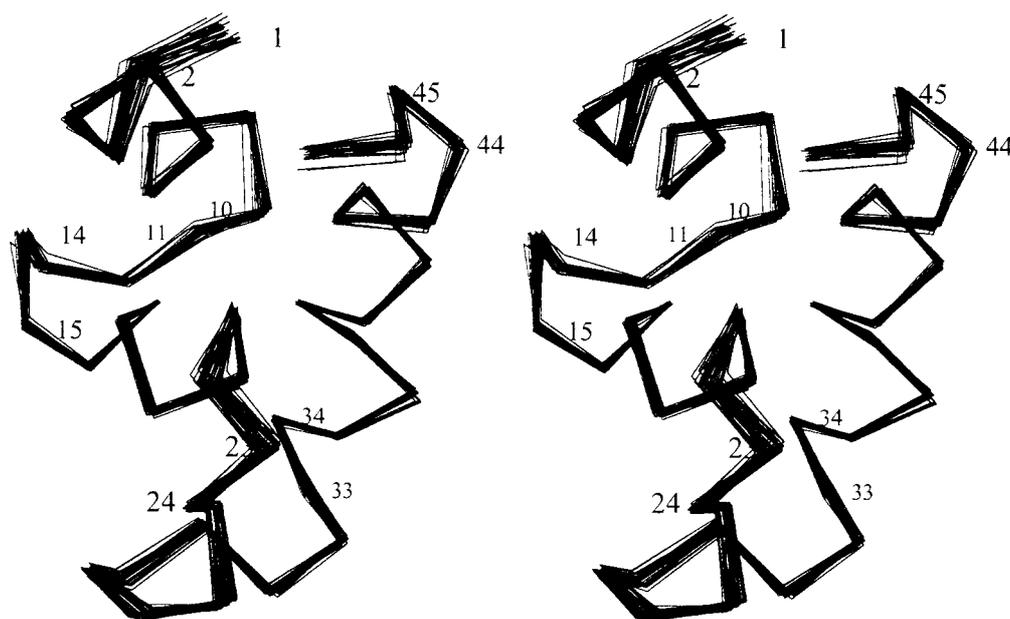
Figure 5. Stereoview of FruR(1-57)*. The α-carbon chains of the 34 final structures (residues 1 to 48) were superimposed from residues 1 to 47 with ANTHEPROT software tools (Geourjon & Deléage, 1995). The disordered C-terminal region (residues 49 to 60) is not represented.

at the ends of the helices obviously ensure their N and C-capping stabilisation, as described in the following.

For helix II, a capping box (Harper & Rose, 1993), likely exists at its N-terminal end. In such a capping box, the side-chain of the Ncap residue accepts a hydrogen bond from the amide proton of the N3 residue. Conversely, the side-chain of the N3 residue accepts a hydrogen bond from the amide proton of the Ncap residue. Therefore, this motif requires the presence of simultaneous polar

residues at the Ncap and N3 positions, i.e. Glu, Asp, Gln, His, Thr, Asn or Ser. This is the case for α-helix II, in which Ser13 and Thr16 occupy the Ncap and N3 positions, respectively. Moreover, the ϕ and ψ dihedral angles of Ser13 are very close to the characteristic geometry of capping boxes (i.e. $\phi = -94(\pm15)°$ and $\psi = 167(\pm5)°$), and the Ser13-NH $\rightarrow$ Thr16-O$^{\gamma1}$ hydrogen bond is observed in all structures. Although the complementary hydrogen bond Thr16-NH $\rightarrow$ Ser13-O$^{\gamma1}$ does not present the correct distance and geometry, modelling and dy-
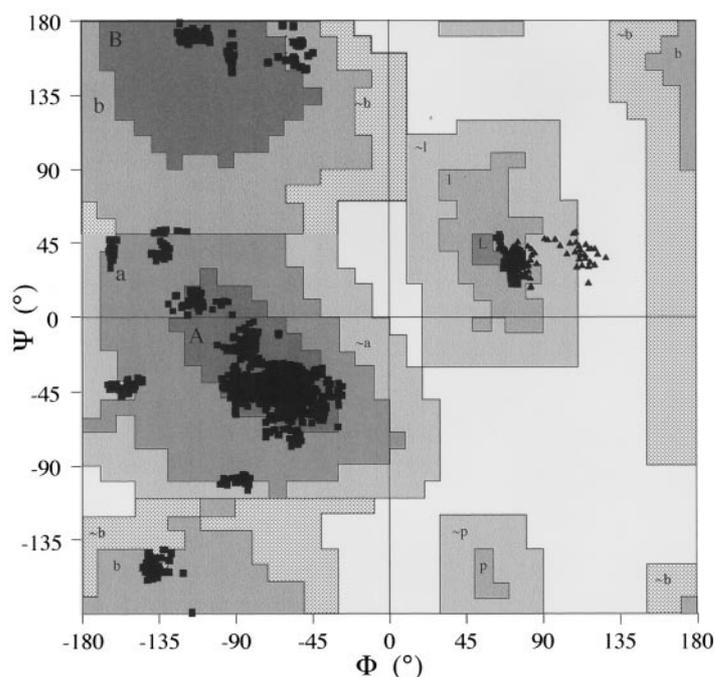
Figure 6. Ramachandran plot for the ϕ/ψ angles of the 34 final FruR(1-57)* structures (residues 1 to 47) drawn with PROCHECK (Laskowski *et al.*, 1993). Glycine residues are represented by triangles and all other residues by squares.
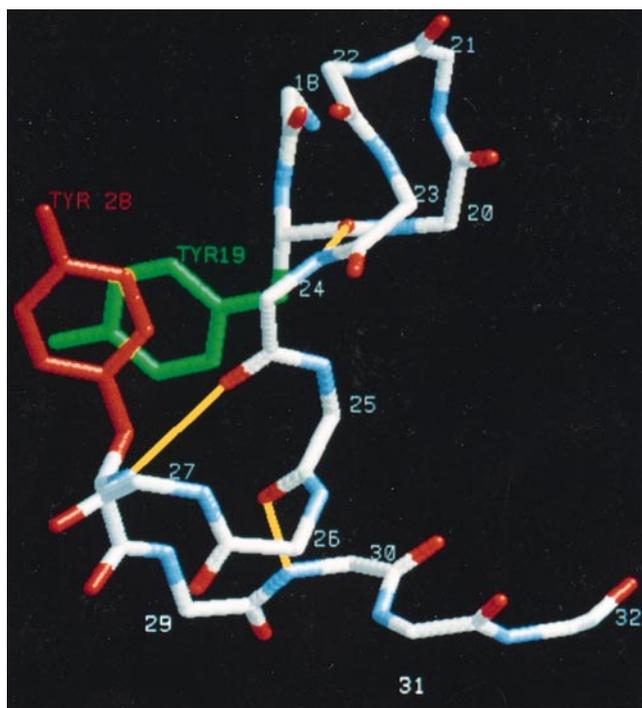
Figure 7. Average structure of the connecting segment between helices II and III in FruR(1 to 57)* showing the stabilisation of the 24 to 30 segment by hydrogen bonds (in yellow) and the stacking of aromatic rings of Tyr19 and Tyr28. Only backbone atoms are shown, except for the side-chains of tyrosine residues (drawn with Rasmol, R. Sayle program).

