The EMBO Journal vol.7 no.13 pp.4305-4309, 1988

Secondary structure determination for the *Antennapedia* homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif

Gottfried Otting, Yan-qiu Qian, Martin Müller[1], Markus Affolter[1], Walter Gehring[1] and Kurt Wüthrich

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich and [1]Biozentrum der Universität Basel, Abt. Zellbiologie, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Communicated by K.Wüthrich

The homeodomain encoded by the *Antennapedia* (*Antp*) gene of *Drosophila* was studied in aqueous solution by nuclear magnetic resonance (NMR). Sequence-specific resonance assignments have been obtained for the complete polypeptide chain of 68 amino acid residues. The secondary structure determined from nuclear Overhauser effects (NOE) and information about slowly exchanging amide protons includes three helical segments consisting of the residues 10-21, 28-38 and 42-52, respectively. Combination of the presently available NMR data with computer modeling provided preliminary evidence for the presence of a helix-turn-helix motif in the homeodomain. Near the turn, this supersecondary structure appears to be very similar to the DNA binding site in the 434 and P22 c2 repressors, but both helices in the homeodomain include 2-3 additional residues when compared with these prokaryotic DNA-binding proteins.

Key words: homeodomain/two-dimensional NMR/NOE/helix-turn-helix motif/sequence-specific DNA-binding protein

Introduction

Based on sequence homologies it has been hypothesized that the structure of the homeodomain of eukaryotic gene regulatory proteins includes a helix-turn-helix motif designed for sequence-specific DNA-binding (Laughon and Scott, 1984; Shepherd *et al.*, 1984). To test this hypothesis a polypeptide with 68 amino acid residues has been expressed in *Escherichia coli* which consists of the *Antennapedia* (*Antp*) homeodomain with an additional methionyl residue at the N terminus and an extra pentapeptide segment at the C-terminal end (Figure 1). The gene construct used for the expression of the protein, its isolation and characterization are described in the preceding paper (Müller *et al.*, 1988). It was demonstrated that this protein binds specifically to certain DNA sequences. In particular, footprinting studies showed that the isolated homeodomain protects the same DNA sequence elements from DNase I attack as the full length *Antp* protein. Since extensive studies with prokaryotic DNA-binding proteins showed that the specificity of binding is largely determined by the polypeptide conformation of the DNA-binding domain, these observations imply that the conformation of the homeodomain in the intact *Antp* protein is maintained in aqueous solution of the isolated homeodomain. Thus a structure determination of the homeodomain is of direct interest also with regard to the DNA-binding properties of the *Antp* protein.

In the preceding paper (Müller *et al.*, 1988) it was shown that under non-oxidizing conditions the homeodomain has a monomeric form in aqueous solution. The relatively low mol. wt and high solubility in water make this protein amenable for a determination of its 3-dimensional structure by NMR (Wüthrich *et al.*, 1982; Wüthrich, 1986). Here we describe the secondary structure determination by NMR for the homeodomain of the *antennapedia* protein from *Drosophila*, and present preliminary evidence for the presence of a helix-turn-helix motif similar to that found in several prokaryotic gene regulatory proteins (Ohlendorf *et al.*, 1982, 1983; Steitz *et al.*, 1982; Zuiderweg *et al.*, 1984; Ptashne, 1986; Mondragon *et al.*, 1988).

Results and discussion

Exploratory circular dichroism (CD) measurements showed a denaturation temperature of ~48°C for the homeodomain at pH values of ~4.5-5.0, and the thermal denaturation was found to be reversible. Based on these observations we decided to use pH 4.3 and a temperature of 20°C for the structure determination by NMR.

Proton NMR assignments for the Antennapedia homeodomain

Sequence-specific ^{1}H-NMR assignments were obtained following the usual strategy (Billeter *et al.*, 1982; Wagner and Wüthrich, 1982a). In a first step the spin systems of the individual amino acid residues were identified in a D_2O solution of the protein at 20°C using two-quantum filtered correlated spectroscopy (2QF-COSY) (Rance *et al.*, 1983), three-quantum filtered correlated spectroscopy (3QF-COSY) (Müller *et al.*, 1986), clean 2-dimensional total correlation spectroscopy (clean-TOCSY) with a mixing time of 100 ms (Griesinger *et al.*, 1988), two-dimensional nuclear Overhauser enhancement spectroscopy (NOESY) with a mixing time of 100 ms (Anil-Kumar *et al.*, 1980) and two-quantum spectroscopy (Braunschweiler *et al.*, 1983). Prior to the sequential assignments, complete spin system identifications were obtained for 34 out of the 68 amino acid residues. Thereby NOESY was used only to connect the spin systems of the aromatic rings with their corresponding αCH-βCH$_2$ fragments.