namic assays suggest that this bond could be easily achieved in all structures. The N terminus of helix III is less defined than helix II, and whereas a consensus capping box sequence is present (Ser31-Thr34), the characteristic hydrogen bonds are not observed. However, the possible salt-bridge between the side-chains of Lys26 and Asp32 may contribute to the stabilisation of the N-terminal end of helix III. For helix I, because of the helix geometry, a salt-bridge between the side-chains of Lys2 and Glu5 likely exists and stabilises the N-terminal end of helix I.

Examination of a sequence alignment of FruR DBD with LacI and PurR (Figure 8), and with the 33 other LacI members identified so far (not shown), reveals the presence of conserved consensus capping box motifs (Harper & Rose, 1993) for the three helices in most cases. For helix I, 61% of sequences present a Thr-X-X-Asp motif, and this increases to 86% when considering the presence of simultaneous polar residues at both Ncap and N3 positions. However, the analysis of 3D structure of PurR and LacI did not clearly show the presence of the reciprocal backbone-side-chain hydrogen-bonding interaction typical of the capping box. For the remaining LacI family members, the N terminus of putative helix I presents a sequence similar to FruR (i.e. a basic and an acidic residue at the Ncap and N3 position, respectively). It is thus probable that

these LacI members exhibit an N-terminal stabilisation of helix I by formation of a salt-bridge, as observed for FruR. The presence of a capping box for helix II is more obvious, since it seems to occur on FruR for the Ser13-X-X-Thr16 sequence which is, on the other hand, strictly conserved in 86% of the LacI members. For helix III, a consensus capping box sequence Ser-X-X-Thr is present in 47% of the LacI members (61% when considering all possible capping box residues), but the existence of a capping box is neither demonstrated here for FruR, nor for LacI DBD (Slijper *et al.*, 1996).

The C termini of α-helices involving glycine (or asparagine) residues have been classified into two major motifs, the Schellman motif and the α_L motif (Aurora *et al.*, 1994). The C-terminal ends of the three FruR(1-57)* helices follow the proposed rules for the Schellman motif, which requires (i) glycine (or asparagine) at the C' position (i.e. Gly11, Gly23 and Asn46 for helices I, II and III, respectively), (ii) either apolar, or arginine, or lysine at the C'' position, (iii) hydrophobic contacts between the C'' and C3 residues (Schellman, 1980). Such contacts are observed: Val12 and Ala7 in helix I, Lys24 and Tyr19 in helix II, and Tyr47 and Val42 in helix III. The Schellman arrangement produces a C'' ← C3 and C' → C2 hydrogen bonding pattern resulting in energetically favourable helix termination (Schellman, 1980; Aurora *et al.*, 1994). The C'' → C3 hydrogen bonds are indeed always observed for helices I, II and III (Val12-NH → Ala7-O', Lys24-NH → Tyr19-O', and Tyr47-NH → Val42-O', respectively). However, even though the distances are correct, the C' → C2 hydrogen bonds are not observed because of an unfavourable geometry. This poor geometric definition can be related to the low number of NOE correlations observed for glycine residues and used for the modelling. Finally, it should be noted that the C-capping of helix II is likely reinforced by the Asn22-H$^{\delta 2}$ → Ser18-O' bond, a backward type of hydrogen bond often observed in C termini of α-helices (Bordo & Argos, 1994). In addition, Asn22-H$^{\delta 2}$ may be involved in a hydrogen bond with Ser18-O$^{\gamma 1}$.

Inspection of sequence alignments of the 36 LacI members reveals the conservation of the Schellman motif for helix I, with glycine (75%) or asparagine (17%) at the C' position and apolar residues at the C3 position (Ala, 100%) and C'' position (Val, 94%). None of these residues has been shown to make contact with DNA in LacI or PurR (Chuprina *et al.*, 1993; Schumacher *et al.*, 1994), but these residues are components of the consensus sequence signature typical of the LacI family (Bairoch, 1993; see the legend to Figure 8). The residue conservation at these positions in the HTH motif can thus be explained by the energetically favourable termination of helix I afforded by this C-capping motif. For helix II, the Schellman motif conservation is not obvious in the LacI family because of the large variability observed in this region. On the contrary, this motif appears to be conserved in the C-capping of helix III for which (i) 67% of C' resi-
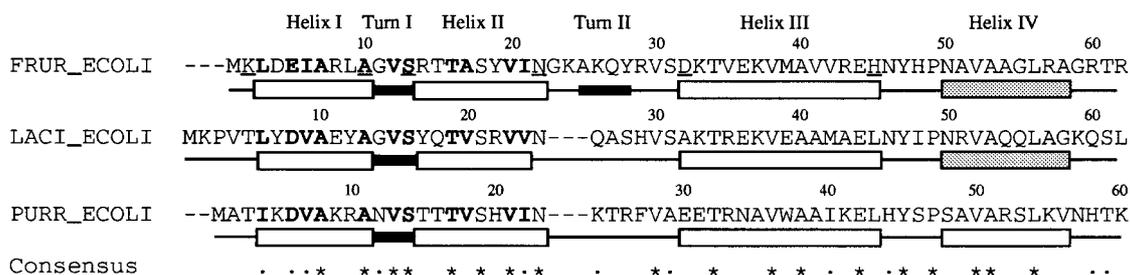
Figure 8. Sequence alignment and comparison of secondary structure of FruR, LacI and PurR DNA-binding domains. The sequences were extracted from Swissprot (Bairoch & Boeckmann, 1994) and termed from their corresponding code entry. The N-terminal parts of FRUR_ECOLI, LACI_ECOLI and PURR_ECOLI were aligned using the Clustalw 1.4 program (Thompson *et al.*, 1994) with default parameters (fixed and variable gap penalties set to 8, filtering level set to 1.5 and window width fixed to 10). Secondary structures were from Slijper *et al.* (1996) for LACI_ECOLI, from Schumacher *et al.* (1994) and Nagadoi *et al.* (1995) for PURR_ECOLI, and from this study for FRUR_ECOLI. White and black boxes indicate α-helix and turn, respectively. Grey boxes indicate the putative helix IV (called the hinge helix), which was characterised in the structure of PurR complexed with its operator (Schumacher *et al.*, 1994), but which is not formed either in free PurR DBD (Nagadoi *et al.*, 1995), or in LacI DBD (Slijper *et al.*, 1996), or in FruR DBD (this study). Helix I - Turn I - Helix II form the HTH motif defined in the LacI family by the following consensus sequence pattern: [LIVM]-x-[DE]-[LIVM]-A-x(2)-[STAG]-x-V-[GSTP]-x(2)-[STAG]-[LIVMA]-x(2)-[LIVMAFYN]-[LIVMC] (LacI family signature, Prosite 13 database; Bairoch, 1993). The corresponding residues are in bold characters in the three sequences. Turn II is a type III turn and is present only in FruR DBD. Ncap and Ccap residues of helices I, II and III are underlined in FruR. Identical and similar amino acid residues for the three DBD are symbolised by an asterisk (*) and a dot (.), respectively.

dues are glycine or asparagine, the remainder being histidine (17%) or alanine or small polar residues, and (ii) apolar residues occupy C″ position (Tyr, 92%; Phe, 8%) and C3 position (Met, 33%; Ile, 31%; Ala, 19%; Val, 17%).

Interestingly, no obvious N or C-capping motifs able to stabilise the helix IV are detected, neither from the 3D structure of PurR complexed with its DNA operator (Schumacher *et al.*, 1994), nor from sequence alignment inspection (Figure 8). This fact is in agreement with the unfolding state of this hinge helix in free PurR, LacI, and FruR DBD, confirming that hinge helix formation is induced by specific binding in DNA minor groove and oligomerisation of the whole repressor (Schumacher *et al.*, 1995; Nagadoi *et al.*, 1995; Spronk *et al.*, 1996).