In a second step, sequential assignments were pursued using two NOESY spectra which had been recorded in H_2O at 10°C with a mixing time of 100 ms, and at 20°C with a mixing time of 150 ms, respectively. Figure 2A shows an overview of the spectral region of the NOESY spectrum at 20°C which was used for the sequential resonance assignments. The spectral region shown comprises along the

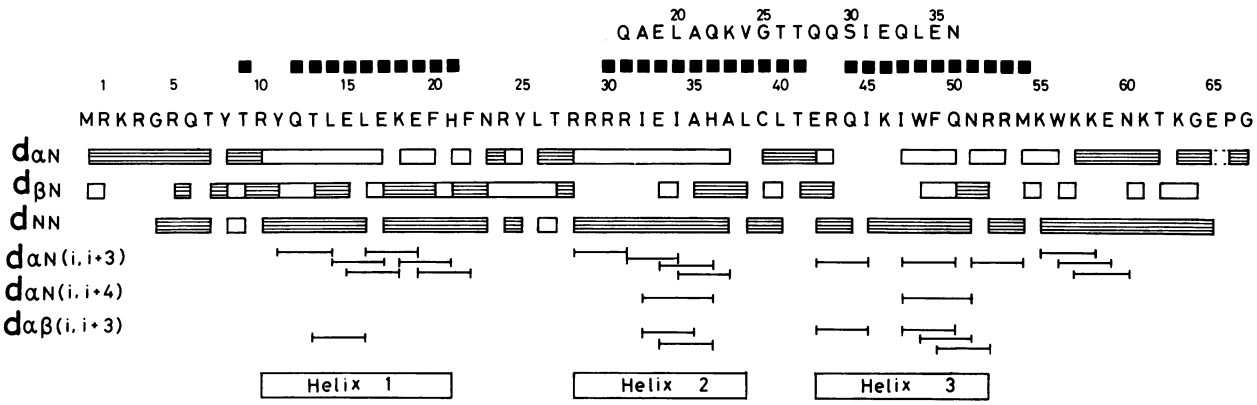

Fig. 1. Amino acid sequence and survey of sequential and medium-range NOE connectivities observed in the *Antp* homeodomain. The top line shows the amino acid sequence of the helix-turn-helix motif in the 434 repressor. Filled squares above the homeodomain sequence mark the location of slowly exchanging amide protons, which were observed in a NOESY experiment recorded at 10°C during 8 h immediately after dissolving the protein in D_2O. Below the sequence, hatched and open bars indicate sequential $d_{\alpha N}$, $d_{\beta N}$ and d_{NN} connectivities represented by strong and weak NOESY cross-peaks, respectively. The $d_{\alpha\delta}$ cross-peaks with Pro 66 are shown with broken lines. $d_{\alpha N}(i,i+3)$, $d_{\alpha N}(i,i+4)$ and $d_{\alpha\beta}(i,i+3)$ connectivities are represented by lines below the sequential connectivities. At the bottom the location of three helices is indicated, which were identified from the data collected in this figure.

ω_2 axis the resonance frequencies of all amide protons, and along ω_1 the entire spectrum. Cross-peaks between pairs of amide protons are found in the region with $\omega_1 = 7.4-9.2$ p.p.m. Most of these cross-peaks correspond to short distances between amide protons of sequentially neighbouring amino acid residues, d_{NN}. Similarly, short sequential distances between $C^{\alpha}H$ of residues i and the amide proton of residue i+1, $d_{\alpha N}$, are manifested by cross-peaks in the ω_1 frequency interval of the C^{α} protons from $\omega_1 = 3.0-4.9$ p.p.m., and the $d_{\beta N}$ connectivities are in the region $\omega_1 = 1.3-4.0$ p.p.m. Figure 2B shows a blow-up of the spectral region containing cross-peaks between different amide protons. In all, 49 cross-peaks between sequentially neighbouring amide protons were identified in this part of the spectrum. As an illustration of the sequential assignment procedure (Wagner and Wüthrich, 1982a), the assignment pathway for the second helix with residues 28–40 is indicated.