Structural comparisons with PurR and LacI counterparts

The comparison of FruR, LacI and PurR DBD, reported in Figure 8, shows the close correspondence of helices I, II and III, and HTH motif. Moreover, the overall topology of the three α-helices of FruR DBD is close to that of LacI and PurR, as illustrated in Figure 9. The superposition of the HTH motifs yields the following C$^\alpha$ r.m.s. deviation: 1.3 Å between LacI HP56 and FruR(1-57)*, 1.6 Å between PurRN56 and FruR(1-57)*, and 1.5 Å between LacI HP56 and PurRN56. Moreover, this r.m.s.d. decreases to about 1 Å when the HTH motif of free FruR was compared to that of LacI and PurR complexed with DNA (Chuprina *et al.*, 1993; Schumacher *et al.*, 1994). In addition, the

superpositions of the three α-helical regions of these proteins yield a C$^\alpha$ r.m.s.d. of 1.5 Å between FruR and LacI, 1.9 Å between FruR and PurR, and 1.4 Å between LacI and PurR. All these data show the close similarity between the common structural elements for the three domains. In contrast, a large difference exists in the conformation of the connecting segment between helix II and III (Figures 8 and 9). It is noticeable that in FruR, this well-structured segment accommodates the presence of three additional residues without significant change in the position of helix III when compared to PurR and LacI. This structure has no equivalent in LacI or PurR, and is of particular interest, since 3D structural studies of these two proteins complexed with their respective operator report numerous contacts with DNA in this region (Chuprina *et al.*, 1993; Schumacher *et al.*, 1994). It is thus likely that the residues protruding at the surface of FruR in this region are involved in DNA binding and specificity of operator recognition. The examination of the 3D FruR(1-57)* structure highlights the fact that the best candidates are (i) the side-chain of Tyr28, which is stacked on that of Tyr19, (ii) Lys24 and Lys26, and (iii) Arg29, which appears to be completely accessible to the solvent. Concerning the residues of the HTH motif involved in DNA binding, PurR and LacI complexed with DNA display intermolecular contacts at the same positions (Chuprina *et al.*, 1993; Schumacher *et al.*, 1994), which likely also exists in the FruR-DNA complex. The corresponding residues can be deduced from sequence alignment: Leu3, Ser13, Thr15, Thr16, Ser18, Tyr19 and Asn22. In addition, Tyr47 and

Figure 9. Comparison of PurR, LacI and FruR DNA-binding domains. The NMR average structure of PurR(1-59) (Nagadoi *et al.*, 1995; PDB entry, 1PRU) is shown in green, LacI(1-56) (Slijper *et al.*, 1996) in white, and FruR(1-57) (this study; PDB entry, 1UXC) in red. The three structures were superimposed by their common helix-turn-helix motifs (residues 3 to 21 for FruR, 4 to 22 for PurR, and 6 to 24 for LacI). The reader is looking down the recognition helix (helix II) on the left, helix III is nearly vertical and helix I is in the background. The Figure was generated and rendered with MOLSCRIPT (P.J. Kraulis program) and Raster3D (Bacon & Anderson program), respectively.

quire only the first three DBD helices to bind specifically to its operator site; in contrast, PurR DBD cannot bind the *purF* operator site (Choi & Zalkin, 1992) and specific hinge helix recognition in the minor groove seems to be essential for specific DNA binding, as stated by Nagadoi *et al.* (1995). Finally, it should be added that the presence of additional residues in the connecting segment between helices II and III is not a distinctive feature of FruR; sequence alignment of the 36 lacI family members described to date highlights insertion of one, two or three residues in this region (one for YGLR_STRCO and ENDR_BACPO;, two for EBGR_ECOLI, and three for FRUR_ECOLI, FRUR_SALTY, SCRR_KLEPN, SCRR_SALTY and RBTR_KMEAE; proteins are termed with their corresponding Swissprot code entry). Taken together, all these observations suggest that different structural strategies are used by the various members of the LacI family to ensure the specific recognition of their respective DNA operators.

Conclusions

The FruR DNA-binding domain, C-terminally fused to a 6xHis sequence, has been overproduced in *E. coli* at high yield and uniformly ^{15}N-labelled. The highly purified and homogeneous recombinant FruR(1-57)* product was able to specifically recognise *in vitro* the natural *ace* operator DNA and exhibited well-resolved NMR spectra, which allowed the complete ^{1}H and ^{15}N resonance assignments. The highly flexible LQHHHHHH C-terminal tail did not complicate the interpretation of the NMR spectra, since the corresponding amide protons were rapidly exchangeable, and since most of the corresponding aliphatic protons did not exhibit any detectable NOE correlation. These observations indicate the absence of any stable interaction of the C-terminal 6xHis sequence with the FruR DBD. They demonstrate that this type of plasmid construction can be advantageously used for protein structure determination by NMR, provided that the absence of structural and functional interference is checked.

The measurement of numerous interproton distances and dihedral angles permitted the molecular modelling of the 3D structure of the 47 first N-terminal residues of FruR. The quality of the stereochemistry in the final set of calculated structures is excellent, with a r.m.s.d. of 0.37 Å at the main-chain level. The FruR DBD exhibits a typical HTH motif stabilised by a third helix, and the topology of these structural elements is very close to that of LacI and PurR. These results confirm the universality of this topology in the whole LacI protein family. However, the structure comparison of the connecting fragment between helix II and helix III in FruR with that of LacI and PurR showed a striking difference. This segment, containing three additional residues in FruR and without equivalent in LacI and PurR, is highly structured and stabil-

Asn52 are also likely to be involved in DNA contacts.