A survey of all experimentally observed sequential connectivities is presented in Figure 1. With the exception of Gln44–Ile 45, where both amide proton resonances have virtually identical chemical shifts, all pairs of sequentially neighbouring amino acid residues are connected by at least one of the sequential connectivities d_{NN}, $d_{\alpha N}$ or $d_{\beta N}$. In the case of Pro 66 the sequential connectivity with $C^{\alpha}H$ of Glu 65 is to δCH_2, which shows that this peptide bond is in the *trans* form (Wüthrich, 1986). The peptide segments identified with the sequential assignments could unambiguously be fitted to the amino acid sequence, which yielded the sequence-specific assignments.

Location of secondary structure elements in the homeodomain sequence

The data used for the secondary structure determination are collected in Figure 1, where strong and weak nuclear Overhauser enhancement (NOE) cross-peak intensities are distinguished. Because the NOESY spectra were recorded with mixing times of 100 ms or longer, relaxation effects and spin diffusion must be considered in any attempt at quantitative evaluation of the cross-peak intensities in terms of inter-proton distances (Wüthrich, 1986). However, keeping these inherent difficulties in mind, a qualitative comparison of relative cross-peak intensities is still meaningful in view of a secondary structure determination (Billeter *et al.*, 1982). Thus, the different successions of several intense sequential NOESY cross-peaks between amide protons are indicative of regions with helical secondary structure in the protein conformation (Wüthrich *et al.*, 1984). Further, direct evidence for helical secondary structure is gained from $d_{\alpha N}(i,i+3)$ and $d_{\alpha\beta}(i,i+3)$ connectivities, and supporting evidence comes from amide proton exchange rates, which are slower in helices due to intra-helix hydrogen bonds (Wagner and Wüthrich, 1982b). Thereby, since the amide protons of the first three residues in an α-helix are not involved in hydrogen bonds, the start of the helix is usually three residues before the first slowly exchanging amide. In agreement with this prediction, the NOE constraints which are indicative for helical structure in the homeodomain extend beyond the first slowly exchanging amide proton (Figure 1). It is interesting to note, however, that in all three helices of the homeodomain only the first two amide protons exchange rapidly. We tentatively attribute the slow exchange for the third helical residues to hydrogen bonding with the side chain of the residue immediately preceding the helix, which is in all three cases a threonine (Figure 1). Overall, the combined information from the sequential NOE connectivities d_{NN}, and the medium-range connectivities $d_{\alpha N}(i,i+3)$ and $d_{\alpha\beta}(i,i+3)$ identify three helical segments extending approximately from residues 10 to 21, 28 to 38 and 42 to 52, respectively. The helices are very likely of the α-type, since $d_{\alpha N}(i,i+4)$ connectivities are expected only for α-helices, and there is no evidence for $d_{\alpha N}(i,i+2)$ NOE connectivities, which would be indicative of 3_{10}-helices.

Extended polypeptide conformations are characterized by short $d_{\alpha N}$ distances and the absence of d_{NN} connectivities (Billeter *et al.*, 1982). For example, in the homeodomain the sequence Cys 39–Leu 40–Thr 41 has clearly a non-helical conformation, so that the breakpoint between the two helical segments is also well defined.

A fourth helical segment comprising the residues 55–60 is indicated by the continuous sequence of strong d_{NN} connectivities and the presence of $d_{\alpha N}(i,i+3)$ connectivities (Figure 1). However, this region also showed strong NOESY

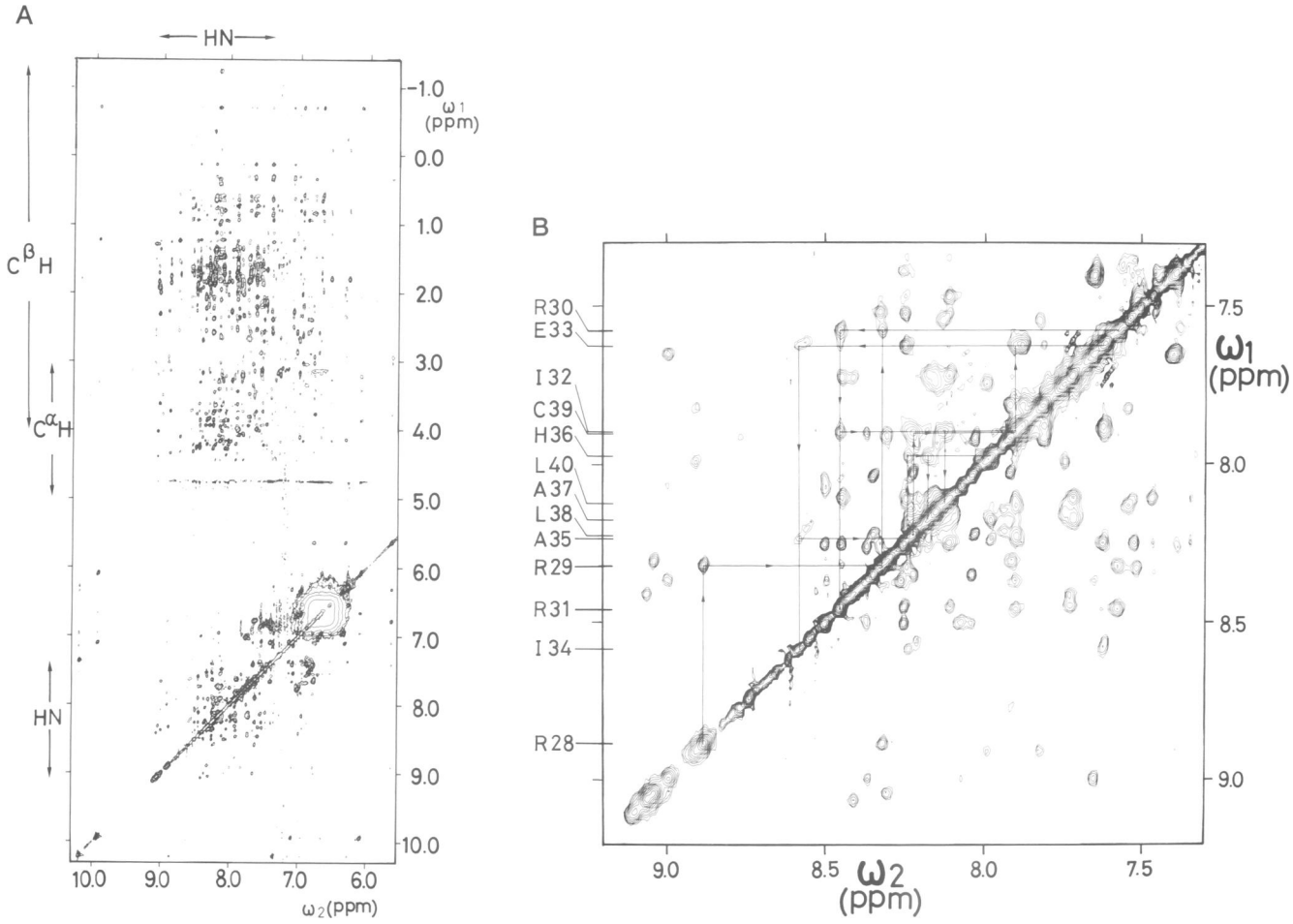

Fig. 2. NOESY spectrum of a 4 mM solution of the *Antp* homeodomain in a mixed solvent of 90% H_2O/10% D_2O, pH 4.3, 20°C, recorded at 600 MHz with a mixing time of 150 ms. (**A**) Spectral region with the resonances from all amide protons along ω_2, and the complete spectrum along ω_1. On the left side of the spectrum the ranges of resonance positions for the amide protons, C^αHs and C^βHs are indicated. In this spectrum the cross-peaks near ω_1 = 4.8 p.p.m. are bleached out due to water pre-saturation, but these missing peaks where observed in a spectrum recorded at 10°C. (**B**) Spectral region containing the amide-proton−amide-proton cross-peaks. The d_{NN} connectivities of the sequential assignment pathway in the helix 2 from residues 28 to 40 are identified with arrows which connect the corresponding cross-peaks.

cross-peaks corresponding to $d_{\alpha N}$, and no slowly exchanging amide protons could be identified in this peptide segment. We therefore defer a further characterization of this part of the polypeptide chain to the determination of the complete 3-dimensional structure. The same holds for the chain terminal segments 0−8 and 61−67, where the present qualitative pattern recognition approach does not yield unambiguous evidence for a particular secondary structure type. For these chain terminal segments there is also evidence for high flexibility, both from the narrow line shapes and from the fact that strong sequential connectivities $d_{\alpha N}$ and d_{NN} are simultaneously observed for many of the residues (Figure 1).