One can wonder whether the connecting segment between helices II and III contains three additional residues that provide a particular conformation to FruR DBD when compared to LacI and PurR counterpart. It can be supposed that this conformation is related to the pleiotropic gene regulation ensured by FruR, i.e. its ability to bind to several different DNA operators in *E. coli* (Nègre *et al.*, 1996). In fact, when compared to LacI and PurR specific DNA operators, FruR operators exhibit a much lower palindromy. A possible explanation may be that the recognition of several pseudo-palindromic operators requires a pre-defined shape of DBD to accommodate its binding on any half-site operator. On the contrary, DBD binding on a nearly perfect palindromic operator, such as the *lac* operator, may require more conformational fitting to ensure selective and specific DNA base recognition. Besides, LacI appears to re-

ised by several hydrogen bonds and the cycle stacking of Tyr19 and Tyr28. It is proposed that the shape of this structural element plays an important role in the binding to the different operators recognised by FruR.

The well-defined structure of FruR(1-57)* allowed us to scrutinize the hydrogen bond possibilities and provided a starting point for helix N and C-capping analyses, which highlight the importance of the residues that fulfil helix-capping roles and ensure the stabilisation of the structure. Similar modes of stabilisation were observed for all DBD LacI family members, which exhibit well conserved key residues for the N- and C-capping of the three helices, except for helices II and III ends flanking the connecting segment because of a large sequence variability in this region. Schellman motifs and N capping boxes provide energetically favourable helix terminations, forming two additional hydrogen bonds at each end, and are a structural justification for residue conservation at the N-cap, N3, C3, C′ and C″ positions. Most of these residues are components of the consensus HTH LacI family signature, and their structural role provides a rational explanation for this signature deduced from protein sequence alignments. As described for free PurR and LacI DBD, the putative hinge helix (helix IV) of free FruR is unfolded. In contrast with helices I, II and III, this helix does not contain any typical residue pattern able to ensure its stabilisation by end capping. This is in agreement with the fact that the hinge helix is formed only upon repressor oligomerisation and complexation with DNA operator. Finally, the structural role of most FruR DBD residues was characterised, and potential residues involved in DNA binding and specific operator recognition were identified. Work is in progress to determine the 3D structure of FruR DBD complexed with a 14 bp half-site consensus DNA operator and to determine the functional role of these residues, in particular for the specificity of DNA recognition.

Materials and Methods

Construction of plasmid pCB4 and expression of FruR(1-57)*

A 171 bp DNA fragment containing the 5′-end of the *fruR* gene was obtained by PCR amplification from *E. coli* chromosomal DNA. Two oligonucleotides were used, one carrying a *Nde*I site within the following sequence: 5′-TAT<u>CATATG</u>AAACTGGATGAAATCG-3′, and the other with a *Pst*I site in a sequence complementary to the region spanning to the 57th codon of the *fruR* gene: 5′-TAT<u>CTGCAG</u>ACGAAGCCCAGCTGCCACGGCG-3′. The PCR fragment obtained was digested with *Nde*I and *Pst*I endonucleases, and cloned between the corresponding sites of the expression vector pT7.7 with a 6xHis tag (Cortay *et al.*, 1994). The resulting plasmid pCB4 encoded the DBD of FruR as a fusion protein with a C-terminal LQHHHHHH sequence. The expression product of pCB4, obtained after transformation in the host strain

BL21(DE3), was called FruR(1-57)* and presented the following sequence:

<pre>
        10   20   30
   MKLDEIARLA GVSRTTASYV INGKAKQYRV

        40   50   60
   SDKTVEKVMA VVREHNYHPN AVAAGLR**LQH** **HHHHH**
</pre>

Longer N-terminal FruR fragments were cloned and overproduced, in particular FruR(1-63)* (i.e. with the LQHHHHHH C-terminal extension) and FruR(1-63) without extension.

Overproduction and purification of FruR(1-57)*

E. coli strain BL21(DE3)[pCB4] was grown exponentially at 37°C in NZCYM medium (BIO 101, Inc.) supplemented with ampicillin (100 μg/ml). The production of FruR(1-57)* was induced by the addition of 0.4 mM isopropyl-β, D-thiogalactopyranoside when the A_{600} of the culture reached 0.8. Three hours after induction, bacteria were disrupted in a French pressure cell at 20,000 psi in 10 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 1 mM EDTA, 10% (v/v) glycerol, 1 mM phenylmethylsulphonyl fluoride, 1% (v/v) Triton and Benzon nuclease (Merck) at 240 units/ml. After centrifugation at 30,000g for 30 minutes, the supernatant was incubated for 30 minutes with Ni^{2+} NTA-agarose beads (Quiagen) equilibrated in a PBS buffer (16 mM Na_2HPO_4, 4 mM NaH_2PO_4 (pH 6.8), 300 mM NaCl) containing 1% Triton. The gel was washed with three gel volumes of PBS buffer supplemented with 1% Triton and at least ten gel volumes of PBS buffer without Triton but containing 10 mM imidazole. After elution with 0.3 M imidazole in PBS buffer, FruR(1-57)* was purified by chromatography on a sulfopropyl CS4 column (elution at 150 mM NH_4Cl, 20 mM sodium phosphate (pH 6.8), 100 mM NaCl, from a linear gradient of 0 M to 0.5 M NH_4Cl; see Scarabel *et al.*, 1995). For NMR samples, the fractions containing highly purified FruR(1-57)* were concentrated and dialysed against 10 mM sodium phosphate buffer (pH 5.9), 50 mM NaCl, 70 mM NH_4Cl, 5% (v/v) ^2H_2O and 0.05% sodium azide by using a Centriprep-3 ultrafiltration device (Amicon). Uniformly ^{15}N-labelled FruR(1-57)* was produced by growing BL21(DE3)[pCB4] cells in Celton N (Martek, 95% ^{15}N-enrichment).