Computer modelling using the present, preliminary NMR data indicates that the homeodomain contains a helix-turn-helix motif

The DNA-binding properties of the homeodomain and the presence of the short segment of irregular structure between helices 2 and 3 (Figure 1) are suggestive of a structural relationship with a group of prokaryotic repressor proteins, which contain a conserved helix-turn-helix motif as the DNA recognition site (Ohlendorf *et al.*, 1982, 1983; Steitz *et al.*, 1982; Zuiderweg *et al.*, 1984; Ptashne, 1986; Mondragon *et al.*, 1988). Since the presently available NMR data are not sufficient as an input for a computation of the complete 3-dimensional structure of the homeodomain, we used a modeling approach to investigate the compatibility of the preliminary experimental observations with a helix-turn-helix motif of the type seen in the prokaryotic proteins. The starting point was crystal structure data on the 434 repressor and NMR data on the P22 c2 repressor. For the DNA-binding domain 1−69 of the 434 repressor a crystal structure was refined to 2.0 Å resolution (Mondragon *et al.*, 1988). For the DNA-binding domain 1−76 of P22 c2 repressor, NMR studies showed extensive homology of the 3-dimensional structure in solution with the crystal structure of the 434 repressor, in particular near the helix-turn-helix fragment (Otting, 1987; Otting *et al.*, 1987). On this basis we used the coordinates from the crystal structure of 434 repressor 1−69, to which hydrogen atoms were added using standard amino acid geometries. All proton−proton distances <5.0 Å were then identified in this structure, as well as the hydrogen bonds involving backbone amide protons. From this structural information we generated a hypothetical NMR data set for the polypeptide segment in the 434

Table I. Pairs of proton resonances, A and B, in the homeodomain, between which long-range and medium-range NOEs were observed that characterize the turn at residues 39–41 and the relative orientation of the two adjoining helices (Figure 1)

Resonance A	Resonance B
A 35 βCH_3	T 41 αH
A 35 βCH_3	E 42 NH
A 35 αH	L 40 NH
A 35 αH	I 45 δCH_3
T 41 NH	Q 44 βCH_2
T 41 βH	Q 44 NH
E 42 αH	I 45 NH
T 41 NH	Q 44 NH

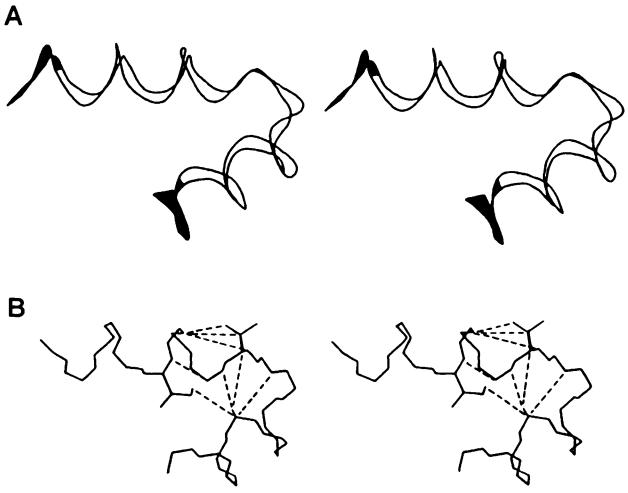

Fig. 3. Stereoviews illustrating the results of a preliminary structure determination for the *Antp* homeodomain polypeptide segment 28–52, which was obtained using NMR data and a modelling approach based on comparison with the homologous helix-turn-helix motif in 434 repressor. (**A**) Ribbon drawing of the polypeptide backbone in the crystal structure of the 434 repressor 1–69 (Mondragon *et al.*, 1988). Helix 2 and helix 3 have been extended by three and two residues, respectively (hatched segments), so that the length of the helices corresponds to that in the *Antp* homeodomain (Figure 1). (**B**) Backbone from the structure of 434 repressor 1–69 with side chains added for those residues which are conserved between 434 repressor and the *Antp* homeodomain. Broken lines indicate the medium-range and long-range NOEs observed in the homeodomain (Table I), which correspond in all cases to $^1H-^1H$ distances from 3.0 to 4.2 Å in the crystal structure of the 434 repressor 1–69.

repressor which corresponds to the residues 28–52 in the homeodomain (Figure 1). This data set consisted of the NOEs expected from near approach between pairs of hydrogen atoms in this polypeptide fragment, and of a list of the hydrogen-bonded and hence slowly exchanging amide protons along the sequence. The predicted, qualitative NMR data for the 434 repressor are nearly fully coincident with the experimental data for the homologous segment 31–50 in the homeodomain. This holds for the sequential and medium-range NOEs, as well as for the long-range NOEs that were so far identified (Table I). The data on slow exchange of amide protons in the segment 31–50 of the homeodomain (Figure 1) are consistent with those predicted from the crystal structure of 434 repressor 1–69 and with the pattern of slowly exchanging amide protons measured in P22 c2 repressor 1–76. The only exception is that the amide proton of Thr 41 in the homeodomain exchanges slowly, whereas fast exchange was observed for the homologous residue Ser 31 in P22 c2 repressor 1–76 (Otting, 1987).

The result of this combined use of NMR data and computer modelling is illustrated in Figure 3. The figure shows that the present preliminary experimental NOE data for the homeodomain (Table I) are fully compatible also with the crystal structure atom coordinates of the 434 repressor (from the aforementioned solution studies with P22 c2 repressor, there is strong evidence that this part of the repressor structure is preserved in non-denaturing aqueous media). The figure displays also a significant difference between the structures of the homeodomain and the 434 repressor, since the two helices adjacent to the turn at residues 39–41 (Figure 3A) are longer by two and three residues, respectively, than those in the helix-turn-helix motif of 434 repressor 1–69 (Figure 3B).

Conclusions

Figure 1 shows that most of the pairs of sequentially neighbouring amino acid residues in the homeodomain have been connected by two or even three sequential NOEs. This puts the sequence-specific resonance assignments for this protein on a firm basis (Billeter *et al.*, 1982). We have thus laid the foundation for a complete determination of the 3-dimensional structure of the homeodomain in solution (Wüthrich *et al.*, 1982).

It is an attractive aspect of the NMR method for protein structure determination that important information on the secondary structure results as a by-product of the resonance assignments (Wüthrich, 1986). For the homeodomain, this provided valuable, preliminary information on homologies of the molecular conformation with prokaryotic gene regulating proteins, which is summarized in Figure 3. We believe that this is the first direct experimental evidence for homologies between the spatial structures of eukaryotic and prokaryotic DNA-binding proteins. From a methodological point of view it may be of interest to add that all the data presented here were obtained using 13.5 mg of the protein, and that the recording and processing of the NMR spectra and the sequence-specific resonance assignments as presented in Figure 1 were completed within 5 weeks.

Materials and methods

The *Antp* homeodomain polypeptide of *Drosophila* was obtained from the corresponding DNA sequence inserted into a vector for expression and overproduction in *E.coli* (Müller *et al.*, 1988). The protein consists of a sequence of 68 amino acid residues, which is located at the C terminus in the *Antp* protein and has a mol. wt of 8545. The amino acid composition is 12 Arg, 8 Lys, 7 Glu, 6 Thr, 5 Leu, 4 Ile, 4 Gln, 3 Gly, 3 Tyr, 3 Phe, 3 Asn, 2 Met, 2 Ala, 2 Trp, 2 His, 1 Cys and 1 Pro. The sequence is shown in Figure 1.