NMR spectroscopy

All NMR experiments were recorded at 500 MHz on a Varian Unity-*plus* spectrometer equipped with *ultra-nmr* shims and using a triple resonance proton-carbon-nitrogen 5 mm probe with a self-shielded *z*-gradient coil. Spectra were acquired at 20°C and additional data at 10°C and 30°C. The carrier frequency was set on the water resonance, and quadrature detection in the indirectly observed dimensions was obtained with States-TPPI method (Marion *et al.*, 1989a).

For 2D homonuclear ^1H experiments, conventional phase sensitive DQF-COSY, Clean-TOCSY, and NOESY were performed (see Lesage *et al.*, 1996 and references therein). For NOESY, mixing times in the 50 to 250 ms range were used. The spectra were recorded with 6000 Hz spectral width and data sets collected as 512 and 2048 points in t_1 and t_2 dimensions, respectively, with 32 or 64 scans per increment. For heteronuclear ^{15}N-^1H experiments, the spectral width was set to

2600 Hz in the ^{15}N dimension. Broad band ^{15}N decoupling during acquisition was accomplished by means of a WALTZ-16 sequence. J-HMQC (Kay & Bax, 1990) and a gradient version of HSQC (Ruiz-Cabello *et al.*, 1992) were recorded with 2048 points in ^{1}H dimension and 512 increments in t_1. 3D ^{15}N NOESY-HMQC (Marion *et al.*, 1989b) and 3D ^{15}N HMQC-TOCSY (Wijmenga *et al.*, 1989) were acquired with 256, 64 and 1024 points in F_1, F_2 and F_3 dimensions, respectively. A 150 ms mixing time was used for ^{15}N NOESY-HMQC and a 80 ms spin lock for 3D ^{15}N HMQC-TOCSY. The HNHA pulse sequence (Vuister & Bax, 1993) was accommodated to hardware and software requirements of the Unity-*plus* 500 MHz spectrometer. Water suppression was carried out using selective, low-power irradiation during the 1.5 s relaxation delay and during the mixing time in NOESY experiments. For NOESY experiments, a SCUBA pulse train was applied after water saturation. Alternatively, water was suppressed using a WATERGATE sequence before detection (Piotto *et al.*, 1992). Homonuclear ^{1}H spectra were carried out with a 3 mM FruR(1-57)* sample, either in 95%H$_2$O / 5% ^{2}H$_2$O or 100% ^{2}H$_2$O solution, and heteronuclear ^{1}H-^{15}N correlation spectra with a 2 mM ^{15}N-labelled FruR(1-57)* sample.

For data collection and processing, VNMR software (Varian) was used. Prior to Fourier transformation, shifted squared sine-bell and sine-bell apodization functions were used in directly and indirectly detected dimensions, respectively. For 3D spectra, linear prediction was used to improve resolution in the indirect dimensions. Zero-filling was applied in all cases. Resulting spectra were baseline-corrected using a spline fitting of pre-defined baseline regions.

NMR derived constraints and structure calculations

NOE intensities used as input for the structures calculations were obtained from the NOESY spectrum recorded with a 150 ms mixing time on the fully protoned sample and checked for spin diffusion on spectra recorded at shorter mixing times (50 and 100 ms). NOEs were partitioned into four categories of intensities that were converted into distances ranging from a common lower limit of 1.8 Å (sum of the van der Waals radii) to upper limits of 2.6 Å, 3.3 Å, 3.8 Å and 5 Å. The cross-peak intensity of H$^\delta$-H$^\varepsilon$ protons of Tyr47 was used as the distance reference (2.45 Å). NOEs back-calculations were performed from calculated structures by using the standard procedure of X-PLOR 3.1 (Brünger, 1992). Sterospecific assignments for γ-methyl protons of valine and for β-methylene protons were done as described (Barsukov & Lian, 1993). Protons without stereospecific assignments were treated as pseudoatoms, and the correction factors were added to the distance constraints according to Wüthrich *et al.* (1983). Coupling constants ^3J NHH$^\alpha$ were measured from both J-HMQC and HNHA spectra and converted into dihedral angle constraints as described (Kay & Bax, 1990; Vuister & Bax, 1993, respectively). Dihedral φ angles were constrained to $-60° \pm 30°$ for ^3J NHH$^\alpha$ that were smaller than 6 Hz, and to $120° \pm 40°$ for ^3J NHH$^\alpha$ values that were larger than 9 Hz.

Three-dimensional structures were generated from the NOE distances and deduced dihedral angles data using X-PLOR 3.1 from Brünger (1992) on an IBM rs6000. The default X-PLOR parameter set was used, except for some minor modifications to increase the duration of the molecular dynamic simulations and the number of energy minimisation steps as described (Lesage *et al.*, 1996).

A final minimisation step was applied with the standard force-field parameters except that K_{bonds} and K_{angles} were fixed to 600 kcal mol^{-1} Å^{-2} and 90 kcal mol^{-1} rad^{-2}, respectively. Ensembles of 50 structures were calculated to widely sample the conformational space, and the structures were compared on the basis of pairwise r.m.s.d. for the backbone atom coordinates (N, C$^\alpha$ and C'). Local analogies were analysed by calculating the local r.m.s.d. of a tripeptide window slided along the sequence. Average structure was calculated using the average X-PLOR procedure, followed by 5000 cycles of Powell energy minimisation.