All the spectroscopic studies were performed with a single sample of 13.5 mg of lyophilized protein. For the NMR measurements the protein was dissolved in water and the pH was adjusted to 4.3 by the addition of minute amounts of HCL. To prevent oxidation of the free cysteine side chain, we added small amounts of $Na_2S_2O_4$, removed oxygen by repeated freeze–thaw cycles under nitrogen, and finally sealed the sample under nitrogen. Spectra in H_2O were acquired in a mixed solvent of 90% H_2O/10% D_2O. Slowly exchanging amide protons were identified in a NOESY spectrum recorded at 10°C with the lyophilized protein freshly dissolved in D_2O. Before the acquisition of further spectra in D_2O, the amide protons were exchanged with deuterium by keeping the solution at room temperature for several days. After the experiments in D_2O the protein had somewhat deteriorated, as manifested by the appearance of

additional NMR lines. Further data collection for the determination of the complete 3-dimensional structure therefore had to be deferred until completion of a new protein preparation.

All ^{1}H-NMR spectra were recorded at 500 MHz on a Bruker AM 500 spectrometer, except for a NOESY spectrum in H_2O which was recorded on a Bruker AM 600 spectrometer. CD spectra were recorded on a JASCO J-500A spectropolarimeter, using a 0.02 mM protein solution.

Acknowledgements

We thank Dr S.Harrison for the coordinates of the 434 repressor 1–69 and a pre-print of a publication on this structure, Spectrospin AG., Fällanden, for measurement time on a 600 MHz NMR spectrometer, Peter Güntert for help with the computer programs used for the structural analysis of the 434 repressor 1–69, and Mrs E.Huber for the careful processing of the typescript. Financial support by the Schweizerischer Nationalfonds (project 3.198.85) is gratefully acknowledged.

References

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

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

Braunschweiler,L., Bodenhausen,G. and Ernst,R.R. (1983) *Mol. Phys.*, **48**, 535–560.

Griesinger,C., Otting,G., Wüthrich,K. and Ernst,R.R. (1988) *J. Am. Chem. Soc.*, in press.

Laughon,A. and Scott,P. (1984) *Nature*, **310**, 25–31.

Mondragon,A., Subbiah,S., Almo,S.C., Drottar,M. and Harrison,S.C. (1988) *J. Mol. Biol.*, in press.

Müller,M., Affolter,M., Leupin,W., Otting,G., Wüthrich,K. and Gehring,W. (1988) *EMBO J.*, **7**, 4299–4304

Müller,N., Ernst,R.R. and Wüthrich,K. (1986) *J. Am. Chem. Soc.*, **108**, 6482–6492.

Ohlendorf,D.H., Anderson,W.F., Fisher,R.G., Takeda,Y. and Matthews,B.W. (1982) *Nature*, **298**, 718–723.

Ohlendorf,D.H., Anderson,W.F., Lewis,M., Pabo,C.O. and Matthews,B.W. (1983) *J. Mol. Biol.*, **169**, 757–769.

Otting,G. (1987) Ph.D. Thesis, ETH Zürich Nr. 8314.

Otting,G., Leupin,W., Glauser,M., Eugster,A., Senn,H. and Wüthrich,K. (1987) Abstracts of the 9th International Biophysics Congress, Jerusalem, p. 61.

Ptashne,M. (1986) *A Genetic Switch, Gene Control and Phage* λ. Cell Press and Blackwell, Cambridge.

Rance,M., Sørensen,O.W., Bodenhausen,G., Wagner,G., Ernst,R.R. and Wüthrich,K. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 479–485.

Sheperd,J.C.W., McGinnis,W., Carrasco,A.E., De Robertis,E.M. and Gehring,W. (1984) *Nature*, **310**, 70–71.

Steitz,T.A., Ohlendorf,D.H., McKay,D.B., Anderson,W.F. and Matthews,B.W. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 3097–3100.

Wagner,G. and Wüthrich,K. (1982a) *J. Mol. Biol.*, **155**, 347–366.

Wagner,G. and Wüthrich,K. (1982b) *J. Mol. Biol.*, **160**, 343–361.

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

Wüthrich,K., Wider,G., Wagner,G. and Braun,W. (1982) *J. Mol. Biol.*, **155**, 311–319.

Wüthrich,K., Billeter,M. and Braun,W. (1984) *J. Mol. Biol.*, **180**, 715–740.

Zuiderweg,E.R.P., Billeter,M., Boelens,R., Scheek,R.M., Wüthrich,K. and Kaptein,R. (1984) *FEBS Lett.*, **174**, 243–247.

Received on August 29, 1988; revised on September 28, 1988