Statistical analysis, superimposition of structures, 3D graphic displays and manipulations were achieved by using ANTHEPROT 2.0 software (Geourjon & Deléage, 1995). The secondary structure elements and Ramachandran plots were analysed according to the Kabsch & Sander (1983) definition rules, as incorporated in the program PROCHEK (Laskowski *et al.*, 1993). To identify hydrogen bonds, a search was undertaken with the X-PLOR 3.1 routine using default distance and angle parameters. To assess the presence of a hydrogen bond, a donor-acceptor distance not greater than 3.5 Å was considered (allowing a 0.5 Å error in the typical bond length) and the hydrogen bonding potential energy should be lower than -0.5 kcal/mol.

References

Aurora, R., Srinivasan, R. & Rose, G. D. (1994). Rules for α-helix termination by glycine. *Science,* **264**, 1126–1130.

Bairoch, A. (1993). The PROSITE dictionary of sites and patterns in proteins, its current status. *Nucl. Acids Res.* **21**, 3097–3103.

Bairoch, A. & Boeckmann, B. (1994). The SWISS-PROT protein sequence data bank: current status. *Nucl. Acids Res.* **22**, 3578–3580.

Barsukov, I. L. & Lian, L. Y. (1993). Structure determination from NMR data. In *NMR of Macromolecules. A Practical approach* (Roberts, G. C., ed.), pp. 315–357, Oxford University Press, Oxford.

Böckmann, A., Penin, F. & Guittet, E. (1996). Rapid estimation of relative amide proton exchange rates of ^{15}N-labelled proteins by straightforward water

selective NOESY-HSQC experiment. *FEBS Letters,* **393**, 191–196.

Bordo, D. & Argos, P. (1994). The role of side-chain hydrogen bonds in the formation and stabilization of secondary structure in soluble proteins. *J. Mol. Biol.* **243**, 504–519.

Brennan, R. G. (1992). DNA recognition by the helix-turn-helix motif. *Curr. Opin. Struct. Biol.* **2**, 100–108.

Brennan, R. G. & Matthews, B. W. (1989). The helix-turn-helix DNA binding motif. *J. Biol. Chem.* **264**, 1903–1906.

Brünger, A. T. (1992). *X-PLOR: A System for X-ray Crystallography and NMR. Version 3.1 Manual* Yale University Press, New Haven, CT.

Choi, K. Y. & Zalkin, H. (1992). Structural characterization and corepressor binding of the *Escherichia coli* purine repressor. *J. Bacteriol.* **174**, 6207–6214.

Chou, P. Y. & Fasman, G. D. (1977). β Turns in proteins. *J. Mol. Biol.* **115**, 135–175.

Chuprina, V. P., Rullman, J. A. C., Lamerichs, M. J. N., van Boom, J. H., Boelens, R. & Kaptein, R. (1993). Structure of the complex of *lac* repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. *J. Mol. Biol.* **234**, 446–462.

Cortay, J. C., Nègre, D., Scarabel, M., Ramseier, T. M., Narendra, B., Vartak, M. H., Saier, M. H. & Cozzone, A. J. (1994). *In vitro* asymmetric binding of the pleiotropic regulatory protein, FruR, to the *ace* operator controlling glyoxylate shunt enzyme synthesis. *J. Biol. Chem.* **269**, 14885–14891.

de Vlieg, J., Scheek, R. M., van Gunsteren, W. F., Berendsen, H. J. C., Kaptein, R. & Thomason, J. (1988). Combined procedure of distance geometry and restrained molecular dynamics techniques for protein structure determination from nuclear magnetic resonance data: application to the DNA binding domain of *lac* repressor from *Escherichia coli*. *Proteins: Struct. Funct. Genet.* **3**, 209–218.

Fersht, A. R. & Serrano, L. (1993). Principle in protein stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol.* **3**, 75–83.

Geourjon, C. & Deléage, G. (1995). ANTHEPROT 2.0: a 3D module fully coupled with protein sequence analysis methods. *J. Mol. Graph.* **13**, 209–212.

Harper, E. T. & Rose, G. D. (1993). Helix stop signals in proteins and peptides: the capping box. *Biochemistry,* **32**, 7605–7609.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577–2637.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. (1985). A protein structure from nuclear magnetic resonance data. *Lac* repressor headpiece. *J. Mol. Biol.* **182**, 179–182.

Kay, L. E. & Bax, A. (1990). New methods for the measurement of NH-C$^\alpha$H coupling constants in ^{15}N-labeled proteins. *J. Magn. Reson.* **86**, 110–126.

Lamerichs, R. M. J. N., Boelens, R., van der Marel, G. A., van Boom, J. H., Kaptein, R., Buck, F., Fera, B. & Rüterjans, H. (1989). ^{1}H-NMR study of a complex between the *lac* repressor headpiece and a 22 base pair symmetric *lac* operator. *Biochemistry,* **28**, 2985–2991.

Lamerichs, R. M. J. N., Boelens, R., van der Marel, G. A., van Boom, J. H. & Kaptein, R. (1990). Assignment of the ^{1}H-NMR spectrum of a *lac* repressor head-piece-operator complex in H$_2$O and identification of NOEs. *Eur. J. Biochem.* **194**, 629–637.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). Procheck: A program to check the stereochemical quality of protein structure. *J. Appl. Crystallog.* **26**, 283–291.

Lesage, A., Penin, F., Geourjon, C., Marion, D. & van der Rest, M. (1996). Trimeric assembly and three-dimensional structure model of the FACIT collagen COL1-NC1 junction from CD and NMR analysis. *Biochemistry,* **35**, 9647–9660.

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G. & Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science,* **271**, 1247–1354.

Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989a). Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J. Magn. Reson.* **85**, 393–399.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989b). Three-dimensional heteronuclear NMR of ^{15}N-labelled proteins. *J. Am. Chem. Soc.* **111**, 1515–1517.

Nagadoi, A., Souichi, M., Nakamura, H., Enari, M., Kobayashi, K., Yamamoto, H., Sampei, G., Mizobuchi, K., Schumacher, M. A., Brennan, R. G. & Nishimura, Y. (1995). Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain. *Structure,* **3**, 1217–1224.

Nègre, D., Bonod-Bidaud, C., Geourjon, C., Deléage, G., Cozzone, A. J. & Cortay, J. C. (1996). Definition of a consensus DNA-Binding site for the pleiotropic regulatory protein, FruR. *Mol. Microbiol.* **21**, 257–266.

Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR,* **2**, 661–665.

Ramseier, T. M., Nègre, D., Cortay, J. C., Scarabel, M., Cozzone, A. J. & Saier, M. H., Jr (1993). *In vitro* binding of the pleiotropic transcriptional regulatory protein, FruR, to the *fru, pps, ace, pts* and *icd* operons of *Escherichia coli* and *Salmonella typhimurium*. *J. Mol. Biol.* **234**, 28–44.

Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of α-helices. *Science,* **240**, 1648–1652.

Ruiz-Cabello, J., Vuister, G. W., Moonen, C. T. W., van Gelderen, P., Cohen, J. S. & van Zijl, P. C. M. (1992). Gradient-enhanced heteronuclear correlation spectroscopy. Theory and experimental aspects. *J. Magn. Reson.* **100**, 282–302.

Saier, M. H., Ramsaier, T. M. & Reizer, J. (1996). Regulation of carbon utilization. In *Escherichia coli and Salmonella, Cellular and Molecular Biology* (Neidhardt, F. C., ed.), pp. 1325–1343, ASM Press, Washington, DC.

Scarabel, M., Penin, F., Bonod-Bidaud, C., Nègre, D., Cozzone, A. J. & Cortay, J. C. (1995). Overproduction, purification and structural characterization of the functional N-terminal DNA-binding domain of the *fru* repressor from *Escherichia coli* K-12. *Gene,* **153**, 9–15.

Schellman, C. (1980). In *Protein Folding* (Jaenicke, R., ed.), pp. 53–61, Elsevier/North-Holland, New York.

Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. (1994). Crystal structure of LacI member,

PurR, bound to DNA: minor groove binding by α helices. *Science,* **266**, 763–770.

Schumacher, M. A., Choi, K. Y., Lu, F., Zalkin, H. & Brennan, R. G. (1995). Mechanism of corepressor-mediated specific DNA binding by the purine repressor. *Cell,* **83**, 147–155.

Serrano, L. & Fersht, A. R. (1989). Capping and α-helix stability. *Nature,* **342**, 296–299.

Slijper, A. M., Bonvin, A. M. J. J., Boelens, R. & Kaptein, R. (1996). Refined structure of *lac* repressor headpiece (1-56) determined by relaxation matrix calculations from 2D and 3D NOE data: change of tertiary structure upon binding to the *lac* operator. *J. Mol. Biol.* **259**, 761–773.

Spronk, C. A., Slijper, M., van Boom, J. H., Kaptein, R. & Boelens, R. (1996). Formation of the hinge helix in the lac repressor is induced upon binding to the lac operator. *Nature Struct. Biol.* **3**, 916–919.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.

Vuister, G. W. & Bax, A. (1993). Quantitative J correlation: a new approach for measuring homonuclear three-bond J(NN-HA) coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.* **115**, 7772–7777.

Weickert, M. J. & Adhya, S. (1992). A family of bacterial regulators homologous to *Gal* and *Lac* repressors. *J. Biol. Chem.* **267**, 15869–15874.

Wijmenga, S. S., Hallenga, K. & Hilbers, C. W. (1989). A three-dimensional heteronuclear multiple quantum coherence homonuclear Hartmann-Hahn experiment. *J. Magn. Reson.* **84**, 634–642.

Wintjens, R. & Rooman, M. (1996). Structural classification of HTH DNA-binding domains and protein-DNA modes. *J. Mol. Biol.* **262**, 294–313.

Wishart, D. S., Sykes, B. D. & Richards, F. M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **222**, 311–333.

Wishart, D. S., Sykes, B. D. & Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry,* **31**, 1647–1651.

Wüthrich, K. (1986) . In *NMR of Proteins and Nucleic Acids,* John Wiley & Sons, New York.

Wüthrich, K., Billeter, M. & Braun, W. (1983). Pseudostructures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.* **169**, 949–961.

Edited by P. E. Wright

JMB
Online

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

Supplementary material comprising two Tables is available from JMB Online.